



The European Journal of Psychology Applied to Legal Context

<https://journals.copmadrid.org/ejpalc>



Measuring and Priming Underreporting Using the General Inventory of Behaviours, Symptoms, and Opinions (GIBSO)

Irena Bošković^{1,4}, Harald Merckelbach², Francesca Bosco³, and Paolo Roma³

¹Erasmus University Rotterdam, Rotterdam, the Netherlands; ²Maastricht University, Maastricht, the Netherlands; ³Sapienza University of Rome, Rome, Italy; ⁴University of Novi Sad, Serbia

ARTICLE INFO

Article history:

Received 13 February 2026

Accepted 24 August 2026

Keywords:

GIBSO

Underreporting

Faking good

Symptom validity assessment

ABSTRACT

Background: Underreporting impacts the self-reports' validity across various domains (e.g., symptom and risk assessments). We evaluated the effectiveness of the General Inventory of Behaviours, Symptoms, and Opinions (GIBSO) in identifying underreporting. **Method:** Participants ($N = 305$), who participated in English or in Dutch, were told to respond honestly. Some participants (non-prime group, $n = 76$) received no further information, whereas others were primed to believe we would assess their ability to be care-takers (caretaker prime, $n = 75$), risk of endangering others (safety prime, $n = 77$), or developing substance-abuse issues (addiction prime, $n = 77$). All participants completed the GIBSO as well as mental and physical symptom inventories and questionnaires about drug use and aggression. **Results:** Groups had comparable scores on all inventories, indicating no significant impact of primes on participants' responses. Both the GIBSO Total and General Underreporting (GU) subscale correlated significantly and negatively with other measures, and 28% of participants exceeded the GIBSO cutoff score (> 20). GIBSO Total and GU scale were equivalent between the international and Dutch participants, whereas Conservatism scale appeared more culturally sensitive. **Conclusions:** Given its negative correlations with symptom measures, the GIBSO, and particularly its GU subscale, holds promise as a screening tool for underreporting, although further systematic research is needed.

Medición e inducción a la disimulación en el General Inventory of Behaviours, Symptoms, and Opinions (GIBSO)

RESUMEN

Antecedentes: La disimulación afecta la validez de los autoinformes en diversos dominios (p. ej., en la evaluación de síntomas y de riesgos). Diseñamos unos estudios con el objetivo de evaluar la eficacia del General Inventory of Behaviours, Symptoms, and Opinions (GIBSO) en la clasificación de la disimulación. **Método:** Se solicitó a los participantes ($N = 305$), que respondieron en inglés o en neerlandés, que contestaran con honestidad. Un grupo de participantes (grupo sin inducción, $n = 76$) no recibieron más información, mientras que a otros grupos se les indujo a creer (*primed*) que evaluaríamos su capacidad para ser cuidadores (inducción de cuidador, $n = 75$), su riesgo de poner en peligro a otros (inducción de seguridad, $n = 77$) o de desarrollar problemas de abuso de sustancias (inducción de adicción, $n = 77$). Todos los participantes completaron el GIBSO, así como inventarios de síntomas físicos y mentales y cuestionarios sobre el uso de drogas y violencia. **Resultados:** Los grupos obtuvieron puntuaciones comparables en todos los inventarios, lo que indica que las inducciones no tuvieron un impacto significativo en las respuestas de los participantes. Tanto la Escala Total del GIBSO como la Subescala de Disimulación General (GU) correlacionaron de forma significativa y negativa con las demás medidas, y el 28% de los participantes superó la puntuación de corte de clasificación de la disimulación del GIBSO (> 20). Los resultados en la Escala Total del GIBSO y la escala GU fueron similares entre los participantes internacionales y neerlandeses, mientras que la Escala de Conservadurismo pareció ser más sensible culturalmente. **Conclusiones:** Dadas sus correlaciones negativas con las medidas de síntomas, el GIBSO, y en particular su subescala GU, se perfila como una herramienta prometedora para la detección de la disimulación, aunque se necesita más investigación sistemática.

People often distort their reports by exhibiting some type of biased responding. For instance, they can portray their health as worse than it actually is (i.e., exhibiting “negative” response bias), or, in contrast, individuals can exhibit “positive” response bias by presenting themselves in a better light. Positive response bias mostly occurs as denial of symptoms, traits or behaviours that might be seen as unfavourable

(Paulhus, 2002; see also Rogers, 2018). Such self-presentation, when captured only in test scores, should be labelled as “underreporting”, because this term does not require having specific information about the motives behind such behaviour (Bošković & Rowlands, 2025). When motives are known, other terms might be more applicable, such as faking good, suggesting intentional, externally driven, dis-

Cite this article as: Bošković, I., Merckelbach, H., Bosco, F., & Roma, P. (2026). Measuring and priming underreporting using the General Inventory of Behaviours, Symptoms, and Opinions (GIBSO). *European Journal of Psychology Applied to Legal Context*, 18, Article e260183. <https://doi.org/10.5093/ejpalc2026a8>

Correspondence: irena.boškovic@ff.uns.ac.rs; Bošković@essb.eur.nl (I. Bošković).

ISSN: 1889-1861/© 2026 Colegio Oficial de la Psicología de Madrid. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

tortion of self-report. However, a plethora of motives can be driving underreporting, ranging from unintentional (e.g., obliging the social norms) to intentional ones (e.g., employment). For instance, research in occupational settings shows that individuals often withhold or minimise symptoms because they fear negative job-related consequences such as loss of opportunities, reduced trust from supervisors, or stigma from colleagues (Bogaers et al., 2021). Further, many different situations in the legal arena, such as parole, risk, and custody hearings that involve clear interests, might be motivating for a claimant, defendant or convict to exhibit underreporting (see Merckelbach & Bošković, 2025).

Although a limited number of studies directly compared the prevalence of underreporting to overreporting behaviour, their findings indicate that underreporting in health reports might be more “prevalent” than its opposite – symptom overreporting. For instance, in an older study involving forensic patients ($N = 159$), higher percentages scored above the cutoff of MMPI indices of underreporting than above the cutoff of overreporting indices: 21% versus 9% (Heilbrun et al., 1990). Looking at Covid reports ($N = 465$), more than 25% of responders stated knowing someone who fabricated having the infection, whereas more than 50% knew someone who concealed having it (Bošković et al., 2024). In another study by Bošković et al. (2025), participants ($N = 215$) were asked to report about their history of faking behaviour. While 33% admitted symptom overreporting in the past, 48% reported having engaged in underreporting. Additionally, participants also exhibited a stronger propensity to underreport than to overreport, as well as higher confidence in their success of convincing others.

