



RESEARCH ARTICLE

How Reliable, How Certain? A Bayesian Reliability Analysis of the Short Dark Tetrad (SD4) Among Indonesian University Students

Aflah Zakinov Irta^{1*)}, Rizal Kurniawan², Anindra Guspa³

Published online: 29 July 2026

Abstract: The Dark Tetrad (subclinical narcissism, Machiavellianism, psychopathy, and sadism) is a central framework in personality psychology and is studied across clinical, organisational, and forensic contexts. The Short Dark Tetrad (SD4) has been adapted into several languages, but Indonesian studies have reported its reliability only as point estimates, without quantifying estimation uncertainty. This study therefore estimated the internal-consistency reliability of the Indonesian SD4 within a Bayesian framework. Responses from 289 Indonesian university students were analysed in JASP (1,000 MCMC samples; 50 burn-in) through a four-step sequence: MCMC trace plots, Posterior Predictive Check (PPP) eigenvalue plots, Bayesian single-factor fit diagnostics, and McDonald's ω posterior distributions with 95% Bayesian credible intervals (BCI). All chains converged, the eigenvalue evidence supported a dominant single factor (F1/F2 ratio = 2.5 to 3.6), and model fit was adequate to excellent (B-CFI = .905 to .965). Reliability was good for Sadism (ω = .813) and Narcissism (ω = .808) and acceptable for Machiavellianism (ω = .725) and Psychopathy (ω = .717). The Indonesian SD4 therefore shows acceptable-to-good internal-consistency reliability, reported here with explicit uncertainty and accompanied by a reproducible Bayesian reporting template. Because this evidence concerns reliability rather than validity, it supports research use in university-student samples, whereas high-stakes application will require further validity and measurement-invariance evidence.

Keywords: Bayesian Reliability; Credible Interval; Dark Personality; Internal Consistency; McDonald's Omega; Short Dark Tetrad

Abstrak: Dark Tetrad (narsisisme subklinis, Machiavellianisme, psikopati, dan sadisme) merupakan kerangka sentral dalam psikologi kepribadian dan dikaji pada konteks klinis, organisasi, serta forensik. Short Dark Tetrad (SD4) telah diadaptasi ke dalam beberapa bahasa, tetapi studi di Indonesia baru melaporkan reliabilitasnya sebagai estimasi titik tanpa mengukur ketidakpastian estimasi. Oleh karena itu, penelitian ini mengestimasi reliabilitas konsistensi internal SD4 versi Indonesia dalam kerangka Bayesian. Data dari 289 mahasiswa Indonesia dianalisis menggunakan JASP (1.000 sampel MCMC; burn-in 50) melalui empat langkah berurutan: plot jejak (trace plot) MCMC, plot nilai eigen Posterior Predictive Check (PPP), diagnostik kecocokan model faktor tunggal Bayesian, dan distribusi posterior ω McDonald beserta interval kredibel Bayesian 95% (BCI). Seluruh rantai konvergen, bukti nilai eigen mendukung satu faktor dominan (rasio F1/F2 = 2,5 hingga 3,6), dan kecocokan model tergolong memadai hingga sangat baik (B-CFI = 0,905 hingga 0,965). Reliabilitas tergolong baik untuk Sadisme (ω = 0,813) dan Narsisisme (ω = 0,808), serta dapat diterima untuk Machiavellianisme (ω = 0,725) dan Psikopati (ω = 0,717). Dengan demikian, SD4 versi Indonesia menunjukkan reliabilitas konsistensi internal yang dapat diterima hingga baik, dilaporkan disertai ketidakpastian yang eksplisit dan templat pelaporan Bayesian yang dapat direproduksi. Karena bukti ini menyangkut reliabilitas, bukan validitas, temuan ini mendukung penggunaan untuk keperluan penelitian pada sampel mahasiswa, sedangkan penerapan berisiko tinggi memerlukan bukti validitas dan invariansi pengukuran lebih lanjut.

Kata kunci: Interval Kredibel; Kepribadian Gelap; Konsistensi Internal; Omega McDonald; Reliabilitas Bayesian; Short Dark Tetrad

^{1*),2} Department of Psychology, Faculty Psychology and Health, Universitas Negeri Padang, Indonesia

³ Department of Psychology, Faculty of Human Development, Universiti Pendidikan Sultan Idris, Malaysia

**) corresponding author*

Aflah Zakinov Irta, S.Psi., M.Si
Department of Psychology, Faculty Psychology and Health,
Universitas Negeri Padang, Indonesia
Komplek Cendana, Blok F 3, Alai Parak Kopi, Padang,
Sumatra Barat - Indonesia

INTRODUCTION

Research on socially aversive personality has increasingly focused on the construct of dark personality, commonly conceptualized within the Dark Tetrad framework, which encompasses subclinical narcissism, Machiavellianism, psychopathy, and sadism (Paulhus & Williams, 2002; Paulhus et al., 2021). These traits are linked to negative interpersonal, organisational, and societal outcomes, including deception, exploitative behaviour, workplace counterproductivity, cyberbullying, academic

dishonesty, intimate partner violence, and forensic misconduct (Gómez-Leal et al., 2024; Plouffe et al., 2023; Rassin et al., 2024). The formal incorporation of everyday sadism as the fourth Dark Tetrad constituent followed Buckels et al.'s (2013) demonstration that sadism is conceptually and empirically distinguishable from and provides incremental predictive validity beyond the original Dark Triad for aggressive and exploitative behaviours.

Paulhus et al.'s (2021) Short Dark Tetrad (SD4) advanced measurement of the construct: a 28-item inventory of four seven-item subscales. The SD4 addressed critical measurement limitations of its predecessor, the Short Dark Triad (SD3; Jones & Paulhus, 2014), by equalising subscale lengths and refining item content to reduce construct overlap. Cross-cultural validation has proceeded across several languages: Chinese (Zhang et al., 2022), Spanish (Ortet-Walker et al., 2024), Brazilian (Guilhermino et al., 2025), Indian (Dubey et al., 2025), and Polish (Kosowski et al., 2026) samples. Plouffe et al.'s (2023) systematic review, conducted using COSMIN methodology (Mokkink et al., 2010), cautioned that psychometric quality varies across cultures and that each adaptation needs independent validation.

Indonesia, the fourth most populous nation, has a collectivist cultural profile marked by strong social-harmony norms, high power distance, and communal identity (Hofstede, 2001), and it remains an under-researched context for Dark Tetrad assessment. Normative constraints on overt grandiosity and interpersonal exploitation differ from those in Western individualist cultures, so the SD4 needs psychometric evaluation in Indonesia before applied use. Two preliminary Indonesian studies have initiated this process: Irta et al. (2025) conducted cross-cultural adaptation following International Test Commission (2017) guidelines on 258 university students, yielding acceptable confirmatory factor analytic fit ($GFI = .975$, $CFI = .81$, $RMSEA = .073$) and aggregate McDonald's $\omega = .806$; Arnietta et al. (2025) reported Cronbach's α coefficients ranging from .621 to .752 on 305 students. While both studies represent substantive initial contributions, they share a critical shared limitation: exclusive reliance on point-estimate reliability coefficients without quantifying the sampling uncertainty present in all psychometric estimation, a practice that recent work has criticised as inadequate (Pfadt et al., 2022, 2023; Sijtsma, 2009; Peters, 2014).

