Personality and Individual Differences 250 (2026) 113551

Contents lists available at ScienceDirect

Personality and Individual Differences

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

Symptom over- and underreporting are relatively stable behavioral opposites in forensic patients and undergraduates

Daniël van Helvoort [a,b,*], Henry Otgaar [a,c], Chijs van Nieuwenhuizen [b,d], Harald Merckelbach [a]

[a] Clinical Psychological Science, Maastricht University, the Netherlands
[b] GGzE Institute for Mental Health Care, Eindhoven, the Netherlands
[c] Leuvens Institute of Criminology, Catholic University of Leuven, the Netherlands
[d] Tranzo, Scientific Center for Care and Wellbeing, Tilburg University, the Netherlands

ARTICLE INFO

Keywords:
Symptom validity
Symptom overreporting
Symptom underreporting
Inattentive responding
Traits

ABSTRACT

The phenomena of symptom overreporting, underreporting, and inattentive responding have often been linked to traits using cross-sectional designs. The preliminary question – the temporal stability of these phenomena – has remained largely unexamined. We conducted a test-retest study in forensic inpatients and students (all $ns \geq 64$), who twice – across a six- to ten-week interval – completed stand-alone measures of symptom overreporting, symptom underreporting, inattentive responding, and Big Five personality domains. Symptom overreporting rates were similar in patients and students (±5 %). The same was true for symptom underreporting in patients (±15–20 %) and students (±5–10 %), whereas inattentive responding was non-trivial in patients (±15 %), but low in students (±0–5 %). In both groups, symptom overreporting and underreporting were relatively stable behavioral opposites. Trait scores of those who engaged in symptom overreporting and underreporting deviated ±0.75 SD from those who passed all validity indices, and in a patterned rather than arbitrary way. The shared variance may reflect a common underlying mechanism. In patients, borderline personality disorder was linked to symptom overreporting, whereas narcissistic personality disorder was associated with symptom underreporting. Overall, our findings suggest that both traits and contextual factors shape how accurately people report their symptoms.

1. Introduction

Distorted symptom expressions – including symptom overreporting, symptom underreporting, and inattentive responding – may compromise the accuracy of psychological assessments in, for example, occupational, clinical or forensic settings (e.g., Lee et al., 2019; Melchers et al., 2020). Research on trait correlates of distorted self-reporting has heavily focused on feigning outside the symptom domain, including socially desirable responding (e.g., exaggerating virtues) and general life adjustment. Aside from studies on MMPI validity indicators (e.g., Novo et al., 2022), the more drastic phenomena of symptom overreporting (i.e., exaggeration/fabrication of symptoms) and symptom underreporting (i.e., minimization/denial of symptoms) and their links with traits have received less attention (Van Helvoort et al., 2022).

1.1. Traits and distorted symptom expression

Context – and the incentives it provides – affects symptom over- and underreporting. For instance, in forensic samples, symptom overreporting is common during pre-trial, but may shift to underreporting during incarceration (Walters, 1988). Beyond context, a scoping review (Van Helvoort et al., 2022) found reasons to assume that traits play a role. In particular, depressivity, dissociation, and alexithymia – correlates of neuroticism – appeared to be inducive to overreporting (e.g., Duberstein & Heisel, 2007), whereas narcissism was, albeit inconsistently, related to underreporting. Also, conscientiousness has been associated with underreporting and social desirability, though causal pathways remain uncertain (e.g., De Vries et al., 2014). Low openness – along with low conscientiousness and low agreeableness – has been related to inattentive responding, but findings are inconsistent (e.g., Berry et al., 2019). Most studies are cross-sectional, sample-dependent,

[*] Corresponding author at: Section Forensic Psychology, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD, Maastricht, the Netherlands.
E-mail address: d.vanhelvoort@maastrichtuniversity.nl (D. van Helvoort).

https://doi.org/10.1016/j.paid.2025.113551
Received 28 July 2025; Received in revised form 3 November 2025; Accepted 7 November 2025

and difficult to interpret in terms of causality (e.g., Holden & Passey, 2010; Van Helvoort et al., 2022).

1.2. *Stability is suggestive of trait contribution*

One key question is whether symptom over- and underreporting are stable across time – a prerequisite for attributing trait-like properties to these phenomena. Germane to this is a study by An et al. (2012; p. 854) in which performance validity tests were administered twice to students ($N = 36$), concluding that "participants who exerted poor effort at one point in time also exerted poor effort at a later time. This could indicate a trait-like component." As the authors noted, this is only suggestive of trait involvement, as alternative explanations remain possible. For example, inattentive responding might spuriously obscure (validity) test indices, a point noted by DeRight and Jorgensen (2015) who replicated the findings of An et al. (2012).