Both symptom under- and overreporting may have a considerable “impact”. For example, whereas overreporting may lead to unnecessary and potentially harmful treatments (Merckelbach & Dandachi-FitzGerald, 2025; van der Heide et al., 2020), underreporting was shown to significantly lower the odds of seeking (needed) treatment (Le Forestier et al., 2024; Pachankis et al., 2020), consequently impacting the safety not only of the evaluatee but their environment as well (Jackson & Harrison, 2018; see also Weissman & Gorlin, 2023). Furthermore, a non-trivial proportion of research participants (12%, Bošković et al., 2024) spontaneously engage in underreporting, even when explicitly told to respond honestly, directly lowering the validity of research findings. Therefore, effective assessment and detection of symptom underreporting is important. Detecting this behaviour, however, is difficult. The content people choose to conceal often depends on the surrounding context, prevailing social norms, and interpretation. Prior studies show that individuals frequently minimize or completely conceal psychological (Bril-Barniv et al., 2016; Martin, 2010) and physical complaints (Quinn et al., 2017), problematic behaviours such as substance use (Clark et al., 2016; Steinhoff et al., 2023), aggression (Arnetz et al., 2015; see also Lobbestael, 2015), and sensitive aspects of sexual behaviour (Bošković et al., 2023; Pachankis et al., 2020). Even among those actively engaged in therapy, about a third concealed the severity of their symptoms, thoughts about suicide, and drug or alcohol use (Farber, 2020). Yet, the available measures of underreporting often are narrow in their focus, dominantly capturing only biased presentation of personality characteristics (Paulino et al., 2024) or symptoms (e.g., Supernormality Scale; Cima et al., 2003), and are often embedded as subscales in broader measures (e.g., MMPI-2-FR; for an overview, see Picard et al., 2023).

In order to address the wider range of domains that are subject to underreporting, the General Inventory of Behaviours, Symptoms, and Opinions (GIBSO) was recently developed (Bošković & Merckelbach, 2024). The GIBSO includes 48 items, of which four serve as an initial appraisal that the test protocol is valid and ready for interpretation. The remaining 44 items cover domains of psychological (depression, anxiety, trauma-related) symptoms, common physical issues (backpain, stomach pain, fatigue), antisocial behaviour (drug use and aggression), as well as domains of sexual fantasies and disclosure of

one's issues. In prior research, these items were shown to fall under two factors: General Underreporting (GU, 33 items), directly measuring symptom and behaviour concealment, and Conservatism (C, 11 items), mostly tapping into perceived stigma and help-seeking attitude. The C scale in itself is not crucial for detecting underreporting but is instead important for understanding the factors that drive it. This is especially relevant when assessing victims of (sexual) violence, who have been shown to be prone to underreporting (Bošković et al., 2023). The proposed cutoffs for underreporting are > 20 for the GIBSO full scale and > 17 for the GU subscale. In preliminary research (Bošković & Rowlands, 2025), these cutoffs were associated with a sensitivity and specificity of .79 and .80, respectively, thereby providing some preliminary evidence for the effectiveness of the GIBSO in signalling symptom underreporting.

One serious limitation of these initial results is that they were obtained in a simulation study involving psychology students explicitly instructed to underreport. Having a sample with a high level of pre-existing knowledge about mental health that was explicitly instructed to manipulate their symptom reports restricts generalizability to realistic settings, such as court hearings (Bianchini et al., 2001; Mirza et al., 2020). In the majority of real-life assessments, a person will not receive direct instructions (e.g., from their attorney; but see Spengler et al., 2020). More commonly, individuals are exposed to subtle prompts, such as implicit expectations or guidance from legal counsel about how their self-presentation should ideally look like. Therefore, we lack research that includes more realistic conditions in which participants have been prompted in various subtle ways (i.e., ecologically valid designs), that would help us examine the utility of available measures.

Current Study

In this study, we explored whether the GIBSO can detect underreporting when individuals are simply exposed to minimal, indirect prompts, while also including corresponding clinical scales to account for the broad psychological, somatic, and behavioural domains the instrument gauges. All participants were told to respond honestly, but a randomly subset also received subtle cues suggesting that their answers would be used to evaluate personal qualities such as parenting ability or risk of harmful behaviour. We included four experimental conditions (non-prime, caretaker prime, safety prime, and addiction prime) and administered questionnaires that tap into mental (depression, anxiety, and stress) and physical (somatic) health, and into antisocial behaviour (drug use and aggression). This design allowed us to assess whether the GIBSO is sensitive to both spontaneous and subtly prompted underreporting. We then compared group scores on the GIBSO and clinical scales, anticipating that prime groups would exhibit significantly higher GIBSO scores than unprimed participants. Furthermore, underreporting tendencies were expected to manifest as negative correlations between the GIBSO and these scales. As a subsidiary aim, as our participants could complete the study in English or in Dutch, we examined whether the English and Dutch versions of the GIBSO generated comparable results.

Method

Participants

Initially, 392 student participants and members of the general public joined the study, but records of participants who 1) did not provide informed consent ($n = 4$), or permission to use their data ($n = 11$), 2) provided incomplete reports ($n = 55$), and/or 3) failed the attention checks ($n = 17$, see Materials and Measures section) were excluded. The final sample consisted of 305 participants, of whom the major-

rity were women ($n = 227$, 74.4%) in their mid-twenties ($M = 26.17$, $SD = 11.91$, range = 18–72). The sample was collected primarily in the Netherlands but also across different countries. The most frequently reported nationalities were Dutch ($n = 208$, 66%), German ($n = 23$, 8%), Spanish ($n = 17$, 6%), Turkish, Polish, and Indian ($< 3\%$). International participants responded in English ($n = 196$), and rated their English proficiency as relatively high ($M = 3.94$, $SD = .76$, range = 2–5). The rest of the sample participated in their native language, Dutch ($n = 109$). All participants rated their overall physical and mental health using a 5-point scale (higher values being indicative of better self-reported health), and their ratings were indicative of moderately good health ($M_{\text{physical}} = 3.68$, $SD_{\text{physical}} = .91$; $M_{\text{mental}} = 3.36$, $SD_{\text{mental}} = .93$; range = 1–5).

Participants were randomly allocated to one of four conditions: non-prime ($n = 76$), caretaker prime ($n = 75$), safety prime ($n = 77$), and addiction prime ($n = 77$). Groups did not differ in terms of age, nor in (mental and physical) health ratings, $F_s < .83$, $p_s > .48$, $\eta^2 < .008$, nor in education, $H(3) = 4.62$, $p = .202$.

Materials and Measures

Questionnaire order was randomized. Overall, the measures were selected to tap into psychological (DASS-21) and physical (Adapted PHQ-15) health, antisocial behaviour (BPAQ and DAST-10), and into underreporting tendencies (GIBSO). Attention checks were embedded in all measures except the GIBSO (which contains built-in validity checks) and were tailored to match each scale's response format (e.g., instructing participants to select *a little* or *sometimes at specific places*). Depending on their preference, participants either received English or Dutch versions of the measures.

Primes

All participants were told to respond honestly. However, participants allocated to the prime groups received additional prompts. Below are the instructions that each group received (the text in italics was identical for each group):

Non-prime control:

Dear participant,

In a moment, we will present a couple of measures that tap into various behaviours, symptoms of physical and mental health, and opinions. Please read each item carefully and respond in accordance with the instructions to each measure. Please note that we inserted a few items that will check whether you are paying attention and if you fail those, your data will be considered invalid and will not be used. So, please carefully read the items and respond to them accordingly. Because the goal of our study is to inspect how common the measured concepts are, we need you to respond to all items as honestly as possible.

Caretaker prompt:

... Prior to filling out our measures, we need to inform you that some of the measures used in this study were designed to assess how capable you would be as a primary caretaker of a child (i.e., we will assess your parenting abilities).

