

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, Self-Report Symptom Inventory, and Inventory of Problems-29**

**This is the author's manuscript**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1876663> since 2022-10-17T10:42:11Z

*Published version:*

DOI:10.1080/23279095.2022.2057856

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

## Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, Self-Report Symptom Inventory, and Inventory of Problems-29

Irena Boskovic, Ali Yunus Emre Akca & Luciano Giromini

To cite this article: Irena Boskovic, Ali Yunus Emre Akca & Luciano Giromini (2022): Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, Self-Report Symptom Inventory, and Inventory of Problems-29, Applied Neuropsychology: Adult, DOI: [10.1080/23279095.2022.2057856](https://doi.org/10.1080/23279095.2022.2057856)

To link to this article: <https://doi.org/10.1080/23279095.2022.2057856>



© 2022 The Author(s). Published with  
license by Taylor & Francis Group, LLC



View supplementary material [↗](#)



Published online: 13 Apr 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

# Symptom coaching and symptom validity tests: An analog study using the structured inventory of malingered symptomatology, Self-Report Symptom Inventory, and Inventory of Problems-29

Irena Boskovic<sup>a,b</sup> , Ali Yunus Emre Akca<sup>c</sup> , and Luciano Giromini<sup>c</sup> 

<sup>a</sup>Forensic Psychology Section, Clinical Psychology Department, Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, Rotterdam, The Netherlands; <sup>b</sup>Forensic Psychology Section, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands; <sup>c</sup>Department of Psychology, University of Turin, Torino, Italy

## ABSTRACT

In this pilot and exploratory study, we tested the robustness of three self-report symptom validity tests (SVTs) to symptom coaching for depression, with and without additional information available on the Internet. Specifically, we divided our sample ( $N = 193$ ) so that each subject received either the Structured Inventory of Malingered Symptomatology (SIMS;  $n = 64$ ), the Self-Report Symptom Inventory (SRSI;  $n = 66$ ), or the Inventory of Problems-29 (IOP-29;  $n = 63$ ). Within each of the three subgroups, approximately one third of participants were instructed to respond honestly (Genuine Condition,  $n_{SIMS} = 21$ ;  $n_{SRSI} = 24$ ;  $n_{IOP-29} = 26$ ) and approximately two-thirds were instructed to feign depression. One half of the feigners were presented with a vignette to increase their compliance with instructions and were given information about symptoms of depression (Coached Feigning,  $n_{SIMS} = 25$ ;  $n_{SRSI} = 18$ ;  $n_{IOP-29} = 21$ ), and the other half were given the same vignette and information about symptoms of depression, plus two Internet links to review before completing the test (Internet-Coached Feigning,  $n_{SIMS} = 18$ ;  $n_{SRSI} = 24$ ;  $n_{IOP-29} = 16$ ). Overall, the results showed that the genuine conditions yielded the lowest total scores on all three measures, while the two feigning conditions did not significantly differ from each other. Looking at the detection rates for all feigning participants, all three measures showed satisfactory results, with IOP-29 performing slightly better than SIMS and SIMS performing slightly better than SRSI. Internet-Coached Feigners scored slightly lower on all three measures than feigners who were coached without the Internet links. Taken together, the results of this preliminary and exploratory study suggest that all three SVTs examined are sensitive to feigned depression even in the presence of symptom coaching, both with and without additional Internet-based information.



## KEYWORDS


Depression; feigning; IOP; malingering; SIMS; SRSI; symptom validity

Whether physical or mental, feigning illness and/or symptoms occurs at a non-trivial rate in both civil (i.e., personal injury claims) and criminal settings. The *Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition* defines “the intentional production of false or grossly exaggerated physical or psychological symptoms motivated by external incentives” as malingering (DSM-5; American Psychiatric Association, 2013). The “external incentives” are a key part of the definition. This allows professionals (i.e., clinicians and researchers) to distinguish malingering from *factitious disorder*, which is motivated by *internal* incentives (Bass & Wade, 2019; Boskovic, 2019). External incentives related to malingering can be financial (e.g., compensation for disability), legal (e.g., avoidance of criminal prosecution, avoidance of civil obligations such as military service), and receipt of privileges (e.g., additional retakes for failed/missed exams)

(Akca et al., 2020; see also Merten & Merckelbach, 2020; Sherman et al., 2020). However, due to the complexity of verifying incentives (see van Impelen et al., 2017), researchers generally prefer the terms “feigning” and “overreporting” to “malingering” because they do not involve assumptions about intentionality or possible incentives (Rogers, 2018b).