Dark personality traits may be expressed, and reported, differently in Indonesia. In collectivistic settings with strong social-harmony norms and relatively high power distance, overt grandiosity, callousness, and interpersonal exploitation draw stronger normative sanction, which can shape both how these traits are expressed and how readily they are disclosed on self-report measures. Two response processes are especially relevant here. First, socially desirable responding and impression management can lead respondents to under-report undesirable attributes. Second, culturally patterned response styles, such as acquiescence and a tendency to avoid extreme endorsement of socially proscribed content, can further compress item responses. Because both processes reduce item variance and inter-item covariance, they affect internal-consistency reliability independently of the trait itself and call for uncertainty-aware reliability appraisal in this population. This expectation aligns with cross-cultural evidence linking high power distance and collectivism to acquiescent and less extreme responding (Johnson et al., 2005; Smith, 2004).

An Indonesian SD4 already exists: prior work adapted the instrument cross-culturally and reported classical reliability (Arnietta et al., 2025; Irta et al., 2025). The gap lies

in how that reliability was reported. Both studies give point estimates with no measure of estimation uncertainty. This study fills that gap by providing Bayesian reliability evidence for the four Indonesian SD4 subscales: McDonald's ω with credible intervals, supported by convergence and model-plausibility diagnostics.

Point-estimate reliability reporting has practical costs, not just technical ones. A reliability coefficient reported without an uncertainty interval implies a false precision and hides how wide the plausible range for the true population reliability can be. In high-stakes settings such as personnel selection, clinical screening, and forensic risk assessment, this can weaken the validity of decisions. A related issue is the common use of Cronbach's α as the only reliability index. Coefficient α is not inherently wrong, and under tau-equivalence it is an appropriate estimator; but when tau-equivalence does not hold, which is common in personality scales such as the SD4, whose items carry substantively varying factor loadings within each subscale, α can underestimate reliability (Flora, 2020; McDonald, 1999; Revelle & Condon, 2019). Reporting α alone therefore shows neither this potential bias nor the sampling uncertainty around the estimate, which is why a model-based index (McDonald's ω) with an explicit uncertainty interval is preferable.

Bayesian reliability analysis in the open-source JASP software (Pfadt et al., 2022, 2023) addresses both limitations. By modelling the posterior distribution of reliability coefficients through Markov Chain Monte Carlo (MCMC) sampling conditional on observed data and uninformative priors, the Bayesian framework yields: (a) McDonald's ω theoretically superior reliability index that is unbiased under the congeneric model; (b) 95% Bayesian Credible Intervals (BCI) with the direct probabilistic interpretation that the true population reliability lies within the stated range with 95% posterior probability, which frequentist confidence intervals cannot provide (Morey et al., 2016); (c) MCMC trace plots enabling direct visual verification of chain convergence and stationarity; and (d) Posterior Predictive Check (PPP) eigenvalue plots that assess the plausibility of the single-factor model underlying ω estimation. Together these outputs give a fuller reliability evaluation than any single point estimate.

A distinctive methodological strength of Bayesian reliability analysis is that its outputs naturally follow a logical inferential sequence: MCMC trace plots must first verify that the sampling algorithm has adequately explored the posterior before any inference is drawn; PPP eigenvalue plots must then establish that the single-factor model assumed in ω estimation is plausible; Bayesian fit indices subsequently quantify model adequacy; and only then do the final ω estimates carry interpretive weight. This order follows the dependency structure of Bayesian inference and gives a level of transparency that classical reliability reporting lacks.

The present study addresses that gap by applying the full sequential Bayesian framework to the Indonesian SD4. Beyond its psychometric contribution, it demonstrates a reproducible analytical protocol, implementable in freely available software, that can serve as a template for uncertainty-aware reliability reporting in Indonesian psychological science. The evidence here concerns internal-consistency reliability, which is a precondition for, but not the same as, the validity evidence (such as criterion-related validity and measurement invariance) that clinical, organisational, or forensic use would also require. The objective is therefore specific: to provide Bayesian internal-

consistency reliability evidence for all four Indonesian SD4 subscales in a university-student sample, reported through four sequential steps: (1) MCMC convergence diagnostics, (2) Posterior Predictive Check (PPP) eigenvalue assessment of single-factor plausibility, (3) Bayesian single-factor fit diagnostics, and (4) posterior McDonald's ω estimation with 95% credible intervals.

METHODS

Participant characteristics and research design

A cross-sectional quantitative psychometric design was employed, consistent with contemporary standards for instrument validation research (Boateng et al., 2018). The analytical framework centred on Bayesian reliability analysis with Markov Chain Monte Carlo (MCMC) sampling, with sequential diagnostic verification preceding the final reliability estimates. Data were collected in a naturalistic setting through an online self-report questionnaire.

Sampling procedures

Participants were active Indonesian university students recruited through purposive online convenience sampling via institutional channels. The questionnaire was administered online using Google Forms over a data-collection period spanning July to September 2025. Inclusion criteria specified: (a) active enrolment in a recognised Indonesian higher education institution; (b) age 18–35 years; and (c) provision of informed electronic consent. A total of 289 responses were received; none were excluded, yielding a final analysed sample of $N = 289$. Because the online questionnaire required a response to every item, the dataset contained no missing values; no attention-check or careless-responding screening was applied. Because recruitment was non-probability (purposive/convenience), claims of generalisability are restricted to Indonesian university-student populations comparable to the present sample and are not extended to the broader Indonesian adult population. The study received institutional ethical approval under reference number 21.02/KEPKH-UNP/V/2024 and was conducted in accordance with the Declaration of Helsinki. Sample demographic characteristics were as follows. The mean age was 20.01 years ($SD = 2.65$). Of the 289 participants, 61 (21.1%) were men and 228 (78.9%) were women. These characteristics define the population to which the present reliability evidence applies.

Sample size, power, and precision

Within a Bayesian framework, sample adequacy is more appropriately framed in terms of posterior precision than of frequentist statistical power. The present $N = 289$ was evaluated against two considerations: the sample size recommended for stable Bayesian reliability estimation and adequate MCMC convergence under congeneric single-factor models (Pfadt et al., 2022), and participant-to-parameter ratio guidance for seven-item subscales (Kyriazos, 2018). Adequacy was then assessed directly through the width of the resulting 95% credible intervals, which index the precision with which each ω is estimated; the observed interval widths (reported in the Results) were sufficiently narrow to support interpretable reliability conclusions for this sample. We therefore use the terms

“sample adequacy” and “posterior precision” rather than “power.”