Safety prompt:

... Prior to filling out our measures, we need to inform you that some of the measures used in the study were designed to assess how likely it would be for you to endanger the safety of yourself or others in the future (i.e., your aggressive tendencies).

Addiction prompt:

... Prior to filling out our measures, we need to inform you that some of the measures used in the study were designed to assess how likely it would be for you to develop an addiction problem in the future (i.e., substance abuse).

Depression, Anxiety, and Stress Scale-21 (DASS-21; Lovibond & Lovibond, 1995)

The DASS-21 was derived from the original, longer version ($n_{\text{items}} = 42$). Its 21 items comprise depression, anxiety, and stress scales. The response format is a 4-point scale (0 = *does not apply to me at all*, 3 = *applies to me very much or most of the time*), and the total score per subscale is then multiplied by two, ranging from 0 to 42 for each subscale, with higher scores being indicative of more complaints. Two attention checks were included and six participants failed them. The internal consistency was good for both the English and Dutch versions, with Cronbach's alphas being .90 and .91, respectively.

Level 2 Somatic Symptom-Adult Patient, Patient Health Questionnaire Physical Symptoms (Adapted PHQ-15; Kroenke et al., 2002)

The origin of this measure is the PHQ-15, but the adapted version covers a variety of physical complaints that are rated on a 3-point response scale (0, 1, or 2). Using this scale, respondents indicate how much they suffer from each complaint. The total score ranges from zero to 30, with higher scores signalling more self-reported somatic symptoms. Based on the attention check, one participant was removed. The internal consistency of this scale was sufficient, with Cronbach's alphas of .80 for the English and .82 for the Dutch version.

Drug-Abuse Screening Test (DAST-10; Skinner, 1982)

The DAST-10 is a self-report measure of drug use in the last 12 months (Skinner, 1982). The response format is "Yes" or "No". Yes answers are summed to obtain a severity index, ranging from 0 to 10 (Bohn et al., 1991). Two participants failed the attention check embedded in this measure. The scale's internal consistency was modest, with Cronbach's alphas being .69 for the English and .68 for the Dutch version. Lower internal consistency values were likely a reflection of both the wide range of substances covered by the measure and the restricted score variance in our sample, given that participants infrequently endorsed DAST-10 items.

Buss-Perry Aggression Questionnaire (BPAQ; Buss & Perry, 1992)

The BPAQ is a 29-item measure of aggressive behaviour. Respondents indicate how characteristic each item is for them using a 5-point Likert scale (1 = *extremely uncharacteristic*, 5 = *extremely characteristic*). The total score ranges from 29 to 145, with higher scores being indicative of stronger self-reported aggression tendencies. We embedded two attention checks and three participants failed them. The internal consistency of the measure was acceptable in both languages, with Cronbach's alpha being .84 for the English and .83 for the Dutch version.

General Inventory of Behaviour, Opinions, and Symptoms (GIBSO; Bošković & Merckelbach, 2024)

The GIBSO is a measure of underreporting, including 48 items to which respondents indicate their dis/agreement using "Yes", "Not

Sure”, and “No” options, which are rated as 1, 0.5, and 0, respectively. Two items serve as attention checks and two as item understanding checks. The total score varies from 0 to 44, with higher scores being suggestive of stronger underreporting tendencies. The proposed cutoff scores are > 20 for the total scale, and > 17 on the General Underreporting (GU) subscale (Bošković & Rowlands, 2025). Based on the attention checks, five participants were excluded from the dataset. The overall internal consistency was satisfactory (Cronbach’s alphas for the English and Dutch version being .75 and .80, respectively).

Motivation, Clarity, and Difficulty Checks

At the end of the study, prior to debriefing, participants rated four aspects of the procedure using a 5-point Likert scale ranging from 1 (*not at all*) to 5 (*a great deal*): (a) motivation to participate, (b) clarity of instructions, (c) clarity of items/questions, and (d) difficulty of the task.

Procedure

The study was conducted online, using Qualtrics, during the period from March until July 2025. Non-students were recruited through snowballing globally, while students were recruited by a student research platform at our institution. After they had followed a link, participants were asked to provide informed consent, respond to demographic questions, and rate their health. Participants were then randomly allocated to one of four conditions and they received the corresponding instructions. Following this, they were administered the questionnaires in a randomized order. Afterward, participants were asked to evaluate the study and their participation. Finally, they were debriefed and asked to provide permission to use their responses. Participants were either rewarded with research credits (when students) or were entered in a raffle for a 10-euro prize. The study was approved by the standing ethics committee of Erasmus University Rotterdam, the Netherlands.

Data Analyses

The data were analysed using the SPSS statical package (version 30.0). We first ran descriptive analyses. Then, we examined potential group differences on all measures using Multivariate Analyses of Variance (MANOVAs) and Bonferroni post-hoc pairwise comparisons. Eta squared (η^2) and Cohen’s *d* were used as measures of effect size, and interpreted according to Cohen’s (1988) guidelines ($\eta^2 \approx .01$ small, .06 medium, $\geq .14$ large; $d \approx 0.20$ small, 0.50 medium, 0.80 large). Pearson’s *r* coefficients were employed to compute correlations between inventories. Beyond correlations, the influence of demographics was evaluated using Welch’s *t*-tests, and differences between the English and Dutch GIBSO versions were assessed using independent-samples *t*-tests. We also conducted exploratory ROC analyses evaluating the diagnostic utility of the GIBSO cutoff score using low-scoring participants across most symptom inventories as a criterion group. However, as these analyses were purely exploratory, they were excluded from this manuscript to maintain focus (see Supplemental file). These results, along with the raw data and analytic outputs, are available on the Open Science Framework (OSF) at: <https://osf.io/sg7pq/overview>.

Results

Participants’ Experience

Participants rated their motivation, clarity of the instructions and questions, and difficulty of their task using a 5-point scale (range from one to five). On average, participants’ ($N = 305$) motivation was moderate ($M = 3.55$, $SD = 0.87$), and clarity of the instructions ($M = 4.48$, $SD = 0.66$) and items ($M = 4.31$, $SD = 0.78$) was relatively high, while the task was perceived as easy ($M = 1.52$, $SD = 1.00$). There were no significant differences between four groups with regard to any of these ratings, $F(4, 296) = 0.44$, $p = .95$, $\eta^2 = .006$.