The exact prevalence (i.e., base rate) of feigning is difficult to determine and therefore tends to be based on estimates. One renowned study has shown that feigning was suspected in approximately 7–31% of more than 33,000 annual examinations by practitioners, depending on the setting (i.e., civil, criminal, or medical/psychiatric; Mittenberg et al., 2002; see also Dandachi-FitzGerald et al., 2020). Indeed, factors such as ambiguous definitions of feigning, referral context (e.g., clinical, forensic), criteria for detection, methods, and standardized measures used, among others,

**CONTACT** Irena Boskovic  [boskovic@essb.eur.nl](mailto:boskovic@essb.eur.nl)  Forensic Psychology Section, Clinical Psychology Department, Erasmus School of Social and Behavioral Sciences, Erasmus University Rotterdam, 3062 PA Rotterdam, The Netherlands.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23279095.2022.2057856>.

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

contribute to the variability of the frequency with which feigning occurs, ranging from 3% to 64% (Young, 2015). A review of the literature considering these contributing factors found a likely base rate of feigning to be  $15\% \pm 15\%$  (Young, 2015).

Prominent researchers and associations have emphasized the importance of paying attention to possible feigning and developing enhanced methods for detecting feigning (e.g., Bush et al., 2005; Chafetz et al., 2015; Heilbronner et al., 2009; Slick et al., 1999). Most commonly, the assessment of possible feigning is conducted using both *Symptom Validity Tests* (SVTs) and *Performance Validity Tests* (PVTs; Larrabee, 2012). PVTs essentially assess *underperformance*, following the rationale that feigners, when presented with a cognitive task (typically with a dichotomous response option of which only one is correct), sometimes *deliberately* choose the wrong answer(s) to present themselves as impaired (Fazio et al., 2015; Merten & Merckelbach, 2012, 2013). In contrast, SVTs are used to detect *over-endorsement* of genuine or fabricated symptoms, as well as the presentation of implausible combinations of symptoms. The rationale behind most SVTs is that individuals who are feigning tend to overendorse illness-related complaints and that they may overreport bizarre or extremely unlikely problems (Boskovic et al., 2020). In turn, this also has the potential to generate unlikely patterns of symptom endorsements.

Some SVTs are administered in the form of symptom inventories that examinees must complete on their own, while others take the form of a (semi)structured interview (for an overview, see Rogers, 2018a). The most commonly used stand-alone SVT is probably the *Structured Inventory of Malingered Symptomatology* (SIMS; Dandachi-Fitzgerald et al., 2013; Martin et al., 2015; Smith & Burger, 1997), which consists of 75 true/false items across five domains: Amnesic disorder, psychosis, low intelligence, neurological impairment, and affective disorders. The majority of items refer to extreme, bizarre, or atypical complaints, although some of them describe more common symptoms (Martin et al., 2015). The SIMS has been shown to be a psychometrically sound measure in a variety of samples (for a meta-analysis see van Impelen et al., 2014), and has high consistency across different cultures (Nijdam-Jones & Rosenfeld, 2017; see also Boskovic et al., 2017). However, it has some drawbacks, such as overestimation of feigning in certain psychopathological groups (for a systematic review, see van Impelen, 2014). In addition, it has been argued that the SIMS has a high face validity due to the easily recognizable bizarre quality of some items and also limited applicability in a civil context where people might frequently report symptoms related to pain or anxiety (Merten et al., 2016).

To overcome some of these limitations, a new measure was developed, the *Self-Report Symptom Inventory* (SRSI; Merten et al., 2016). The SRSI includes 107 items, most of which are divided into two main scales, one capturing *genuine* symptoms, and the other capturing *pseudo-complaints*. The bizarre quality of the SIMS items has been replaced with more plausible content, making the SRSI less recognizable as a symptom validity measure. In addition, each of the

main SRSI scales includes five subscales to measure commonly reported psychological and physical complaints, which improves its use in the civil context as well. The SRSI includes items that capture cognitive problems, depression, anxiety/post-traumatic stress disorder (PTSD), nonspecific somatic symptoms, and pain symptoms. To date, research has provided strong empirical support for the use of the SRSI in a variety of cultures and contexts (see Giger & Merten, 2019; Merten et al., 2016; van Helvoort et al., 2019; for a review, see Merten et al., 2021).