Measures

The SD4 was developed by Paulhus et al. (2021). This study used the Indonesian adaptation of Irta et al. (2025), produced under the International Test Commission (2017) translation and adaptation guidelines. It comprises 28 items across four seven-item subscales (Machiavellianism, Narcissism, Psychopathy, and Sadism), each rated on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree), with higher scores indicating stronger endorsement of the trait. Because the items came from the present authors' own preliminary adaptation study (Irta et al., 2025), no additional permission was needed, and they were administered verbatim. Following the default JASP single-test reliability implementation (Pfadt et al., 2022, 2023), the five-point responses were treated as continuous and modelled with the Pearson covariance matrix in the congeneric single-factor model underlying McDonald's ω .

Data analysis

Three independent MCMC chains were sampled using the Gibbs sampler. Each chain was run for 1,000 post-burn-in iterations following a burn-in of 50 iterations. Because trace plots alone are insufficient to establish convergence, convergence was additionally evaluated using numerical diagnostics: the potential scale reduction factor (R-hat; Gelman & Rubin, 1992), the effective sample size (ESS) of each ω posterior. Convergence was considered satisfactory where $R\text{-hat} \leq 1.01$ and ESS was adequate for stable interval estimation. All chains met these criteria: R-hat ranged from 1.000 to 1.004 (all ≤ 1.01), and effective sample sizes were adequate across subscales (Machiavellianism, $R\text{-hat} = 1.001$, $ESS = 1,127$; Narcissism, $R\text{-hat} = 1.000$, $ESS = 2,124$; Psychopathy, $R\text{-hat} = 1.001$, $ESS = 1,833$; Sadism, $R\text{-hat} = 1.004$, $ESS = 2,135$). The choice of 1,000 post-burn-in samples and 50 burn-in iterations followed the JASP defaults validated by Pfadt et al. (2022, 2023); their adequacy for the present data was confirmed by the numerical diagnostics above rather than assumed. For the prior, an inverse-Wishart prior was placed on the covariance matrix with the JASP default hyperparameters (an identity scale matrix with degrees of freedom equal to the number of items plus one, $df = k + 1$); this weakly informative specification yields posterior estimates substantively equivalent to frequentist bootstrap intervals under unrestrictive conditions, so that substantive conclusions are driven by the data rather than by the prior. A formal prior sensitivity analysis was not conducted; because this weakly informative prior yields estimates equivalent to frequentist bootstrap intervals under unrestrictive conditions (Pfadt et al., 2022), the reported ω estimates are unlikely to be materially driven by the prior, and a formal sensitivity check remains a useful confirmation for future work. Analyses were then conducted and reported in the following sequential order, which reflects the logical dependency structure of Bayesian reliability inference:

MCMC Convergence Verification. Trace plots were inspected across three independent chains to confirm stationary mixing and absence of trending, non-stationarity, or chain divergence (Gelman & Rubin, 1992). Convergence is a mandatory prerequisite for valid posterior inference; inferences drawn from non-converged chains are unreliable regardless of their apparent plausibility.

Posterior Predictive Check (PPP) of Single-Factor Model Plausibility. PPP eigenvalue plots were examined to

evaluate whether the single-factor model underlying ω estimation is plausible given the observed inter-item correlation structure. A dominant first eigenvalue substantially exceeding the second, with non-overlapping 95% uncertainty intervals, provides Bayesian posterior evidence supporting the unidimensionality assumption.

Bayesian Single-Factor Fit Diagnostics. Bayesian fit indices (B-LR, B-SRMR, B-RMSEA, B-CFI, B-TLI) were evaluated against a priori acceptable criteria: B-CFI $\geq .90$, B-TLI $\geq .90$, B-RMSEA $\leq .08$, B-SRMR $\leq .08$ (Hu & Bentler, 1999). Stringent criteria (B-CFI $\geq .95$, B-RMSEA $\leq .06$) were additionally evaluated.

Bayesian McDonald's ω Reliability Estimation. Only after Steps 1–3 confirmed sampling validity and single-factor model adequacy were McDonald's ω posterior distributions extracted. Primary outputs included: posterior mean estimates of ω_t , 95% Bayesian credible intervals (BCI), and full posterior density plots visualising distributional shape and precision. Reliability benchmarks followed Nunnally (1978): $\omega \geq .70$ = acceptable; $\omega \geq .80$ = good.

All analyses were conducted in JASP Version 0.17 (JASP Team, 2023) using the Bayesian Reliability module (Pfadt et al., 2023). McDonald's total omega (ω_t) was selected as the

primary reliability index rather than Cronbach's α , given its theoretical superiority under the congeneric single-factor model and its robustness to violations of tau-equivalence (Flora, 2020; McDonald, 1999; Revelle & Condon, 2019). The Bayesian estimation approach was preferred over frequentist point-estimate approaches for three convergent reasons: (a) it yields full posterior distributions rather than single-valued estimates; (b) it produces 95% Bayesian credible intervals with direct probabilistic interpretations; and (c) it supports visual diagnostic verification of sampling convergence and model plausibility through trace and PPP plots (Morey et al., 2016; Pfadt et al., 2022; Wagenmakers et al., 2018).

RESULTS OF STUDY

MCMC Convergence Verification

Figure 1 presents MCMC trace plots for all four Indonesian SD4 subscales, displaying three independent chains (blue, red, green) across 1,000 post-burn-in iterations.