Table 1. Means and Standard Deviations for Each Group and the Univariate Results on All Measures

		Prime					Univariate Results		
		Total	None	Care-taker	Safety	Addiction			
		305	76	75	77	77	$F(3, 301)$	p	η^2
		$M(SD)$	$M(SD)$	$M(SD)$	$M(SD)$	$M(SD)$			
Symptom inventories									
DASS-21	Depression (0 – 42)	7.25 (6.94)	6.26 (6.20)	6.91 (6.06)	7.81 (7.12)	8.02 (8.12)	1.07	.36	.011
	Anxiety (0 – 42)	6.79 (6.51)	7.16 (5.85)	6.45 (5.81)	6.60 (7.21)	6.96 (7.12)	0.18	.91	.002
	Stress (0 – 42)	12.25 (7.76)	11.55 (8.03)	12.14 (7.47)	12.96 (7.99)	12.36 (7.64)	0.43	.74	.004
PHQ-15 (0 -30)		7.75 (4.73)	7.61 (4.68)	7.76 (3.76)	7.68 (5.40)	7.96 (5.01)	0.08	.97	.001
Drug use and aggression measures									
DAST-10 (0 – 10)		2.02 (1.55)	1.75 (1.31)	2.04 (1.72)	1.99 (1.33)	2.29 (1.74)	1.57	.20	.015
BPAQ (0 – 145)		66.25 (13.47)	67.67 (13.25)	66.37 (14.42)	65.06 (13.57)	65.92 (12.73)	0.49	.69	.005
Underreporting (GIBSO)									
Total Score (0 – 44)		17.46 (4.95)	18.17 (4.61)	18.01 (3.77)	17.22 (4.12)	16.27 (5.29)	1.85	.14	.048
GU (0 – 33)		14.85 (4.67)	14.89 (4.09)	15.01 (3.91)	14.80 (4.25)	13.54 (4.74)	1.26	.29	.019
C (0 – 11)		2.62 (1.80)	3.28 (2.25)	3.00 (1.76)	2.43 (1.63)	2.73 (1.88)	1.83	.14	.028
Cutoff		Exceeding the Cutoff % (n)					$\chi^2(df = 3)$		p
Total > 20		28% (85)	31.8% (27)	17% (20)	23.5% (20)	24.7% (21)	3.38		.34
GU > 17		29% (91)	27.5% (25)	18.7% (17)	24.2% (22)	29.7% (27)	3.25		.36

Note. GU = general underreporting; C = conservatism.

Table 2. Correlations (Pearson's *r*) between GIBSO Total Score and Subscales with Symptom Inventories (DASS-21 & PHQ-15), Drug Use and Aggression Measures (DAST-10 & BPAQ), Health Ratings, and Participants' Age

N = 305		GIBSO Total	95%CI [LL-UL]	GU	95%CI [LL-UL]	C	95%CI [LL-UL]
DASS-21	Depression	-.40	[-.49, -.30]	-.49	[-.57, -.40]	.16	[-.05, -.26]
	Anxiety	-.46	[-.54, -.36]	-.51	[-.59, -.42]	.07	[-.04, .17]
	Stress	-.52	[-.60, -.43]	-.56	[-.63, -.48]	.03	[-.08, .14]
Somatic health (PHQ-15)		-.46	[-.54, -.37]	-.52	[-.60, -.44]	.08	[-.03, .19]
Drug use (DAST-10)		-.38	[-.47, -.28]	-.37	[-.46, -.26]	-.09	[-.20, .02]
Aggression (BPAQ)		-.35	[-.45, -.26]	-.51	[-.60, -.43]	.36	[-.24, .44]
Mental Health (0-5)		.31	[-.20, .40]	.35	[-.25, .44]	-.06	[-.17, .05]
Physical Health (0-5)		.43	[-.33, .52]	.51	[-.42, .59]	-.14	[-.24, .02]
Participants' age (18-72)		.22	[-.11, .32]	.23	[-.12, .34]	.004	[-.12, .11]

Note. GU = general underreporting; C = conservatism; values in bold are significant with $p_s < .001$.

Prime Impact on Symptom and Drug Use and Aggression Inventories

Overall, participants obtained relatively low scores on the Depression and Anxiety subscales of the DASS-21 and moderate scores on its Stress subscale. Participants also self-reported low levels of somatic complaints on the PHQ-15. The groups did not significantly differ with regard to DASS-21 or PHQ-15 ratings, $\lambda = .97$, $F(12, 788.72) = 0.72$, $p = .73$, $\eta^2 = .01$. Participants reported relatively low use of substances on DAST-10, but moderate level of aggression on BPAQ. Again, groups did not significantly differ with respect to the ratings on these two tests, $\lambda = .97$, $F(6, 598) = 1.20$, $p = .304$, $\eta^2 = .012$, suggesting that primes did not promote underreporting. For descriptives and univariate results see Table 1.

Prime Impact on GIBSO and its Subscales

Whether or not participants had been exposed to subtle primes did not affect their GIBSO scores; this was true for both full scale scores and the subscale scores, $\lambda = .98$, $F(6, 600) = 1.03$, $p = .403$, $\eta^2 = .010$. Univariate tests confirmed that there were no significant differences between groups (see Table 1). When applying the GIBSO cutoff score of > 20 , 28% of participants exhibited underreporting. A similar rate was found when applying the General Underreporting cutoff point of > 17 (29%). There were no significant differences in these prevalences across groups.

Correlations between GIBSO and Symptom and Behaviour Inventories

Considering the lack of group differences, we collapsed the groups and, using the total sample ($N = 305$), calculated Pearson product-moment correlations of GIBSO full and subscale scores, symptom inventories, self-reported drug use and aggression, and self-reported health (see Table 2).

Both the Total and the GU GIBSO scores correlated significantly and negatively with scores on other inventories, and positively with self-reported health ratings (see Table 2). The C subscale showed only one significant (medium and positive) correlation, with self-reported aggression. The observed correlations were replicated when the analyses were conducted for the English and Dutch subsamples, separately (see Supplemental Table 1).

Participants' Demographics and GIBSO Scores

We examined potential demographic influences by correlating GIBSO scores with participants' age (see Table 2), and evaluating gender differences. Participants' age was modestly, positively associated

with GIBSO Total and GU scores, whereas its correlation with the C scale was not significant. These findings show that older participants endorsed more items on GIBSO. Furthermore, gender differences were evaluated using Welch's *t*-tests due to unequal group sizes ($n_{\text{women}} = 227$, $n_{\text{men}} = 72$). Although men scored descriptively higher across all scales, these differences were not statistically significant for the GU, $t(106.70) = 1.45$, $p = .15$, Cohen's $d = .21$, or C scales, $t(117.47) = 1.53$, $p = .13$, Cohen's $d = 0.21$. A similar pattern emerged for the Total score, yielding a marginal *p* value and a small-to-medium effect size, $t(107.24) = 1.93$, $p = .06$, Cohen's $d = 0.28$. These findings suggest that older men in our sample might have shown slightly stronger tendency towards underreporting, which warrants further investigation.

Equivalence of English and Dutch GIBSO Versions

We examined whether GIBSO scores differed between the two language groups (English vs. Dutch). The international sample and the Dutch sample obtained similar GIBSO Total, $t(303) = .19$, $p = .85$, Cohen's $d = 0.02$, and GU scores, $t(303) = 1.37$, $p = .17$, Cohen's $d = 0.17$; see Table 3. The international participants did, however, report more resistance to help-seeking and disclosure than the Dutch participants on the C subscale, $t(303) = 3.27$, $p = .001$, Cohen's $d = 0.39$.