1.3. *Traits and 'mixed feigning' cases*

A related issue is to what extent symptom overreporting and underreporting are distinct phenomena. Historically, the broader response styles of faking bad and good have been conceptualized as opposite ends of a continuum (Rogers, 2008). Meanwhile, it is conceivable they are not always mutually exclusive. For example, in civil litigation, it is plausible that plaintiffs may exaggerate case-related symptoms (e.g., memory impairments), while emphasizing virtues to appear honest. Such hybrid response styles have been observed in employment applicants (Melchers et al., 2020), yet remain rarely investigated in the clinical domain. Notable exceptions include a study identifying concurrent symptom overreporting and socially desirable responding in disability claimants (Whitman et al., 2023) and a simulation study relying on students showing that instructed 'mixed' symptom over- and underreporting can produce validity test scores alike those of honest respondents (Bošković et al., 2024). Arguably, only when the relationship between symptom over- and underreporting is well-understood can a robust theory of the associations with traits be formulated.

1.4. *Current study*

The studies by An et al. (2012) and DeRight and Jorgensen (2015) are commendable for their test-retest design. However, they did not simultaneously consider symptom overreporting, underreporting, and inattentive responding along with their trait correlates. With this in mind, we conducted a test-retest study to examine these phenomena and their stability in two contexts with differing motivational contingencies: one with relatively higher and assumedly increasing incentives – forensic psychiatric inpatients at intake (T1) and six to ten weeks later near their first treatment evaluation (T2), where assessment results may influence treatment-related decisions – and one with relatively lower incentives: undergraduate students.

1.4.1. *Hypothesis 1*

We expected symptom overreporting and underreporting to be more prevalent in the forensic inpatient sample compared with the student sample (Van Impelen et al., 2016), and inattentive responding to be more common in the student sample (Meade & Craig, 2012).

1.4.2. *Hypothesis 2*

We tested whether symptom overreporting and symptom underreporting are negatively related to each other (i.e., behavioral opposites; Rogers, 2008).

1.4.3. *Hypothesis 3*

We examined whether in both subsample contexts all three types of distorted symptom presentation show temporal stability.

1.4.4. *Hypothesis 4*

We expected symptom overreporting to related to 'neuroticism', symptom underreporting to 'conscientiousness', and inattentive responding to 'openness'.

2. Methods

2.1. *Participants*

Forensic inpatients were recruited from De Woenselse Poort, a high-security forensic psychiatric hospital in Eindhoven, the Netherlands, providing court-imposed treatment to mentally disordered offenders who committed violent crimes. Students were recruited from the Catholic University of Leuven, Belgium, and Maastricht University, the Netherlands, and participated voluntarily or for course credits/small financial compensation. The study was approved by the Ethics Review Committee of the Faculty of Psychology and Neuroscience, Maastricht University and the Institute for Mental Health Care Eindhoven (GGzE), the Netherlands. We preregistered the study (see https://osf.io/4qxuw).

2.1.1. *Eligibility*

All forensic inpatients that enrolled between April 1, 2022, and August 27, 2024 were screened for eligibility by (first author) DH, a licensed (forensic) psychologist. Supplemental File A provides a flow chart of the screening procedure and outlines the exclusion criteria (prior diagnoses of severe cognitive disorders and suspected states that could impair the ability to complete questionnaires). Recruitment ended when $n = 64$ patients and students completed T2, as required for our confirmatory analyses. See Supplemental File B for our statistical approach.

2.1.2. *Background data*

As part of standard assessment at intake (T1), 78 patients participated ($M_{age} = 35.6$, $SD = 9.66$, range = 20–61, 70 men). After a six- to ten-week interval (T2), 14 patients had dropped out due to uncooperativeness. Thus 64 patient participants (82.1 %) remained ($M_{age} = 36.3$; $SD = 9.98$, range = 20–61, 57 men). At T1, 127 students completed the questionnaires ($M_{age} = 22.2$; $SD = 2.93$, range = 18–40, 24 men). At T2, 63 students opted out of participation, with 64 students (50.4 %) remaining ($M_{age} = 22.2$; $SD = 2.54$, range = 19–37, 16 men).

In accordance with the GGzE quality statute, patients provided passive informed consent to use of their assessment data for research purposes. A subsample of 36 patients provided written informed consent to the use of treatment details and DSM-5 diagnoses (made by trained clinicians prior to their admission) from their patient files (see Supplemental File C for the data).