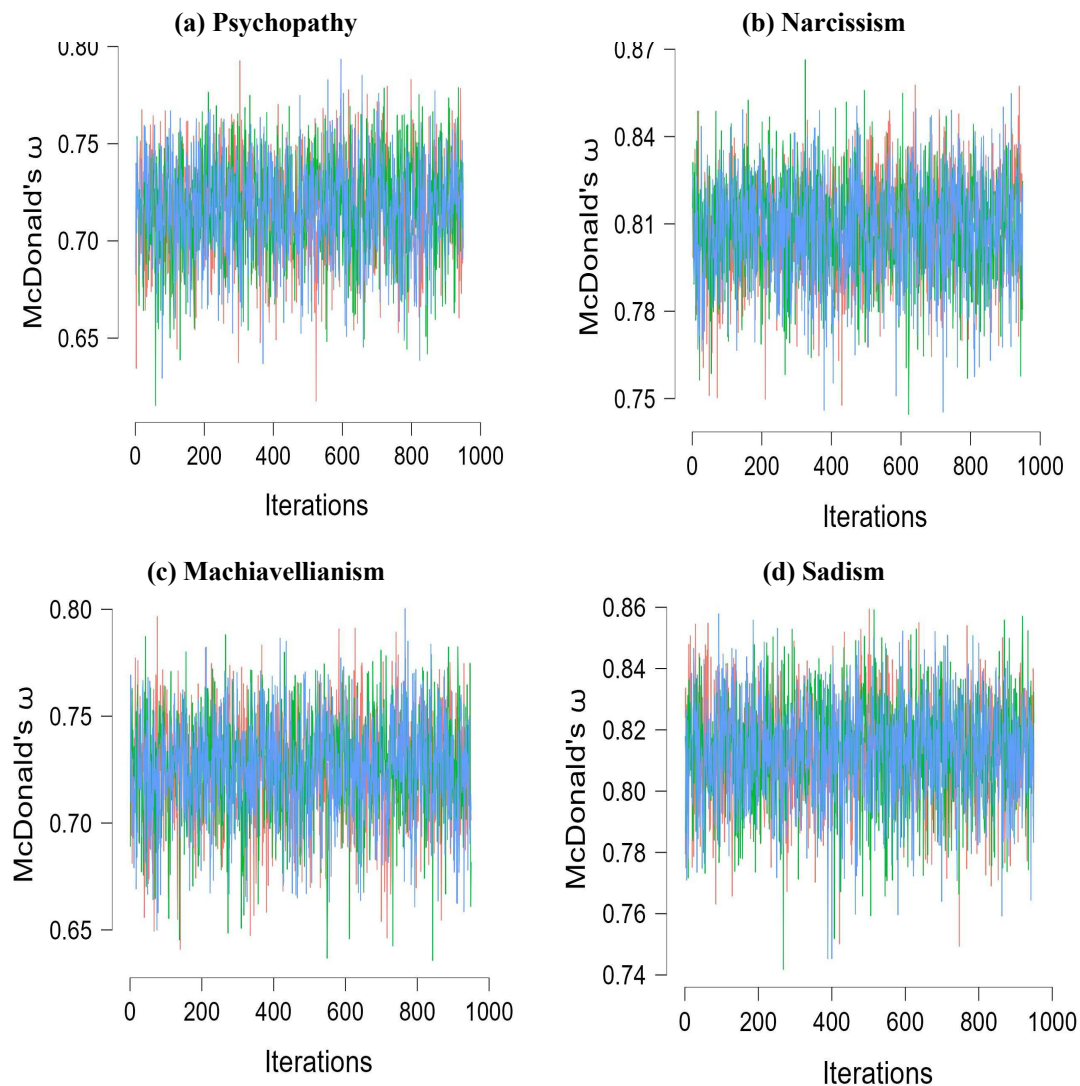


Figure 1. MCMC Trace Plots for McDonald's ω – Indonesian SD4 Subscales

Note. Each panel shows three independent MCMC chains for one subscale's ω posterior across 1,000 post-burn-in iterations. Convergence is indicated by stable horizontal mixing without systematic drift and by free overlap of the three chains.

All chains showed stationary mixing across all four subscales, fluctuating around a stable level, with no systematic drift and free overlap of the three chains throughout sampling. Narcissism traces (Figure 1b) oscillate within $\omega \approx .75-.87$ around the posterior mean of .808; Sadism (Figure 1d) within .74–.86 around .813; Psychopathy (Figure 1a) within .65–.80 around .717; and Machiavellianism (Figure 1c) within .65–.80 around .725.

Table 1 summarises the MCMC convergence diagnostics. The stationary mixing pattern across all subscales satisfies the Gelman–Rubin convergence criterion (Gelman & Rubin, 1992) and supports the Bayesian inferences reported in Steps 2–4. Table 1 reports the numerical convergence diagnostics (R-hat and effective sample size) for each subscale; all R-hat values were ≤ 1.01 with adequate effective sample sizes, documenting convergence quantitatively.

Posterior Predictive Check (PPP) Eigenvalue Assessment

Figure 2 presents Bayesian PPP eigenvalue plots evaluating the plausibility of the single-factor model that underlies McDonald's ω computation. Each plot displays posterior mean eigenvalues (black dots) with 95% posterior uncertainty intervals (grey bars) for the inter-item correlation matrix of the respective seven-item subscale. The theoretical interpretive standard follows from factor analytic conventions (Kaiser, 1960): a dominant first eigenvalue substantially exceeding 1.0, with the second eigenvalue near or below 1.0, indicates unidimensional latent structure. The Bayesian extension (Pfadt et al., 2023) further requires that the 95% uncertainty intervals for the first and second eigenvalues do not meaningfully overlap.

Table 1. MCMC Convergence Summary for Indonesian SD4 Subscales

Subscale	Chain Range (ω)	Posterior Mean	R-hat	ESS
Psychopathy	.65 – .80	.717	1.001	1.833
Narcissism	.75 – .87	.808	1.000	2.124
Machiavellianism	.65 – .80	.725	1.001	1.127
Sadism	.74 – .86	.813	1.004	2.135

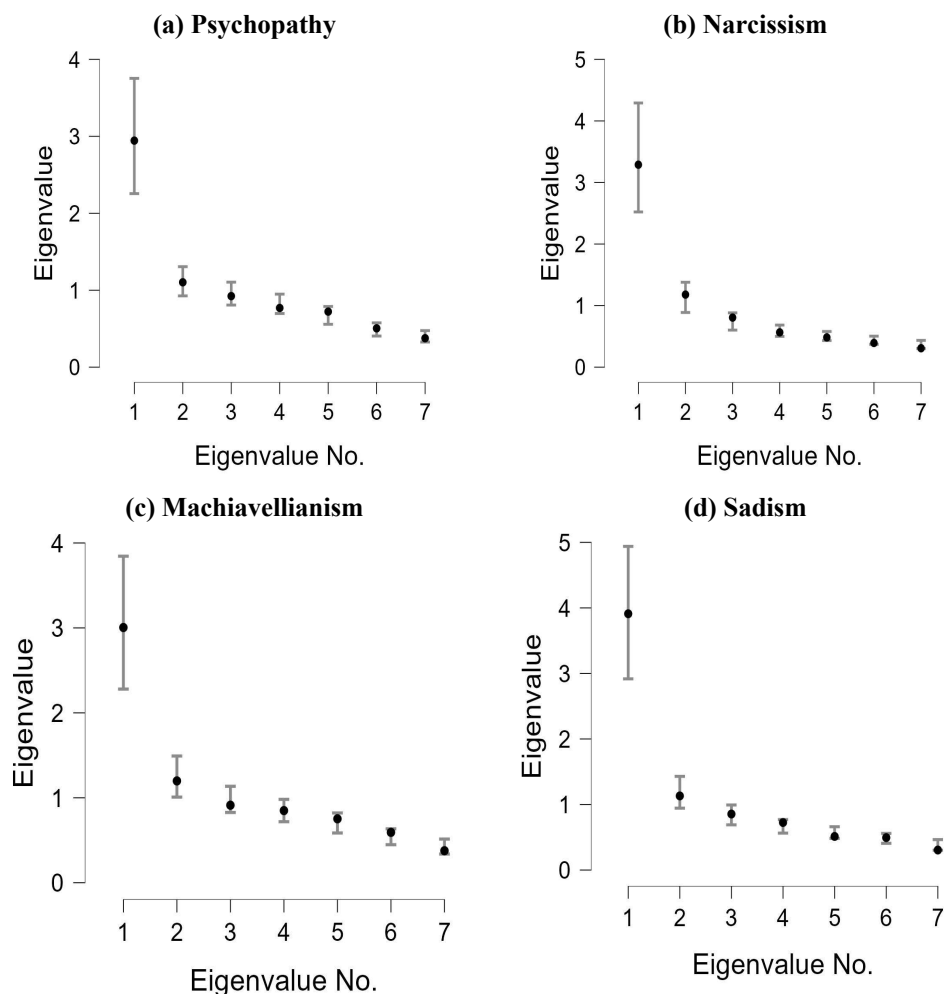


Figure 2. Posterior Predictive Check (PPP) Eigenvalue Plots – Indonesian SD4 Subscales