Table 3. Mean Scores (and Standard Deviations) on English and Dutch Versions of GIBSO (Total, General Underreporting - GU), and Conservatism - C)

Language Groups	Gender	n	GIBSO		
			Total	GU	C
			<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
English	Total	196	17.42 (4.50)	14.56 (4.27)	2.86 (1.90)
Dutch	Total	109	17.54 (5.68)	15.37 (5.30)	2.17 (1.50)

Given the different group sizes, we repeated the analyses using the Mann-Whitney *U* test, and the results did not change. It is also worth noting that the Dutch participants ($M = 29.80$, $SD = 14.24$, range: 18-72) were significantly older than those in the international sample ($M = 24.15$, $SD = 9.87$, range: 18-63; $t(303) = 3.67$, $p < .001$, Cohen's $d = .48$). Therefore, we repeated the analyses using GIBSO scores while controlling for age, but this did not change the pattern, with the difference on the C scale remaining the only significant one ($p < .001$).

Discussion

The main results can be summarized as follows. First, our attempts to promote underreporting by providing participants with subtle prompts failed. Of course, one could argue that our prompts were too weak and that had we used more explicit instructions, we would have observed higher underreporting levels. However, we refrained from using straightforward instructions to underreport precisely because

they may produce trivial results. Rather, we were interested in forms of underreporting that are more intrinsically tied to the person and less artificially imposed.

Alternatively, the lack of prime effects might have to do with our sample. That is, the baseline level of underreporting tendencies (i.e., close to 30%), as indexed by GIBSO scores above previously established cutoff points, surpassed the rates found in a previously studied control group (13%, Bošković et al., 2024; 20%, Bošković & Rowlands, 2025). However, prior studies relied on a different design, and potential differences in baseline tendencies towards symptom minimisation should be taken into account. Still, such high rates of underreporting in our sample might have resulted in a “ceiling effect” (Cohen, 1988), thereby limiting detection of further suppression of disclosure. Arguments in favour of this interpretation can be found in participants’ self-reported symptoms, drug use, and aggression, which diverge from statistics that one would normally expect.

For example, the relatively low levels of depression and anxiety, average levels of stress, and minimal somatic complaints in our sample deviate from the heightened levels of mental (Campbell et al., 2022; Duffy et al., 2019; Storrie et al., 2010) and somatic health problems (e.g., Elnegaard et al., 2015; Rasmussen et al., 2020) researchers observed in general population and student samples. Additionally, the exceptionally low levels of drug use and the moderate level of aggression contradict global rising trends in drug use (Ignaszewski, 2021) and aggression (Iennaco et al., 2024). Admittedly, these behaviours vary by country, socioeconomic factors (Yu & Chen, 2025), and, particularly in the case of aggression, by the specific type of behaviour that is surveyed. Specifically, research shows that physical aggression has decreased over time, while cyber aggression has increased (Van der Laan et al., 2023). With these considerations in mind, it is reasonable to assume that underreporting tendencies were already relatively pronounced in our sample, thereby likely constraining the influence of subtle primes. It should also be emphasized that underreporting per se does not necessarily denote intentional deception but may instead reflect heightened social desirability and the normative societal adjustment of our participants. Thus, it is also possible that the observed prevalence rates might indicate issues with GIBSO’s validity and its utility in differentiating individuals’ good adjustment from underreporting behaviour.

Second, an argument in favour of GIBSO’s validity is provided by the results of the correlation analyses. GIBSO was significantly and negatively associated with scores on symptom and antisocial behaviour inventories, and moderately and positively correlated with participants’ self-reported mental and physical health ratings. This pattern supports the idea that the GIBSO taps into underreporting tendencies, and replicates previously published data on the relationship between GIBSO and symptom measures (Bošković & Rowlands, 2025). Together these findings suggest that GIBSO could have broad utility as a screening tool for underreporting across various domains, as it appears to flag a general tendency toward minimization of psychological and somatic issues.

Third, particularly interesting is our finding of a positive correlation between Conservatism subscale and aggression score, suggesting that this scale captures a broader defensive interpersonal style. Masuda and Boone (2011) showed that self-concealment and reluctance to seek help are associated with heightened distress, interpersonal difficulties, and maladaptive coping. These factors can manifest in, and relate to, elevated aggression scores on the BPAQ, especially considering items that tap into interpersonal functioning, such as “When frustrated I let my irritation show” and “I often find myself disagreeing with people”. Further, both help-seeking attitudes and aggression are strongly influenced by cultural norms (Altweck et al., 2015).

Fourth, we found no differences in the GIBSO Total and General Underreporting scores between the international and Dutch participants. However, Dutch participants did exhibit significantly lower levels on the Conservatism subscale than the international partici-

pants. Previous research suggests that Dutch individuals tend to be more open about mental health and face less stigma related to treatment than individuals from other cultures (Hoshmand et al., 2024; Kotera et al., 2020). This outcome is unsurprising because our sample was recruited globally rather than being restricted to non-Dutch residents residing within the Netherlands. However, the possibility that the linguistic equivalence of items related to help-seeking attitudes was poor requires further consideration. Thus, although the pattern of GIBSO scores demonstrates overall cross-language similarity, the Conservatism subscale may be sensitive to cultural differences, calling for further cross-cultural research. Taken together with the previous correlation results, these findings suggest that the C scale is better understood as a contextual metric that can aid evaluators in interpreting a respondent’s cultural background and interpersonal style.

Fifth, beyond cultural factors, our findings indicate that participants’ demographic characteristics may affect GIBSO scores. Specifically, the positive correlations between the underreporting scales (Total and GU) and participants’ age indicate that older individuals demonstrated a stronger tendency toward underreporting on the GIBSO. This pattern aligns with existing literature demonstrating that socially desirable responding in mental health self-reports is positively correlated with age (Hitchcott et al., 2020). Alternatively, rather than indicating systematic concealment, these correlations may reflect a base-rate artifact: older participants naturally exhibit lower baseline rates of behavioural and clinical issues (e.g., lower impulsivity or substance use), resulting in lower symptom endorsement that co-varies with GIBSO scores. Regarding gender, men exhibited a slightly increased tendency toward underreporting on the GIBSO, a finding that is also supported by previous literature (see Wagner & Reifegerste, 2024). Based on these combined trends, it remains unclear whether the elevated scores among older men reflect a genuine demographic propensity toward underreporting, or GIBSO scales’ sensitivity to baseline reporting styles of this specific cohort. Future studies should investigate this demographic intersection to determine if specific normative adjustments are warranted.

Strengths, Limitations, and Future Directions

A methodological strength of this study lies in our attempt to employ more ecologically valid instructions rather than the usual simulation scenarios. However, our subtle manipulation may have been “too” subtle, as the primed and non-primed groups exhibited similar scores across all measures. Two plausible explanations exist for this outcome. First, participants may have engaged in spontaneous underreporting regardless of their assigned condition, a notion consistent with the elevated minimization tendencies already observed in the sample. Alternatively, the lack of group differences may reflect limitations in the GIBSO’s sensitivity to subtle shifts in underreporting, rather than an ineffective prime. To resolve this ambiguity, future studies should consider experimental manipulations that are highly self-relevant and evoke a stronger, explicit motivation to underreport.

A second strength is the composition of our sample; while students comprised the majority, we also included members of the general public to enhance generalisability. Nonetheless, this heterogeneous composition may have introduced systematic noise. Previous research indicates that students are typically more inclined to disclose health issues (e.g., 13% in Bošković et al., 2024; 20% in Bošković & Rowlands, 2025). Consequently, the higher underreporting rates observed in the current sample might reflect the inclusion of these general population participants. Future validation efforts would benefit from testing samples composed exclusively of general-public participants to isolate these baseline reporting styles.