2.2. *Measures and materials*

2.2.1. *Structured Inventory of Malingered Symptomatology (SIMS) – 21-items*

The 21-item SIMS (Orrù et al., 2022) is a short version of the SIMS (Smith & Burger, 1997), a 75-item true/false keyed self-report questionnaire tapping into symptom overreporting (i.e., endorsement of bizarre/unlikely symptoms). The full-scale SIMS has satisfactory psychometric properties, but tends to generate a relatively high number of false positives in individuals with genuine severe psychopathology (Shura et al., 2022). The items of the short version were selected by Orrù et al. (2022) through machine learning.[1] We considered the summed 21-item SIMS a dimensional index of symptom overreporting. As to prevalence rates, we estimated full-scale SIMS scores[2] and used a conservative cut-point (i.e., >23 pseudosymptoms; Shura et al., 2022). In the present study, aggregated across samples and time points, αs for 21-item SIMS were 0.53–0.68. Test-retest *r*s were 0.63 (patients) and 0.62 (students).

2.2.2. Supernormality Scale (SS)

The SS is a 37-item true/false keyed self-report scale measuring denial of mundane problems and symptoms (Cima et al., 2003). 'Yes' answers are summed – after reversed coding and disregarding distractor items – to obtain scores for the social desirability subscale (11 items), the supernormality subscale (21 items), and the combination of both scales (total SS). We considered the supernormality subscale a dimensional index of symptom underreporting. The SS has sound psychometric properties, though its accuracy is suboptimal. For prevalence rate estimates, we used the stringent cut-point (≥22 on the total SS) to maintain a specificity of ≥0.90. Across our samples and time points, αs for total SS were 0.77–0.84 and test-retest *r*s were 0.81 (patients) and 0.82 (students).

2.2.3. Infrequency Scale (IFS)

The IFS (Chapman et al., 1982) is a 13-item true/false keyed measure of (in)attentive responding, indexing endorsement of highly infrequent experiences. Endorsement of >2 items indicates inattentive responding. We administered the IFS in three blocks of four to five items, interspersed with the other questionnaires. In the current study, test-retest *r*s were 0.43 (patients) and 0.35 (students).

2.2.4. Big Five Inventory–2 Short Form (BFI-2-S)

The BFI-2-S (Soto & John, 2017) is an abbreviated version of the BFI-2, and gauges 'negative emotionality' (formerly known as neuroticism), 'extraversion', 'open-mindedness' (formerly known as openness), 'agreeableness', and 'conscientiousness'. The BFI-2-S comprises 30 items – six items per dimension – on a 5-point scale, ranging from (1) 'disagree strongly' to (5) 'agree strongly'. The BFI-2-S shows satisfactory psychometric properties (Soto & John, 2017). In our study, αs for the domain subscales, taken together for both time points, varied from 0.62 to 0.81 (patients) and 0.54 to 0.86 (students). Test-retest *r*s for the subscales ranged from 0.66 to 0.78 (patients) and 0.82 to 0.91 (students).

2.2.5. Additional measures

We also administered the 18-item *Cognitive Emotion Regulation Questionnaire –short version* (CERQ-short; Garnefski & Kraaij, 2006) and 20-item *Behavioral Emotion Regulation Questionnaire* (BERQ; Kraaij & Garnefski, 2019), which measure cognitive and behavioral coping strategies, respectively. The data obtained with these scales will be considered elsewhere.

2.3. Procedure

Eligible patients (see 2.1.1) were approached by DH within a week of their hospital admission to establish rapport and schedule the assessment. Within two weeks of admission, DH conducted the first assessment (T1): patients completed a 30–45 min booklet about various symptoms and their ways of coping with stress. They were explicitly told that their responses might contribute to decisions on treatment intensity, ward placement, and crisis management. The questionnaires were completed on a laptop using Qualtrics. After six to ten weeks – near their first treatment evaluation (i.e., when these clinical decisions are made) – DH revisited the patients, who then completed the survey again (T2). Thus, at T2, response consequences were likely evaluated as more imminent. Students, invited by course coordinators and following informed consent, completed the same Qualtrics survey twice, over a six- to ten-week interval. They were informed that the questionnaires concerned personal experiences and symptoms and were sometimes administered in settings relevant to their coursework.

Participants completed one of four randomly assigned orders twice and consistently. The orders counterbalanced the sequence of the first two (i.e., the 21-items SIMS and the SS) and last two scales (i.e., the CERQ-short and BERQ) of the battery, with the BFI-2-S administered in-between. This design aimed at minimizing test-order carry-over effects (Meade & Craig, 2012).

3. Results[3]

We checked whether the four different orders affected the results, which was not the case. These and other checks on the data are detailed in Supplemental File B.

3.1. Planned analyses

3.1.1. Group differences[4]

As to the 21-item SIMS scores (i.e., overreporting tendency), forensic inpatients did not statistically differ from students, neither at T1, nor at T2: $t(203) = 0.17$, $p = .43$, $d = 0.02$, and $t(126) = -1.10$, $p = .14$, $d = 0.19$, respectively. Similarly, on the SS supernormality subscale scores (i.e., underreporting tendency), neither at T1, nor at T2 did patients and students statistically significantly differ from each other: $t(203) = 0.68$, $p = .25$, $d = 0.10$, and $t(126) = 1.29$, $p = .10$, $d = 0.23$. In categorical terms, symptom overreporting (i.e., based on SIMS cutoff) ranged between 3 and 6 % across groups and times; symptom underreporting (based on SS cutoff) between 4 % and 23 %. No statistically significant group differences emerged.