The Sadism subscale (Figure 2d) exhibited the strongest unidimensional evidence, with the largest first eigenvalue, a pronounced drop to the second, and non-overlapping 95% credible intervals for the first and second eigenvalues. As an approximate reading of the plots, $F1 \approx 4.0$ and $F2 \approx 1.1$ ($F1/F2 \approx 3.6$). This pattern corroborates the comparatively high Bayesian reliability of Sadism ($\omega = .813$) reported in Step 4. Psychopathy (Figure 2a) and Machiavellianism (Figure 2c) showed adequate ratios (approximately 2.7 and 2.5, respectively; $F1 \approx 3.0$) with marked drop-offs to subsequent eigenvalues. Narcissism (Figure 2b) showed $F1 \approx 3.3$ and $F1/F2 \approx 2.7$; however, the second through fourth eigenvalues remained near or slightly above 1.0 with partially overlapping credible intervals, providing mild but not conclusive posterior evidence of a secondary factor. This pattern should be interpreted cautiously. Because the SD4 Narcissism items were not designed to operationalise grandiose and vulnerable narcissism as separate facets, the mild secondary signal is not direct evidence of those specific facets; it is better described as non-specific structural complexity, broadly consistent with the documented multidimensionality of subclinical narcissism (Ortet-Walker et al., 2024; Plouffe et al., 2023). Table 2 summarises the PPP eigenvalue estimates.

Bayesian Single-Factor Fit Diagnostics

Having confirmed MCMC convergence (Step 1) and established single-factor model plausibility (Step 2), Bayesian fit diagnostics were evaluated to provide quantitative confirmation of model adequacy. Table 3 presents Bayesian fit indices for all four subscales. Psychopathy demonstrated excellent fit ($B-CFI = .956$, $B-TLI = .933$, $B-RMSEA = .056$, $B-SRMR = .043$), meeting the stringent criteria established by Hu and Bentler (1999) on both $B-CFI (\geq .95)$ and $B-RMSEA (\leq .06)$. Machiavellianism similarly achieved excellent fit ($B-CFI = .965$, $B-TLI = .945$, $B-RMSEA = .056$, $B-SRMR = .039$), with the lowest $B-SRMR$ value of all subscales—indicating the tightest alignment between the posited single-factor model and the observed inter-item covariance structure. Sadism demonstrated good fit ($B-CFI = .913$, $B-TLI = .901$, $B-RMSEA = .068$, $B-SRMR = .063$), meeting all acceptable thresholds with comfortable margins.

The Narcissism subscale showed a more mixed fit pattern. While $B-CFI = .905$ and $B-RMSEA = .064$ met acceptable criteria, $B-TLI = .851$ fell marginally below the .90 threshold recommended by Hu and Bentler (1999). This finding converges with the mild secondary eigenvalue signal observed in Step 2 (Figure 2b), providing two independent diagnostic indicators that the Narcissism subscale exhibits mild structural complexity not fully

captured by the strict single-factor model. Nevertheless, three of five indices ($B-CFI$, $B-RMSEA$, $B-SRMR$) meet acceptable thresholds, supporting the continued interpretability of single-factor ω estimation for this subscale while flagging the need for caution in high-stakes applications. Steps 1–3 together provide the basis for the Step 4 reliability estimates.

Bayesian McDonald's ω Reliability Estimates

With MCMC convergence confirmed (Step 1), single-factor model plausibility established (Step 2), and model fit quantitatively verified (Step 3), the reliability estimates can now be interpreted. Table 4 presents posterior mean estimates of McDonald's ω and 95% Bayesian credible intervals for all four Indonesian SD4 subscales, and Figure 3 displays the corresponding posterior density plots.

The Sadism subscale yielded the highest reliability ($\omega = .813$, 95% BCI [.777, .843]) and the narrowest credible interval (BCI width = .066), so it was also the most precisely estimated. The whole credible interval lies above $\omega = .77$, above the conventional benchmark for good reliability (Nunnally, 1978). As shown in Figure 3d, the posterior density is near-Gaussian and symmetric, corroborating the unidimensional structural evidence from Step 2 ($F1/F2$ ratio ≈ 3.6) and the good fit from Step 3 ($B-CFI = .913$). Across all three diagnostic steps, Sadism was the strongest subscale of the Indonesian SD4.

The Narcissism subscale demonstrated equally good reliability ($\omega = .808$, 95% BCI [.774, .841]; BCI width = .067). However, Figure 3b reveals a mild left-skewness in the posterior density, an asymmetry that is a third converging indicator, alongside the mild secondary eigenvalue in Step 2 and the marginal $B-TLI$ in Step 3, of latent structural complexity not fully captured by the single-factor model. These three indicators together suggest that, although Narcissism shows good overall reliability, its multidimensional structure deserves attention in applied use.

The Machiavellianism subscale demonstrated acceptable reliability ($\omega = .725$, 95% BCI [.678, .772]; BCI width = .094). Figure 3c shows a unimodal, approximately normal posterior distribution with moderate width, and the lower BCI bound of .678 falls close to but below the .70 threshold, indicating limited posterior probability of unacceptably low true reliability. This result, combined with Machiavellianism's excellent fit in Step 3 ($B-CFI = .965$), suggests that the modest reliability reflects the inherently broader bandwidth of the Machiavellianism construct rather than structural misspecification (McDonald, 1999).

Table 3. Bayesian Single-Factor Fit Indices for Indonesian SD4 Subscales

Subscale	B-LR	B-SRMR	B-RMSEA	B-CFI	B-TLI
Psychopathy	26.61	.043	.056	.956	.933
Narcissism	29.21	.058	.064	.905	.851
Machiavellianism	26.39	.039	.056	.965	.945
Sadism	31.87	.063	.068	.913	.901

Note. All estimates are posterior means from JASP Version 0.17 (Pfadt et al., 2023). Acceptable fit criteria (Hu & Bentler, 1999): $B-CFI \geq .90$, $B-TLI \geq .90$, $B-RMSEA \leq .08$, $B-SRMR \leq .08$. Stringent criteria: $B-CFI \geq .95$, $B-RMSEA \leq .06$. $B-LR$ = Bayesian Likelihood Ratio; $B-SRMR$ = Bayesian Standardised Root Mean Square Residual; $B-RMSEA$ = Bayesian Root Mean Square Error of Approximation; $B-CFI$ = Bayesian Comparative Fit Index; $B-TLI$ = Bayesian Tucker–Lewis Index.