Furthermore, because the GIBSO is one of only a few scales developed specifically to capture underreporting, empirical investi-

gations into its psychometric boundaries are exceptionally valuable. Although the significant negative correlations with symptom measures replicate previous findings and testify to the construct validity of the GIBSO, this is the first study to compare different language versions of the instrument. While we observed no mean differences between the English and Dutch versions, this absence of significant differences does not in itself demonstrate measurement equivalence. Because our study was not statistically powered or designed for such analyses, formal tests of measurement invariance (e.g., Putnick & Bornstein, 2018; Wang et al., 2021) remain an essential next step across different linguistic, cultural, and demographic groups. Importantly, language mastery also varied within our sample: English-language respondents differed in their proficiency, whereas all Dutch-language respondents were native speakers. Further research should evaluate language equivalence using groups that include both native and non-native speakers of the same target language. This is a critical consideration for researchers implementing other active language versions of the GIBSO beyond those tested here, such as the Italian, French, German, Spanish, and Serbian translations.

Ultimately, future work should employ designs that allow for direct comparison between the GIBSO and other well-established instruments. By incorporating additional underreporting measures or external behavioural criteria, future research can establish the GIBSO's incremental utility when administered alongside established validity scales, such as those embedded in the MMPI-2-RF (Paulhus, 2002; Picard et al., 2023). Still, our study provides provisional data on the English and Dutch versions of the GIBSO across non-clinical populations. This is particularly important given the limited number of tools available to practitioners screening for underreporting.

Conclusion

Our subtle priming instructions were too weak to evoke significantly different response styles among participants. Still, the negative associations between GIBSO scores and self-reported symptoms, drug use, and aggression underscore that the GIBSO captures a general tendency toward minimization across multiple domains. The results of this study offer preliminary but encouraging evidence for GIBSO's utility, particularly for its shorter (GU) version, which may prove to be more robust across different cultural contexts. However, further systematic investigation remains necessary to establish the instrument's psychometric precision while simultaneously clarifying the conceptual nature of underreporting across diverse cultural and demographic populations. Finally, it is important to note that the GIBSO is not designed to be the sole source of information for determining the presence of a positive response bias.

Conflict of Interest

The authors of this article declare no conflict of interest.

Acknowledgements

We thank our master students who collected data for this project: Durga Manju Rajeev, Floor van Dalen, Teresa Ojeda Bosch, Teresa Ojeda Bosch, and Lucas Vogelsang.

Data Availability

The data and the outputs are available on the Open Science Framework platform (<https://osf.io/sq7pq/overview>).

Authors' Contribution

Irena Bošković: Conceptualisation, Data collection, Data analyses, and Writing and Editing. Harald Merckelbach: Data analyses, Writing, and Editing. Francesca Bosco: Data collection, Writing, and Editing. Paolo Roma: Writing and Editing.

References

- Altweck, L., Marshall, T. C., Ferenczi, N., & Lefringhausen, K. (2015). Mental health literacy: A cross-cultural approach to knowledge and beliefs about depression, schizophrenia and generalized anxiety disorder. *Frontiers in Psychology*, 6, Article 1272. <https://doi.org/10.3389/fpsyg.2015.01272>
- Arnetz, J. E., Hamblin, L., Ager, J., Luborsky, M., Upfal, M. J., Russell, J., & Essenmacher, L. (2015). Underreporting of workplace violence: Comparison of self-report and actual documentation of hospital incidents. *Workplace Health & Safety*, 63(5), 200-210. <https://doi.org/10.1177/2165079915574684>
- Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15(1), 19-45. <https://doi.org/10.1076/clin.15.1.19.1907>
- Bogaers, R., Geuze, E., Van Weeghel, J., Leijten, F., Rüsch, N., van de Mheen, D., Varis, P., Rozema, A., & Brouwers, E. (2021). Decision (not) to disclose mental health conditions or substance abuse in the work environment: A multiperspective focus group study within the military. *BMJ Open*, 11(10), Article e049370. <https://doi.org/10.1136/bmjopen-2021-049370>
- Bohn, M., Babor, T., & Kranzler, H. (1991). Validity of the Drug Abuse Screening Test (DAST-10) in inpatient substance abusers. In Harris (Ed.), *Problems of drug dependence* (pp. 233-235). U.S. Department of health and human services.
- Bošković, I., Giromini, L., Akca, A. Y. E., Mazza, C., & Roma, P. (2025). Faking bad, faking good and the dark tetrad: Relationship between spontaneous faking, history of faking behaviour, propensity to fake and dark tetrad. *International Journal of Forensic Mental Health*, 24(2), 155-166. <https://doi.org/10.1177/14999013251326582>
- Bošković, I., & Merckelbach, H. (2024). *The General Inventory of Behaviour, Symptoms, and Opinions (GIBSO)*. Unpublished manual. <https://doi.org/10.1037/t94834-000>
- Bošković, I., Orthey, R., Otgaar, H., Mangiulli, I., & Rassin, E. (2023). # StudentsToo: prevalence of sexual assault reports among students of three European universities and their actions post-assault. *PLoS One*, 18(4), Article e0283554. <https://doi.org/10.1371/journal.pone.0283554>
- Bošković, I., Puente-López, E., Dandachi-FitzGerald, B., & Merckelbach, H. (2024). The prevalence of feigning and concealment of Covid-19 infections in an international sample. *Journal of Health Psychology*, 29(6), 595-607. <https://doi.org/10.1177/13591053231226033>
- Bošković, I., & Rowlands, M. T. (2025). New measure of underreporting on horizon: Initial psychometric evaluation of the General Inventory of Behaviour, Symptoms, and Opinions (GIBSO). *Psychological Injury and Law*, 18(1), 84-96. <https://doi.org/10.1007/s12207-024-09527-2>
- Bril-Barniv, S., Moran, G. S., Naaman, A., Roe, D., & Karnieli-Miller, O. (2016). A qualitative study examining experiences and dilemmas in concealment and disclosure of people living with serious mental illness. *Qualitative Health Research*, 27(4), 573-583. <https://doi.org/10.1177/1049732316673581>
- Buss, A. H., & Perry, M. (1992). The aggression questionnaire. *Journal of Personality and Social Psychology*, 63(3), 452-459. <https://doi.org/10.1037/0022-3514.63.3.452>
- Campbell, F., Blank, L., Cantrell, A., Baxter, S., Blackmore, C., Dixon, J., & Goyder, E. (2022). Factors that influence mental health of university and college students in the UK: A systematic review. *BMC Public Health*, 22(1), Article 1778. <https://doi.org/10.1186/s12889-022-13943-x>
- Cima, M., Merckelbach, H. L. G. J., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., & Muris, P. E. H. M. (2003). The other side of malingering: Supernormality. *The Clinical Neuropsychologist*, 17(2), 235-243. <https://doi.org/10.1076/clin.17.2.235.16507>
- Clark, C. B., Zyambo, C. M., Li, Y., & Cropsey, K. L. (2016). The impact of non-concordant self-report of substance use in clinical trials research. *Addictive Behaviors*, 58, 74-79. <https://doi.org/10.1016/j.addbeh.2016.02.023>
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Duffy, A., Saunders, K. E., Malhi, G. S., Patten, S., Cipriani, A., McNevin, S. H., MacDonald, E., & Geddes, J. (2019). Mental health care for university students: A future forward? *The Lancet Psychiatry*, 6(11), 885-887. [https://doi.org/10.1016/S2215-0366\(19\)30275-5](https://doi.org/10.1016/S2215-0366(19)30275-5)
- Elnegaard, S., Andersen, R. S., Pedersen, A. F., Veldt-Larsen, P., Sondergaard, J., Rasmussen, S., Balasubramaniam, K., Svendsen, R. P., Vedsted, P., & Jarbol, D. E. (2015). Self-reported symptoms and healthcare seeking in the general population -exploring "The Symptom Iceberg". *BMC Public Health*, 15, 685-699. <https://doi.org/10.1186/s12889-015-2034-5>