Regarding IFS scores (i.e., inattentive responding), 13 of 78 forensic inpatients (16.7 %) exceeded the cut-point of >2 at T1, as compared to 8 of 127 students (6.3 %). This difference was not statistically significant after applying Bonferroni correction[3], $\chi^2(1, N = 205) = 5.65$, $p = .02$ $V = 0.17$. However, at T2, 9 out of 64 forensic inpatients (14.1 %) exceeded the cut-point against 0 out of 64 students (0 %). This difference was statistically significant, $\chi^2(1, N = 128) = 9.68$, $p = .002$, $V = 0.28$. Table 1 shows group differences with regard to the *total* scale scores, including rates of participants exceeding cut-points, and the stability of scores.

3.1.2. Correlations between symptom over- and underreporting indices

In forensic inpatients, the 21-item SIMS and SS supernormality scores were inversely and statistically significantly correlated, both at T1 and T2: $r(76) = -0.50$, $p < .001$, 95 % CI [−Inf, −0.34] and $r(62) = -0.53$, $p < .001$, 95 % CI [−Inf, −0.35], respectively. A similar pattern emerged for students at T1 and T2: $r(125) = -0.42$, $p < .001$, 95 % CI [−Inf, −0.29] and $r(62) = -0.53$, $p < .001$, 95 % CI [−Inf, −0.35]. As

[1] A so-called 'wrapper analysis' on the data ($N = 329$) of various samples (i.e., healthy controls, inmates, injury litigants, and inpatients) isolated a subset of 21 items, capturing 92 % of the SIMS variance.

[2] By using the corresponding formula as described by Orrù et al. (2022) and taking into account the correlation of.961 between 21- and 75-item SIMS scores in their samples.

[3] For each set of analyses, we applied Bonferroni corrections to adjust alpha levels based on the number of used tests: for group differences ($p = .05/10 = 0.005$); correlations of over- and underreporting indices (per sample; $p = .05/15 = 0.003$); stability tests ($p = .05/8 = 0.006$); and BFI-2-S correlations (per sample; $p = .05/6 = 0.008$).

[4] As suggested by one reviewer, we also ran 2 (Group: patients vs. students) x 2 (Time: T1 vs. T2) ANOVAs on the 21-item SIMS and SS supernormality scores. This yielded a similar pattern of results as the *t*-tests reported here.

Table 1
Converted full-version SIMS scores (i.e., symptom overreporting), SS total scale scores (i.e., symptom underreporting), and IFS scores (i.e., inattentive responding) in forensic inpatients and students at T1 (intake or baseline) and T2 (after a 6–10 week interval).

		T1			T2			Test stat.	Effect size
		M (*SD*)	Range	*n* (%) > cutoff	*M* (*SD*)	Range	*n* (%) > cutoff		
Forensic inpatients	SIMS	9.86 (5.57)	4–24	3 (3.8)	9.08 (5.11)	4–26	2 (3.1)	$t = 0.54$	$d = 0.07$
	SS	15.49 (5.62)	3–30	10 (12.8)	17.16 (5.96)	3–28	15 (23.4)	$t = -2.55$	$d = 0.32$
	IFS	1.35 (1.03)	0–3	13 (16.7)	1.16 (1.29)	0–5	9 (14.1)	$t = 2.03$	$d = 0.25$
Students	SIMS	9.72 (6.35)	4–38	8 (6.3)	10.10 (5.38)	4–29	2 (3.1)	$t = -0.96$	$d = 0.12$
	SS	14.26 (4.99)	3–26	5 (3.9)	14.55 (5.39)	3–26	6 (9.4)	$t = 0.00$	$d = 0.00$
	IFS	0.75 (1.12)	0–7	8 (6.3)	0.58 (0.71)	0–2	0	$t = 0.12$	$d = 0.02$

Notes. Patient T1 data are based on $n = 78$, T2 on $n = 64$. Student T1 data are based on $n = 127$, T2 on $n = 64$. All ps = ns (based on Bonferroni corrections of $p < .05/10 = 0.005$). Prevalence rates of participants exceeding cutoffs are denoted by n (%).

follow-up checks, the analyses were re-run while correcting for IFS (i.e., with partial correlations). The negative associations remained statistically significant, with *r*s ranging between −0.47 and −0.56, all $ps < 0.001$. Using a categorical approach to the data (based on cut-points) essentially yielded similar patterns, see Supplemental File D.