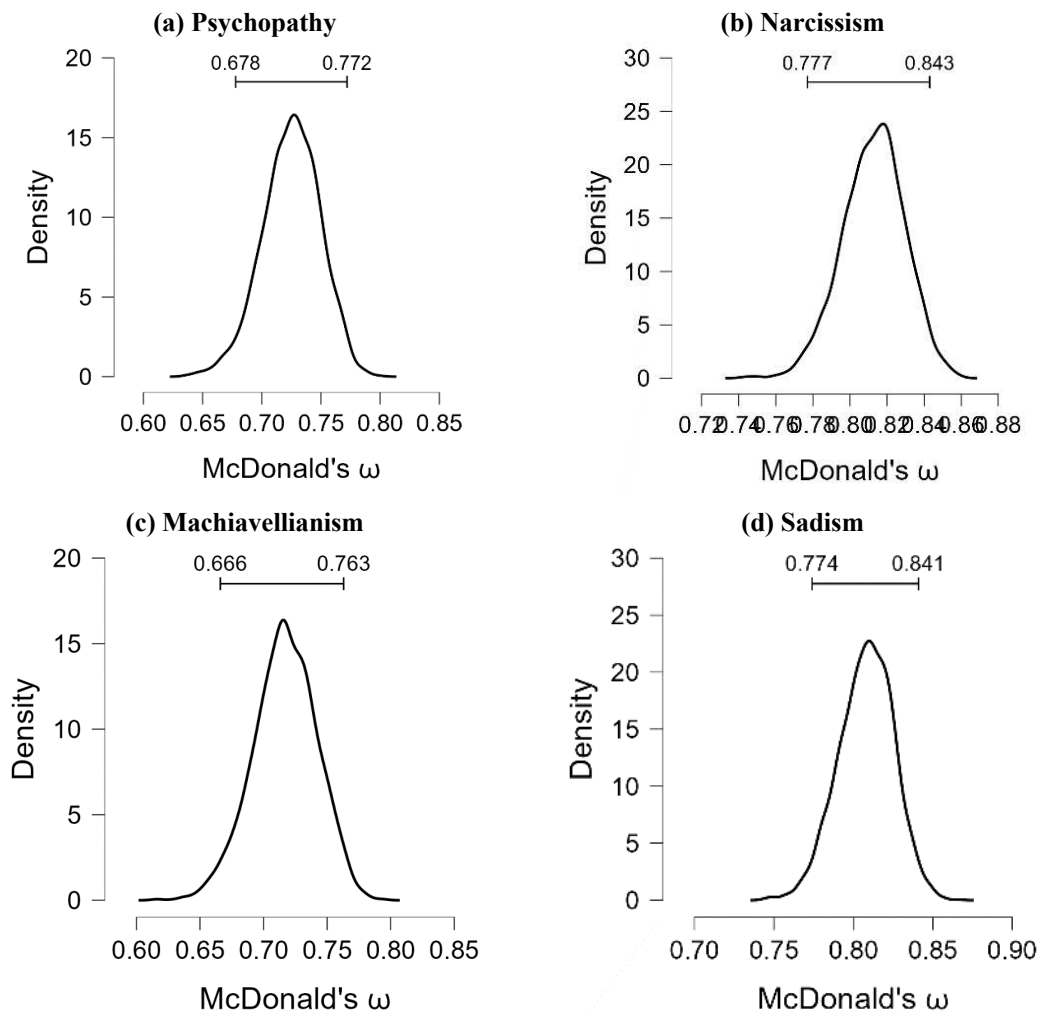


Figure 3. Bayesian Posterior Density Plots of McDonald's ω – Indonesian SD4 Subscales

Note. Each panel shows the posterior density of McDonald's ω for one subscale; the central tendency is the posterior mean and the spread reflects estimation uncertainty, summarised by the 95% credible interval.

Table 4. Bayesian McDonald's ω Reliability Estimates for Indonesian SD4 Subscales (N = 289)

Subscale	ω	Lower 95%	Upper 95%	Reliability Category
Psychopathy	.717	.666	.763	Acceptable (BCI width = .097)
Narcissism	.808	.774	.841	Good (BCI width = .067)
Machiavellianism	.725	.678	.772	Acceptable (BCI width = .094)
Sadism	.813	.777	.843	Good (BCI width = .066)

The Psychopathy subscale yielded acceptable reliability ($\omega = .717$, 95% BCI [.666, .763]; BCI width = .097)—the widest BCI of all subscales, indicating the greatest posterior estimation uncertainty (see table 4). The lower BCI bound of .666 falls marginally below the .70 threshold, suggesting some posterior probability that true population reliability in the Indonesian context is slightly lower than conventional acceptability criteria.

As visualised in Figure 3a, the posterior distribution is nonetheless unimodal and symmetric. Psychopathy showed the widest credible interval (BCI width = .097), indicating the greatest posterior estimation uncertainty among the four subscales. Because the present study did not conduct

item-level analysis or tests of differential item functioning, the source of this wider interval cannot be identified here; it is therefore offered only as a hypothesis for future research. Candidate explanations include psychometric factors such as greater item-content heterogeneity or lower and more variable item-total relations within the seven-item subscale, and, possibly, culturally patterned responding to items describing callous disregard for social norms. Adjudicating between these explanations would require item-level and measurement-invariance analyses (e.g., DIF) that are beyond the scope of the present reliability study.

DISCUSSION

The present study provides the first sequentially structured Bayesian psychometric evaluation of the Indonesian SD4, applying a four-step inferential procedure that reflects the logical dependency architecture of Bayesian reliability analysis. This sequential analytic framework, encompassing MCMC convergence diagnostics, model plausibility and fit evaluation, and subsequent reliability estimation, ensures that the derived ω estimates are supported by a thoroughly validated inferential process. To date, this level of inferential transparency has not been applied to the Indonesian SD4, thereby representing a substantive methodological advancement beyond the point-estimate approaches adopted in prior Indonesian studies (Arnietta et al., 2025; Irta et al., 2025).

A key finding with direct theoretical implications is the convergence of three diagnostic indicators for the Narcissism subscale. The presence of a mild secondary eigenvalue in Step 2 (PPP: $F1/F2 \approx 2.7$, with the $F2$ uncertainty interval overlapping eigenvalue 1.0), the marginally suboptimal B-TLI in Step 3 (.851), and the left-skewed posterior density in Step 4 (Figure 3b) collectively point to a consistent pattern. Taken together, these independent indicators suggest that the Indonesian Narcissism subscale demonstrates a degree of structural complexity, aligning with the well-documented multidimensional nature of subclinical narcissism across cultural contexts. This finding converges with observations from Spanish (Ortet-Walker et al., 2024), Chinese (Zhang et al., 2022), and Polish (Kosowski et al., 2026) SD4 adaptations that the Narcissism subscale frequently shows marginal structural indices compared to its tetrad companions. This structural complexity should be interpreted cautiously and without strong cultural claims. Because the SD4 Narcissism items were not constructed to separate grandiose and vulnerable facets, and because the present study conducted no item-level or measurement-invariance analysis, the pattern is best read as non-specific multidimensionality consistent with prior SD4 adaptations, rather than as evidence of culturally specific facet salience. Whether collectivistic norms differentially shape narcissism items is a question for future studies using item-level and DIF analyses (Ortet-Walker et al., 2024; Paulhus et al., 2021).