- Farber, B. A. (2020). Disclosure, concealment, and dishonesty in psychotherapy: A clinically focused review. *Journal of Clinical Psychology*, 76(2), 251–257. <https://doi.org/10.1002/jclp.22891>
- Heilbrun, K., Bennett, W. S., & White, A. J. (1990). An MMPI-based model of malingering and deception. *Behavioral Sciences & Law*, 8(1), 45–53. <https://doi.org/10.1002/bsl.2370080106>
- Hitchcott, P. K., Penna, M. P., & Fastame, M. C. (2020). Age trends in well-being and depressive symptoms: The role of social desirability. *Psychiatric Quarterly*, 91(2), 463–473. <https://doi.org/10.1007/s11126-020-09711-y>
- Hoshmand, S., Knehtering, H., & Spoelstra, S. K. (2024). Cultural competence of mental health practitioners in The Netherlands. *International Journal of Social Psychiatry*, 70(2), 282–288. <https://doi.org/10.1177/00207640231206062>
- Iennaco, J. D., Molle, E., Lauvud, C., Palmstierna, T., Nijman, H., & Almvik, R. (2024). Measurement and prevalence of aggression. In N. Hallett, R. Whittington, D. Richter, & E. Eneje (Eds.) *Coercion and violence in mental health settings* (pp. 93–123). Springer. https://doi.org/10.1007/978-3-031-61224-4_5
- Ignaszewski, M. J. (2021). The epidemiology of drug abuse. *The Journal of Clinical Pharmacology*, 61(S2), S10–S17. <https://doi.org/10.1002/jcph.1937>
- Jackson, R. L., & Harrison, K. S. (2018). Assessment of law enforcement personnel: The role of response styles. In R. Rogers & D. S. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 552–570). Guilford Publications.
- Kotera, Y., Van Laethem, M., & Ohshima, R. (2020). Cross-cultural comparison of mental health between Japanese and Dutch workers: Relationships with mental health shame, self-compassion, work engagement and motivation. *Cross Cultural & Strategic Management*, 27(3), 511–530. <https://doi.org/10.1108/CCSM-02-2020-0055>
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2002). The PHQ-15: Validity of a new measure for evaluating the severity of somatic symptoms. *Psychosomatic Medicine*, 64(2), 258–266. <https://doi.org/10.1097/00006842-200203000-00008>
- Le Forestier, J. M., Chan, E. W., Shephard, R., Page-Gould, E., & Chasteen, A. L. (2024). Why is concealment associated with health and wellbeing? An investigation of potential mechanisms. *Social Science & Medicine*, 344, Article 116529. <https://doi.org/10.1016/j.socscimed.2023.116529>
- Lobbetael, J. (2015). Challenges in aggression assessment: The gap between self-report and behavior, and a call for new valid behavioral paradigms. *Journal of Socialomics*, 5(1), 141–152. <https://doi.org/10.4172/2167-0358.1000141>
- Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, 33(3), 335–343. [https://doi.org/10.1016/0005-7967\(94\)-00075-U](https://doi.org/10.1016/0005-7967(94)-00075-U)
- Martin, J. M. (2010). Stigma and student mental health in higher education. *Higher Education Research & Development*, 29(3), 259–274. <https://doi.org/10.1080/07294360903470969>
- Masuda, A., Boone, M. S. (2011). Mental health stigma, self-concealment, and help-seeking attitudes among Asian American and European American college students with no help-seeking experience. *International Journal for the Advancement of Counselling*, 33(4), 266–279. <https://doi.org/10.1007/s10447-011-9129-1>
- Merckelbach, H., & Bošković, I. (2025). Symptoms? Which symptoms? Distorted symptom presentation and its detection. In Bošković, I. (Ed.), *Legal and forensic psychology: What is it and what it is not* (pp. 93–109). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-75875-1_7
- 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>
- Mirza, N., Cinel, J., Noyes, H., McKenzie, W., Burgess, K., Blackstock, S., & Sanderson, D. (2020). Simulated patient scenario development: A methodological review of validity and reliability reporting. *Nurse Education Today*, 85, Article 104222. <https://doi.org/10.1016/j.nedt.2019.104222>
- Pachankis, J. E., Mahon, C. P., Jackson, S. D., Fetzner, B. K., & Bränström, R. (2020). Sexual orientation concealment and mental health: A conceptual and meta-analytic review. *Psychological Bulletin*, 146(10), 831–861. <https://doi.org/10.1037/bul0000271>
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 49–69). Erlbaum.
- Paulino, M., Edens, J. F., Moniz, M., Moura, O., Rijo, D., & Simões, M. R. (2024). Personality assessment inventory (PAI) in forensic and correctional settings: A comprehensive review. *Journal of Forensic and Legal Medicine*, 103, Article 102661. <https://doi.org/10.1016/j.jflm.2024.102661>
- Picard, E., Aparcero, M., Nijdam-Jones, A., & Rosenfeld, B. (2023). Identifying positive impression management using the MMPI-2 and the MMPI-2-RF: A meta-analysis. *The Clinical Neuropsychologist*, 37(3), 545–561. <https://doi.org/10.1080/13854046.2022.2077237>
- Quinn, D. M., Weisz, B. M., & Lawner, E. K. (2017). Impact of active concealment of stigmatized identities on physical and psychological quality of life. *Social Science & Medicine*, 192(C), 14–17. <https://doi.org/10.1016/j.socscimed.2017.09.024>
- Rasmussen, S., Jensen, C. T., Rosendal, M., Vægter, H. B., Søndergaard, J., & Jarbøl, D. E. (2020). Multiple physical symptoms and individual characteristics—A cross-sectional study of the general population. *Journal of Psychosomatic Research*, 131, Article 109941. <https://doi.org/10.1016/j.jpsychores.2020.109941>
- Rogers, R. (2018). An introduction to response style. In R. Rogers & D. S. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 3–17). Guilford Publications.
- Skinner, H. A. (1982). *Guide for using the drug abuse screening test (DAST)*. Centre for Addiction and Mental Health. <https://doi.org/10.1037/t03979-000>
- Spengler, P. M., Walters, N. T., Bryan, E., & Millspaugh, B. S. (2020). Attorneys' attitudes toward coaching forensic clients on the MMPI-2: Replication and extension of attorney survey by Wetter and Corrigan (1995). *Journal of Personality Assessment*, 102(1), 56–65. <https://doi.org/10.1080/00223891.2018.1501568>
- Steinhoff, A., Shanahan, L., Bechtiger, L., Zimmermann, J., Ribeaud, D., Eisner, M. P., Baumgartner, M. R., & Quednow, B. B. (2023). When substance use is underreported: comparing self-reports and hair toxicology in an urban cohort of young adults. *Journal of the American Academy of Child & Adolescent Psychiatry*, 62(7), 791–804. <https://doi.org/10.1016/j.jaac.2022.11.011>
- Storrie, K., Ahern, K., & Tuckett, A. (2010). A systematic review: Students with mental health problems—a growing problem. *International Journal of Nursing Practice*, 16(1), 1–6. <https://doi.org/10.1111/j.1440-172X.2009.01813.x>
- Wagner, A., & Reifegerste, D. (2024). Gender-specific non-disclosure of mental distress: A cross-sectional survey on reasons and associated factors. *European Journal of Health Communication*, 5(2), 60–80. <https://doi.org/10.47368/ejhc.2024.204>
- Weissman, S., & Gorlin, E. I. (2023). Exploring the (Mal) adaptive Consequences of Self-Deceptive Enhancement: A Narrative Review. *Journal of Theoretical Social Psychology*, 2023, Article 2382005. <https://doi.org/10.1155/2023/2382005>
- Van der Heide, D., Bošković, I., van Harten, P., & Merckelbach, H. (2020). Overlooking feigning behavior may result in potential harmful treatment interventions: Two case reports of undetected malingering. *Journal of Forensic Sciences*, 65(4), 1371–1375. <https://doi.org/10.1111/1556-4029.14320>
- Van der Laan, C. M., van de Weijer, S. G., Nivard, M. G., & Boomsma, D. I. (2023). Familial clustering of trends in aggression. *Journal of Quantitative Criminology*, 39(4), 1–19. <https://doi.org/10.1007/s10940-021-09523-8>
- Yu, C., & Chen, J. (2025). Global Burden of substance use disorders among adolescents during 1990–2021 and a forecast for 2022–2030: An analysis for the Global Burden of Disease 2021. *BMC Public Health*, 25, Article 1012. <https://doi.org/10.1186/s12889-025-22107-6>