3.1.3. *Temporal stability*[5]

In forensic inpatients, the 21-item SIMS scores across T1 and T2 were correlated, $r(62) = 0.63$, $p < .001$, 95 % CI [0.47, Inf] and did not statistically significantly change over time, $t(63) = -0.54$, $p = .30$, $d = 0.07$. SS supernormality scores of patients also correlated across time points, $r(62) = 0.83$, $p < .001$, 95 % CI [0.75, Inf], but contrary to expectations, statistically significantly increased over time, $t(63) = -3.13$, $p = .001$, $d = 0.39$. Finally, IFS scores of the patients at T1 and T2 were correlated, $r(62) = 0.43$, $p < .001$, 95 % CI [0.25, Inf], and remained stable over time after applying Bonferroni correction[3], $t(63) = 2.03$, $p = .02$, $d = 0.25$. In students, the 21-item SIMS, SS supernormality, and IFS scores were correlated across time points, $rs(62) = 0.62$, 0.84, and 0.35, $ps \leq 0.002$, respectively, and were stable over time, $t(63) = -0.96$, $p = .17$, $d = 0.12$; $t(63) = -1.00$, $p = .16$, $d = 0.13$; and $t(63) = 0.12$, $p = .45$, $d = 0.02$. Table 1 shows scale scores across T1 and T2. Table 2 gives an overview of correlations at each time point.

Table 2
Correlations of the 21-item SIMS scores, SS supernormality subscale scores, and IFS scores in forensic inpatients and students at T1 and T2.

	T1 SIMS	T1 SS	T1 IFS	T2 SIMS	T2 SS	T2 IFS
Forensic inpatients						
T1 SIMS		−0.50*	−0.13	0.63*	−0.42*	0.02
T1 SS			0.24	−0.47*	0.83*	0.05
T1 IFS				−0.03	0.16	0.43*
T2 SIMS					−0.53*	0.02
T2 SS						0.06
T2 IFS						
Students						
T1 SIMS		−0.42*	−0.04	0.62*	−0.44*	0.03
T1 SS			−0.11	−0.51*	0.84*	0.09
T1 IFS				0.02	0.03	0.35*
T2 SIMS					−0.53*	0.19
T2 SS						0.19
T2 IFS						

Notes. Patient T1 data are based on $n = 78$, T2 data on $n = 64$. Student T1 data are based on $n = 127$, T2 data on $n = 64$. *$p < .003$ (based on Bonferroni corrections of $p < .05/15 = 0.003$). SIMS refers to the 21-item SIMS version, SS to the SS supernormality subscale. Correlations are one-tailed (confirmatory), except IFS-SIMS and IFS-SS correlations (exploratory). Correlations involving SIMS data are Spearman-ranked, all other correlations Pearson's *r*.

3.1.4. *Correlations with BFI-2-S scales of interest*

As expected, 21-item SIMS scores and BFI-2-S 'Negative Emotionality' were statistically significantly associated at T1 and T2, both in patients: $r(76) = 0.44$, $p < .001$, 95 % CI [0.27, Inf] and $r(62) = 0.45$, $p < .001$, 95 % CI [0.26, Inf], respectively, and in students: $r(125) = 0.41$, $p < .001$, 95 % CI [0.28, Inf] and $r(62) = 0.41$, $p < .001$, 95 % CI [0.21, Inf], respectively. SS supernormality scores and BFI-2-S 'Conscientiousness' were statistically significantly correlated in the patients at both time points: $r(76) = 0.56$, $p < .001$, 95 % CI [0.41, Inf], and $r(62) = 0.55$, $p < .001$, 95 % CI [0.39, Inf], but not in students after applying Bonferroni correction[3]: $r(125) = 0.21$, $p = .009$, 95 % CI [0.06, Inf], and $r(62) = 0.25$, $p = .02$, 95 % CI [0.05, Inf]. Finally, against expectations, IFS and BFI-2-S 'Open-mindedness' were not statistically significantly linked at T1 or T2. Neither in the patients, $r(76) = 0.13$, $p = .13$, 95 % CI [−0.06, Inf] and $r(62) = -0.02$, $p = .44$, 95 % CI [−0.23, Inf], respectively; nor in students, $r(125) = 0.04$, $p = .33$, 95 % CI [−0.11, Inf] and $r(62) = 0.11$, $p = .21$, 95 % CI [−0.10, Inf], respectively. Partial correlations controlling for IFS yielded similar patterns.