Bayesian credible intervals add information that point estimates leave out. The contrast between Sadism's narrow BCI (width = .066) and Psychopathy's wider BCI (width = .097) is directly interpretable as a difference in posterior certainty: the data support greater confidence in Sadism's reliability than in Psychopathy's, even though both posterior means meet acceptability thresholds. Such nuance is directly consequential for applied instrument selection: a researcher selecting a subscale for a high-stakes screening application should reasonably prefer the subscale with demonstrably greater estimation precision, all else being equal. This type of decision-relevant information cannot be obtained from frequentist confidence intervals, which carry no direct probabilistic interpretation about the location of population parameters (Morey et al., 2016).

The Indonesian reliabilities are broadly consistent with other adaptations (Table 5). The Sadism estimate ($\omega = .813$) is close to the Spanish value ($\omega = .79$; Ortet-Walker et al., 2024), and the Narcissism estimate ($\omega = .808$) meets or exceeds several counterparts. The lower Machiavellianism ($\omega = .725$) and Psychopathy ($\omega = .717$) values fall within the range reported across SD4 studies and resemble the pattern in the Indian Dark Tetrad at Work validation (Dubey et al., 2025), which suggests these two subscales are more variable in reliability across samples, likely reflecting

construct-specific item functioning rather than a weakness of the Indonesian adaptation.

The study also shows that Bayesian reliability analysis, covering MCMC convergence diagnostics, PPP eigenvalue assessment, and credible-interval estimation, is reproducible in free software (JASP). This accessibility is critical for Indonesian psychological science, where methodological advancement is frequently constrained by software licensing costs and technical barriers. The sequential template shown here can be adopted across personality, educational, clinical, and organisational research, improving the quality of reliability evidence reported in Indonesia.

The findings apply mainly to research use of the Indonesian SD4. The Sadism and Narcissism subscales gave the most precise reliability estimates, whereas for Machiavellianism and Psychopathy researchers should report the credible interval alongside the point estimate and note the wider posterior uncertainty. Because the present evidence concerns internal-consistency reliability in a single university-student sample, and not criterion-related validity, predictive validity, or measurement invariance, it does not by itself support using the SD4 in high-stakes settings such as clinical screening, forensic assessment, or personnel evaluation. Those uses would need further validity and invariance evidence; the reliability framework here is a necessary first step toward them, not a sufficient basis for them (Pfadt et al., 2023; Plouffe et al., 2023).

Three limitations should be noted. First, the sample was limited to university students, which may constrain generalisability to Indonesian adults with different educational and cultural backgrounds. Second, the analysis reports subscale-level reliability but not measurement invariance across gender, age, or region, which is the next step for valid group comparisons. Third, criterion-related validity for the Indonesian SD4 is still absent; establishing it, by linking SD4 scores to behavioural, clinical, or organisational outcomes, is a priority for applied use.

Future research priorities follow directly from these limitations: (a) extend the Bayesian reliability framework to diverse Indonesian populations including working adults, clinical samples, and forensic samples; (b) conduct Bayesian measurement invariance testing across gender, age, and regional subgroups; (c) investigate bifactor modelling for the Narcissism subscale to formally test whether a hierarchical factor structure better represents its internal organisation given the convergent three-step evidence of mild structural complexity; (d) establish criterion validity through concurrent and predictive validation studies; and (e) examine differential item functioning analyses to identify items whose functioning may be culturally modulated, particularly within the Psychopathy subscale.

A further question is why the Sadism and Narcissism subscales showed higher reliability than the Machiavellianism and Psychopathy subscales. The present design does not permit a definitive answer, because no item-level analysis was conducted; the following accounts are therefore advanced only as hypotheses for future item-level work. Sadism and Narcissism items tend to describe relatively homogeneous, affectively salient content (e.g., enjoyment of others' suffering; self-enhancement), which can yield more consistent inter-item covariation and thus higher internal consistency. Machiavellianism is, by design, a broad-bandwidth construct spanning cynicism, strategic manipulation, and long-term planning; such breadth can depress internal consistency even when content validity is high (McDonald, 1999). Psychopathy items span impulsivity and callousness, which may be endorsed less consistently in a normative student sample. Each account is testable

through item-total correlations, item-level factor loadings, and DIF analyses, which we recommend for future research.

Reliability and validity are distinct: internal-consistency reliability quantifies the precision and consistency of measurement, whereas validity concerns whether scores represent the intended construct and support intended inferences. The Bayesian ω estimates reported here speak only to the former. They do not establish criterion-related validity, predictive validity, or measurement invariance across groups, and the single-factor checks reported here provide only limited structural-validity information. Accordingly, the contribution of this study is reliability evidence for research use, not a comprehensive validation of the Indonesian SD4.

CONCLUSION AND IMPLICATIONS

A sequentially conducted Bayesian reliability analysis showed that the Indonesian SD4 has acceptable-to-good internal-consistency reliability across all four subscales in a university-student sample. After confirming MCMC convergence (Step 1), single-factor plausibility through PPP eigenvalue assessment (Step 2), and adequate Bayesian model fit (Step 3), the posterior omega estimates (Step 4) placed Sadism and Narcissism in the good range and Machiavellianism and Psychopathy in the acceptable range, so reliability is highest for Sadism and Narcissism. These results support research use of the instrument in Indonesian university samples. They do not, on their own, validate it for all assessment contexts, and high-stakes use would require further validity and measurement-invariance evidence. The study contributes uncertainty-quantified reliability evidence for the Indonesian SD4 and a reproducible sequential Bayesian reporting template for Indonesian psychometric research.

Recommendation. On the basis of these findings, studies using the Indonesian SD4 should report McDonald's omega together with its credible interval rather than a single point estimate, so that measurement uncertainty stays visible. The Sadism and Narcissism subscales can be interpreted with greater confidence, and Machiavellianism and Psychopathy more cautiously. Until validity and invariance evidence is available, the instrument should be used for research rather than for high-stakes individual decisions such as selection or clinical.

DECLARATIONS

Conflict of Interest Statement

The authors declare that they have no conflicts of interest to disclose. This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement

The findings of this study are supported by data that can be obtained by contacting the corresponding author, NF, upon reasonable request.

Generative artificial intelligence

To enhance the linguistic clarity and overall readability of this manuscript, the authors utilized Claude AI (Anthropic) and DeepL for paraphrasing and grammatical refinement. All AI-assisted modifications were strictly reviewed and edited by the authors, who assume full

responsibility for the integrity and final content of this publication.

Informed consent

Written consent was obtained from all participants after they received a thorough explanation of the research procedures, potential risks, and expected benefits. Furthermore, all data management and handling processes were carried out in accordance with the applicable ethical standards in psychological research.

Funding

This research did not receive funding from any party.

AUTHORS' CONTRIBUTIONS.

Aflah Zakinov Irta: Conceptualization; Methodology; Formal Analysis; Validation; Visualization; Writing Original Draft; Data Analysis; Writing Review & Editing.