Appendix

Supplementary Materials

Table S1. Correlations (Pearson's r) between GIBSO Total Score and Subscales with Symptom Inventories (DASS-21 & PHQ-15), Behaviour Measures (DAST-10 & BPAQ), and Health Ratings in English ($n = 196$) and Dutch Subsamples ($n = 109$)

		GIBSO Total		General Underreporting		Conservatism	
		Eng.	Dutch	Eng.	Dutch	Eng.	Dutch
DASS-21	Stress	-.49	-.57	-.54	-.62	.04	.06
	Anxiety	-.43	-.52	-.49	-.62	.09	-.05
	Depression	-.36	-.50	-.47	-.54	.20	-.004
Somatic Symptoms (PHQ-15)		-.41	-.54	-.48	-.57	.11	.02
Drug Use (DAST-10)		-.40	-.37	-.39	-.36	-.06	-.11
Aggression (BPAQ)		-.32	-.40	-.49	-.52	.33	.34
Mental Health (0-5)		.38	.27	.47	.28	-.15	.03
Physical Health (0-5)		.34	.52	.39	.56	-.07	-.03

Note. Values in bold are significant with $p_s < .005$.

Exploratory Analysis: Diagnostic Accuracy of GIBSO

We additionally examined the usefulness of the GIBSO for identifying participants with consistently low (i.e., 1 *SD* below means) score on both symptom, drug use, and aggression measures (i.e., Low scorers). Because only one participant scored low on all six measures, we gradually relaxed the criteria until we obtained a subgroup of reasonable size. When the criterion required participants to have at least four low scores (out of the possible six), 12 participants met this threshold.

We included Low scorers ($n = 12$) as a criterion group whereas the remaining participants ($n = 293$) served as controls in the Receiver Operating Characteristics (ROC) analysis. Overall, the AUCs were .93 (95%CI [.87–.99]) for the Total score and .95 (95%CI [.91–.99]) for the GU score. The optimal cutoff for the Total score was >20, yielding sensitivity = .92 and specificity = .75 (Youden's index = .66; LR = 3.68). For the GU score, the previously proposed cutoff of >17 produced perfect sensitivity (= 1.00) and acceptable specificity (.73; Youden's index = .73; Likelihood Ratio [LR] = 3.70; see Table S2).

Table S2. Cutoff Scores on GIBSO Total and GIBSO General Underreporting scores in detecting Low Scorers on both Symptom and Behaviour Measures ($n = 12$)

GIBSO Total				General Underreporting			
AUC	.93			AUC	.95		
Cutoff	Sensitivity	Specificity	Youden's index	Cutoff	Sensitivity	Specificity	Youden's index
> 16	1.00	.42	.42	>16	1.00	.63	.63
> 17	1.00	.50	.50	>17	1.00	.73	.73
> 18	1.00	.56	.56	>18	1.00	.82	.82
> 19	.92	.67	.58	>19	.84	.85	.68
> 20	.92	.75	.66	>20	.84	.90	.73
> 21	.92	.82	.73	>21	.67	.95	.61
> 22	.75	.86	.61	>22	.58	.97	.55
> 23	.75	.89	.64	>23	.42	.98	.39
> 24	.67	.94	.61	>24	.34	.99	.40
> 25	.58	.96	.55	>25	.34	1.00	.33
> 26	.42	.97	.39	>26	.17	1.00	.17

Notably, increasing both cutoffs by one point (Total > 21, Youden's index = .73; GU >18, Youden's index = .82) appeared to maintain the sensitivity but improve the specificity rates.

Overview of Exploratory Analyses Findings

We examined the utility of the GIBSO in identifying a subgroup of consistently low scoring participants used here as a noisy proxy for underreporting. The GIBSO performed reasonably well, showing high sensitivity (> .92) and equivalent specificity (> .75) to prior research (Bošković & Rowlands, 2024). Slightly adjusting the cutoff (Total > 21; GU > 18) further increased specificity without affecting sensitivity. However, the lower specificity relative to sensitivity may suggest that the GIBSO requires further refinement to better identify true negatives, or that underreporting, particularly in subtle or unconscious forms, is widespread and difficult to detect. This pattern aligns with findings from other measures, such as the Supernormality Scale (Cima et al., 2003). Further research across diverse samples and forms of underreporting is needed to determine whether reduced specificity reflects a structural limitation of the GIBSO or a broader measurement challenge. Future studies could include follow up interviews or additional assessments of individuals scoring low on symptom measures to more precisely evaluate specificity. Importantly,

Appendix

Supplementary Materials (continued)

despite these encouraging findings, the GIBSO, like any symptom validity test, should not be used in isolation. Our results do show (LR of 3,70) that the results on GIBSO could be useful and meaningful, but far from decisive. Thus, conclusions about report authenticity must rely on corroborative information from multiple measures and sources. Finally, it is necessary to acknowledge the following limitation of this analysis: low symptom scores may simply reflect genuinely low symptomatology rather than active underreporting. Therefore, further investigation into the utility of the GIBSO and this approach using more appropriate samples is required.