3.2. *Exploratory analyses*

3.2.1. *Links with personality diagnoses*

We tested whether forensic inpatients who were diagnosed with a DSM-5 'borderline personality disorder' (at T1: 6 of 36 patients; at T2: 5 of 36 patients) would score higher on the 21-item SIMS than patients without this diagnosis. Both at T1 and T2 such association emerged, $z(34) = 2.51$, $p = .01$, $r = 0.42$, and $z(34) = 2.16$, $p = .03$, $r = 0.36$, respectively. Also, we explored whether a DSM-5 'narcissistic personality disorder' diagnosis (at T1 and T2: 5 patients) was linked to higher SS supernormality scores. This was the case, at both T1, $t(34) = 2.73$, $p = .01$, $d = 1.31$, and T2, $t(34) = 2.28$, $p = .03$, $d = 1.10$.

4. Discussion

The literature on trait correlates of distorted symptom expression is mostly cross-sectional (Van Helvoort et al., 2022). Notable exceptions are An et al. (2012) and DeRight and Jorgensen (2015), who showed stability of cognitive underperformance in students, suggesting a potential – yet uninvestigated – trait contribution to symptom overreporting. In our test-retest study in forensic inpatients and undergraduate students, we examined such links among their symptom overreporting and underreporting (i.e., exaggeration/fabrication vs. minimization/denial of psychiatric symptoms) tendencies, inattentive responding, and traits.

4.1. *Group comparisons*

Our main findings were as follows. Contradicting our first

[5] We also computed intra-class correlations (i.e., two-way mixed; absolute agreement; single measures), which yielded similar results to the Pearson's *r*s reported here.

hypothesis, forensic inpatients and students showed comparable rates of symptom overreporting (±5 %) and symptom underreporting (patients ±15–20 %; students ±5–10 %). These over- and underreporting rates parallel those previously found in Dutch forensic outpatients (i.e., 9 % and 23 %, respectively; Van Impelen et al., 2016). Overall, the similarities – even though patients likely faced relatively stronger (and at T2 increasing) incentives – support the possibility that trait factors and symptom distortion may share a conceptual space (Merckelbach et al., 2019). At a minimum, our data show that, despite differences in background, context and incentives, both groups showed a remarkable similarity in their manifestations of over- and underreporting. However, this interpretation must remain tentative given the low base rate of SIMS scores >23.

In contrast to our first hypothesis, inattentive responding was nontrivial in patients (±15 %), but low in students (±0–5 %). Inattentive responding did not obscure the inverse relationship between symptom over- and underreporting observed in both samples. These findings support our second hypothesis and, more broadly, the *bipolarity assumption* – that symptom overreporting and underreporting represent opposite poles of a single behavioral continuum (e.g., Rogers, 2008). While this assumption has been explored in the context of virtues and personal qualities, it has received little empirical scrutiny within the symptom domain. Yet, the broader literature provides indications of mixed response styles in applied settings: among job applicants (e.g., Melchers et al., 2020), in disability evaluations (Whitman et al., 2023), and in instructed feigning paradigms where blended symptom over- and underreporting mimicked validity profiles of honest respondents (Bošković et al., 2024). Thus, the extent to which over- and underreporting act as true opposites may depend on context and clinical characteristics (e.g., like the "good-old-days-bias" in litigants with mild head injury; Iverson et al., 2010). Our results imply that over- and underreporting do not easily go together, but the cited studies highlight the theoretical and clinical need to further investigate the circumstances under which symptom over- and underreporting do co-occur.

4.2. *Stability and trait correlates*

Our test-retest data largely support our third hypothesis, showing that symptom distortions are relatively stable across time. Although this does not confirm trait involvement, it is consistent with such involvement (see also An et al., 2012). The one exception was that underreporting increased among forensic patients near their first treatment evaluation – suggesting that imminent incentives may outweigh dispositional tendencies. This aligns with documented studies showing that contextual shifts can drastically alter the accuracy of symptom self-reports (e.g., Walters, 1988). Taken together, the findings suggest that distorted symptom expression can be both stable and volatile, depending on the context.

Trait correlations were only partly consistent with our fourth hypothesis: symptom overreporting was consistently linked to 'negative emotionality', while symptom underreporting was associated with 'conscientiousness' in patients but not students and inattentive responding was unrelated to 'open-mindedness'. Over- and underreporters differed ±0.75 SD on their scale scores in the direction of prima facie negative and positive traits, respectively, compared to those who did not exceed the relevant cutoffs (see Supplemental File E). Although modest, the correlations make a patterned rather than arbitrary impression. This might reflect response style distortion, but also leaves room for the possibility that certain traits and response styles share an underlying conceptual domain. If self-reported trait scores were mere response artefacts, one would expect an amorphous constellation of undifferentiated correlations across all trait scales.