Rizal Kurniawan: Conceptualization; Investigation; Data Curation; Resources; Project Administration; Data Analysis.

Anindra Guspa: Literature Review; Methodology; Validation; Writing – Review & Editing.

ABOUT THE AUTHORS

Aflah Zakinov Irta completed a Master of Science in Personality Psychology at the University of Indonesia and is currently a lecturer and researcher in psychology at Universitas Negeri Padang, where he is affiliated with the Psychometrics-based research group. His research interests focus on psychometric assessment and personality measurement.

Rizal Kurniawan holds a Master of Arts in Psychology from Universitas Gadjah Mada and is currently a researcher and lecturer at Universitas Negeri Padang. His research concentration lies in social psychology.

Anindra Guspa is currently in her second semester as a PhD student in Psychology at Universitas Pendidikan Sultan Idris (UPSI), Malaysia, and also serves as a lecturer in the Psychology Department at Universitas Negeri Padang.

REFERENCES

- Arnietta, S. N., Natalya, L., & Siaputra, I. B. (2025). Exploring the shadows: Validation of the Short Dark Tetrad (SD4) in Indonesian. *Psikoborneo: Jurnal Ilmiah Psikologi*, 13(2), 218–225. <https://doi.org/10.30872/psikoborneo.v13i2.19167>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Buckels, E. E., Jones, D. N., & Paulhus, D. L. (2013). Behavioral confirmation of everyday sadism. *Psychological Science*, 24(11), 2201–2209. <https://doi.org/10.1177/0956797613490749>
- Dubey, A., Choubisa, R., & Philipose, J. V. (2025). Psychometric validation of Dark Tetrad at Work scale in an Indian context. *Acta Psychologica*, 255, Article 104918. <https://doi.org/10.1016/j.actpsy.2025.104918>

- Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, 3(4), 484–501. <https://doi.org/10.1177/2515245920951747>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–472. <https://doi.org/10.1214/ss/1177011136>
- Gómez-Leal, R., Fernández-Berrocal, P., Gutiérrez-Cobo, M. J., Cabello, R., & Megías-Robles, A. (2024). The Dark Tetrad: Analysis of profiles and relationship with the Big Five personality factors. *Scientific Reports*, 14, Article 4443. <https://doi.org/10.1038/s41598-024-55074-w>
- Guilhermino, C. S., Bonfá-Araujo, B., de Azevedo, T. G., Silva, J. L. P., Hauck-Filho, N., Bandeira, P. F. R., Tondin, C. F., & Alvarenga, M. A. S. (2025). Cross-cultural adaptation of the Short Dark Tetrad (SD4) in the Brazilian context. *Psychological Test Adaptation and Development*, 6(1), 77–82. <https://doi.org/10.1027/2698-1866/a000100>
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Sage.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd ed.). <https://www.intestcom.org>
- Irta, A. Z., Adri, Z., Lubis, S. I. S., Akbar, Z. M., & Ramadhini, Z. S. (2025). Cross-cultural adaptation of the dark personality instrument for the Indonesian context: A preliminary study. In *Proceedings of the 3rd International Conference on Psychology and Health Issues, ICoPHI 2024, 2 November 2024, Padang, West Sumatera, Indonesia*. EAI. <https://doi.org/10.4108/eai.2-11-2024.2354545>
- JASP Team. (2023). JASP (Version 0.17) [Computer software]. <https://jasp-stats.org>
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-cultural psychology*, 36(2), 264–277. <https://doi.org/10.1177/0022022104272905>
- 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>
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>
- Kosowski, P., & Janeczke, P. (2026). Psychometric Properties of the Polish Adaptation of the Short Dark Tetrad (SD4-PL). *Journal of clinical psychology*, 82(3), 273–286. <https://doi.org/10.1002/jclp.70066>
- Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 19(4), 539–549. <https://doi.org/10.1007/s11136-010-9606-8>
- Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103–123. <https://doi.org/10.3758/s13423-015-0947-8>
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
- Ortet-Walker, J., Garofalo, C., Vidal-Arenas, V., Bogaerts, S., Mezquita, L., Ortet, G., & Ibáñez, M. I. (2024). The Spanish Short Dark Tetrad (SD4): Association with personality and psychological problems. *Psicothema*, 36(2), 195–204. <https://doi.org/10.7334/psicothema2023.172>
- 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., & 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](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity: Why and how to abandon Cronbach's alpha and the route towards more comprehensive assessment of scale quality. *European Health Psychologist*, 16(2), 56–69. <https://www.ehps.net/ehp/index.php/contents/article/view/ehp.v16.i2.p56>
- Pfadt, J. M., van den Bergh, D., Sijsma, K., Moshagen, M., & Wagenmakers, E.-J. (2022). Bayesian estimation of single-test reliability coefficients. *Multivariate Behavioral Research*, 57(4), 620–641. <https://doi.org/10.1080/00273171.2021.1891855>
- Pfadt, J. M., van den Bergh, D., Sijsma, K., Moshagen, M., & Wagenmakers, E.-J. (2023). A tutorial on Bayesian single-test reliability analysis with JASP. *Behavior Research Methods*, 55(3), 1069–1078. <https://doi.org/10.3758/s13428-021-01778-0>
- Plouffe, R. A., Saklofske, D. H., & Smith, M. M. (2023). Assessing Dark Tetrad traits: A COSMIN systematic review of measurement instruments available. *Personality and Individual Differences*, 214, Article 112330. <https://doi.org/10.1016/j.paid.2023.112330>
- 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, Article 311. <https://doi.org/10.1038/s41598-023-50968-7>
- Revelle, W., & Condon, D. M. (2019). Reliability from α to ω : A tutorial. *Psychological Assessment*, 31(12), 1395–1411. <https://doi.org/10.1037/pas0000754>
- Sijsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107–120. <https://doi.org/10.1007/s11336-008-9101-0>

- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of cross-cultural psychology*, 35(1), 50-61. <https://doi.org/10.1177/0022022103260380>
- Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25(1), 35–57. <https://doi.org/10.3758/s13423-017-1343-3>
- Zhang, J., Zhu, M., Yue, X., & Chen, Z. (2022). The super-short Dark Tetrad: Development and validation within the Chinese context. *Personality and Individual Differences*, 184, Article 111195. <https://doi.org/10.1016/j.paid.2021.111195>

ADDITIONAL INFORMATION

Correspondence All inquiries and requests for additional materials should be directed to the Corresponding Author.

Publisher's Note Utan Kayu Publishing maintains a neutral stance regarding territorial claims depicted in published maps and does not endorse or reject the institutional affiliations stated by the authors.

Open Access This article is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0), which permits others to share, adapt, and redistribute the material in any medium or format, even for commercial purposes, provided appropriate credit is given to the original author(s) and the source, a link to the license is provided, and any changes made are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original. To view a copy of this license, visit <https://creativecommons.org/licenses/by-sa/4.0/>.

© The Author(s) 2026

This page has been intentionally left blank