Meanwhile, in a subsample of patients whose diagnostic data were available (n = 36) provide support for a specific relationship between traits and symptom distortions: borderline personality disorder was linked to overreporting, while narcissistic personality disorder was related to underreporting. Admittedly, the subsample size was small, but even so, the differences were considerable: e.g., 86.4 % of those with a narcissistic personality disorder scored higher on the SS supernormality scale than the mean score of those without this diagnosis (Cohen's d = 1.10). In sum, then, the interpretation that traits (e.g., affective instability; self-enhancement motives) might co-occur with, or predispose to, particular response styles is not refuted by our data. Overall, our findings are consistent with the view that the correlational patterns reflect shared or bi-directional underpinnings (see also Van Helvoort et al., 2022).

5. Limitations

First, we did not assess (or manipulate) perception of incentives. Although we assumed forensic patients faced relatively stronger incentives – they were explicitly told that their responses might influence treatment decisions (e.g. treatment intensity; ward placement) – patients likely varied in their perceptions, which may also have shifted over time. In fact, at T2 the imminent treatment evaluation may have increased underreporting in the patient group. If this interpretation is correct, it would be a contextual effect that weakens trait-like stability.

Second, in part we used brief scales (i.e., the 21-item SIMS and BFI-2-S) to minimize inattentive or fatigued responding in forensic inpatients. However, and likely due to their short lengths (Schmitt, 1996), their internal consistency coefficients fell below expectations (αs = 0.53–0.68 for the 21-item SIMS; α < 0.60 in students for some BFI-2-S subscales), which may limit the inferences that we can draw from the data. Still, one could argue here that when measures have modest reliability, estimates of relationships will be correspondingly attenuated, which would mean that our correlations are underestimations (Schmitt, 1996). Moreover, test-retest reliability of the IFS was low, but may be theoretically appropriate given that inattentive responding – by nature – fluctuates over time. In sum, we believe that while the measurement precision was imperfect for some instruments, the consistency and structure of the observed patterns still offer meaningful – though tentative – insights into the constructs of interest.

Third, while our patient sample suffered from a wide range of serious psychopathology, generalizability to other patient groups (e.g., with severe cognitive symptoms) remains unclear. Fourth, carry-over effects cannot be ruled out, as prior assessment experience varied across patients. Fifth, the scales were administered electronically, although studies suggest minimal differences between digital and traditional (validity) testing (e.g., Giromini et al., 2024). Finally, given that we used the stringent SS cutoff (i.e., ≥22), sensitivity might be low. Thus, the prevalence rates of symptom underreporting may be an underestimation.

6. Conclusions

Our findings suggest that when it comes to symptom distortion, individual differences matter, even under circumstances that impose considerable incentives to distort symptoms. Clinically, it seems prudent to develop guidelines for identifying and discussing with patients the internal and external factors that might drive their symptom distortion, and that foster a joint narrative or case conceptualization. This might help avoid one-sided, stigmatizing interpretations, such as the misassumption that symptom distortion in high incentive contexts is primarily indicative of faking bad or good (e.g., see Merckelbach et al., 2019). An approach which takes trait correlates into account may also reduce risks. Preliminary evidence suggests that failing to adequately address symptom distortions may result in misdiagnoses, drop-out, harmful interventions, and large financial costs (Merckelbach & Dandachi-FitzGerald, 2025). More research on the determinants and consequences of symptom distortions is needed to do justice to the precise role of individual differences.

CRediT authorship contribution statement

Daniël van Helvoort: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Henry Otgaar:** Writing – review & editing, Investigation, Formal analysis. **Chijs van Nieuwenhuizen:** Writing – review & editing, Resources, Methodology. **Harald Merckelbach:** Writing – review & editing, Supervision, Methodology, Formal analysis, 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.2025.113551.

Data availability

Student data can be found at https://osf.io/j9mvf. Patient data are confidential and available only upon reasonable request.

References

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*(8), 849–857. doi:10/1093/arclin/acs085.

Berry, K., Rana, R., Lockwood, A., Fletcher, L., & Pratt, D. (2019). Factors associated with inattentive responding in online survey research. *Personality and Individual Differences, 149*, 157–159. https://doi.org/10.1016/j.paid.2019.05.043

Bošković, I., Giromini, L., Katsouri, A., Tsvetanova, E., Fonse, J., & Merckelbach, H. (2024). The spectrum of response bias in trauma reports: Overreporting, underreporting, and mixed presentation. *Psychological Injury and Law, 17*, 117–128. https://doi.org/10.1007/s12207-024-09503-w

Chapman, L. J., Chapman, J. P., & Miller, E. N. (1982). Reliabilities and intercorrelations of eight measures of proneness to psychosis. *Journal of Consulting and Clinical Psychology, 50*(2), 187–195.

Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., & Muris, P. (2003). The other side of malingering: Supernormality. *The Clinical Neuropsychologist, 17*(2), 235–243. https://doi.org/10.1076/clin.17.2.235.16507

De Vries, R. E., Zettler, I., & Hilbig, B. E. (2014). Rethinking trait conceptions of social desirability scales: Impression management as an expression of honesty-humility. *Assessment, 21*(3), 286–299.

DeRight, J., & Jorgensen, R. S. (2015). I just want my research credit: Frequency of suboptimal effort in a non-clinical healthy undergraduate sample. *The Clinical Neuropsychologist, 29*(1), 101–117. https://doi.org/10.1080/13854046.2014.989267

Duberstein, P. R., & Heisel, M. J. (2007). Personality traits and the reporting of affective disorder symptoms in depressed patients. *Journal of Affective Disorders, 103*(1–3), 165–171. https://doi.org/10.1016/j.jad.2007.01.025

Garnefski, N., & Kraaij, V. (2006). Cognitive emotion regulation questionnaire – Development of a short 18-item version (CERQ-short). *Personality and Individual Differences, 41*(6), 1045–1053. https://doi.org/10.1016/j.paid.2006.04.010

Giromini, L., Pignolo, C., Zennaro, A., & Sellbom, M. (2024). Using the MMPI-2-RF, IOP-29, IOP-M, and FIT in the in-person and remote administration formats: A simulation study on feigned mTBI. *Assessment, 31*(8), 1626–1642. https://doi.org/10.1177/10731911241235465

Holden, R. R., & Passey, J. (2010). Socially desirable responding in personality assessment: Not necessarily faking and not necessarily substance. *Personality and Individual Differences, 49*(5), 446–450. https://doi.org/10.1016/j.paid.2010.04.015

Iverson, G. L., Lange, R. T., Brooks, B. L., & Lynn Ashton Rennison, V. (2010). "Good old days" bias following mild traumatic brain injury. *The Clinical Neuropsychologist, 24*(1), 17–37. https://doi.org/10.1080/13854040903190797

Kraaij, V., & Garnefski, N. (2019). The Behavioral Emotion Regulation Questionnaire: Development, psychometric properties and relationships with emotional problems and the Cognitive Emotion Regulation Questionnaire. *Personality and Individual Differences, 137*, 56–61. https://doi.org/10.1016/j.paid.2018.07.036

Lee, P., Joo, S.-H., & Fyffe, S. (2019). Investigating faking effects on the construct validity through the Monte Carlo simulation study. *Personality and Individual Differences, 150*, Article 109491. https://doi.org/10.1016/j.paid.2019.07.001

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

Melchers, K. G., Roulin, N., & Buehl, A. K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment, 28*(2), 123–142. https://doi.org/10.1111/ijsa.12280

Merckelbach, H., & Dandachi-FitzGerald, B. (2025). Symptom overreporting and its consequences for treatment. *Current Opinion in Psychology, 65*, Article 102091. https://doi.org/10.1016/j.copsyc.2025.102091

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science, 28*(3), 321–326. https://doi.org/10.1177/0963721419837681

Novo, R., Gonzalez, B., & Roberto, M. (2022). Beyond personality: Underreporting in high-stakes assessment contexts. *Personality and Individual Differences, 184*, Article 111190. https://doi.org/10.1016/j.paid.2021.111190

Orrù, G., De Marchi, B., Sartori, G., Gemignani, A., Scarpazza, C., Monaro, M., & Roma, P. (2022). Machine learning item selection for short scale construction: A proof-of-concept using the SIMS. *The Clinical Neuropsychologist, 37*(7), 1371–1388. https://doi.org/10.1080/13854046.2022.2114548

Rogers, R. (2008). An introduction to response styles. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (3rd ed., pp. 3–13). The Guildford Press.

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment, 8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured Inventory of Malingered Symptomatology: A psychometric review. *Psychological Injury and Law, 15*(1), 64–78. https://doi.org/10.1007/s12207-021-09432-y

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Van Helvoort, D., Merckelbach, H., van Nieuwenhuizen, C., & Otgaar, H. (2022). Traits and distorted symptom presentation: A scoping review. *Psychological Injury and Law, 15*(2), 151–171. https://doi.org/10.1007/s12207-022-09446-0

Van Impelen, A., Merckelbach, H., Niesten, I., Jelicic, M., Huhnt, B., & Campo, J.Á. (2016). Biased symptom reporting and antisocial behaviour in forensic samples: A weak link. *Psychiatry, Psychology, and Law, 24*(4), 530–548. https://doi.org/10.1080/13218719.2016.1256017

Walters, G. D. (1988). Assessing dissimulation and denial on the MMPI in a sample of maximum security, male inmates. *Journal of Personality Assessment, 52*(3), 465–474. https://doi.org/10.1207/s15327752jpa5203_8

Whitman, M. R., Gervais, R. O., & Ben-Porath, Y. S. (2023). Virtuous victims: Disability claimants who over-and under-report. *The Clinical Neuropsychologist, 38*(7), 1584–1607. https://doi.org/10.1080/13854046.2023.2185686