Another relatively new stand-alone SVT is the *Inventory of Problems-29* (IOP-29; Viglione et al., 2017). The IOP-29 includes 29 self-administered items that assess the credibility of psychiatric and cognitive symptom presentations. Specifically, the items tap into various symptom presentations associated with conditions such as depression, PTSD, cognitive impairment, schizophrenia, and combinations thereof. Since its introduction, many studies have aimed to test its psychometric properties and cross-cultural validity. The results of these studies have been particularly encouraging (e.g., Banovic et al., 2021; Carvalho et al., 2021; Giromini, Barbosa, et al., 2019; Giromini, Lettieri, et al., 2019; Ilgunaite et al., 2020; Roma et al., 2020; Sömen et al., 2021; Winters et al., 2020; for a quantitative review, see Giromini & Viglione, 2022).

Although each of the SVTs described above is successful in detecting feigning, coaching presents a problem that could threaten their success. *Coaching* implies the use of information about validity testing (e.g., through self-preparation) or the provision of information about validity measures (e.g., by third parties) that could confound the detection of feigning (Rogers & Bender, 2018). In a recent meta-analysis on neurocognitive testing, data showed that coaching about target symptoms was significantly more detrimental to the utility of SVTs than test coaching (i.e., providing information about the SVTs; Crişan et al., 2021). Some of the potential sources of coaching are (a) feedback provided by professionals (e.g., psychologists or psychiatrists) on symptom presentations or (b) attorneys themselves (see Essig et al., 2001), but also (c) the *Internet* (Suhr & Gunstad, 2007). Thanks to the Internet, information about various symptom presentations, clinical criteria for disorders, and physical or mental sequelae of deficits is readily available to anyone. Further, research has shown that the validity of PVT results was compromised by online information. Specifically, results showed that 8–26% of websites visited by researchers posed a threat (i.e., a moderate to high level threat) to the validity of test scores (Bauer & McCaffrey, 2006; see also Ruiz et al., 2002).

Given the impact of coaching on the validity of symptom inventories, it is important to determine the extent to which commonly used SVTs are robust to Internet-facilitated coaching. To our knowledge, no study has yet examined the effects of Internet coaching on SIMS, SRSI, and IOP-29 scores. However, several studies have investigated the psychometric robustness of SIMS, SRSI, and IOP-29 to other types of coaching. For example, Jelcic et al. (2007) tested the utility of SIMS in detecting feigned cognitive decline

between honest, naïve (i.e., uncoached), and coached participants. Results showed that SIMS correctly classified 86% of feigners, who were informed and warned about SVTs, suggesting that SIMS is resistant to coaching (Jelicic et al., 2011). In a recent study that included students either responding honestly or being coached to feign depression or ADHD, the SIMS showed perfect specificity (i.e., no false positive errors), but low to medium sensitivity (36–52%), depending on the cutoff point applied. Because feigning groups in this study received, in addition to disorder-specific symptom information, a warning that one or more tests were designed to detect feigning (Grant et al., 2020), the authors concluded that coaching instructions may result in lower sensitivity (see also Jelicic et al., 2011). Regarding the SRSI and coaching, Boskovic et al. (2019) examined the ability of the SRSI to detect fabricated PTSD-related complaints. In addition to honest participants, the study included a group of college students who were instructed to feign PTSD without any additional information and a group of actors who were required to watch videos of people talking about trauma-related experiences before completing the SRSI. Findings showed that the SRSI performed well, correctly classifying 89% of the coached actors. Finally, the IOP-29 was also tested with respect to the effects of coaching. For instance, Gegner et al. (2021) included honest participants and two groups of participants who were instructed to feign mild traumatic brain injury (TBI), half of whom also received a warning to avoid detection. The IOP-29 was able to detect over 90% of feigning participants, with only a small difference (approximately 2%) in favor of the coached participants. In summary, previous studies have supported the robustness of the SIMS, SRSI, and IOP-29 to coaching. However, no study has used the exact same coaching instructions for all three tests, and no study has focused specifically on Internet coaching.

In light of the above, the current study was designed with the goal of separately examining the SIMS, SRSI, and IOP-29 with respect to their robustness to coaching, using the same simulation paradigm and instructions for all three measures (see below). We chose these stand alone SVTs because they are frequently researched and used in practice (see Giromini et al., 2022). Specifically, we investigated whether symptom coaching and additional Internet coaching (i.e., providing additional Internet links to gain information concerning the disorder) had an impact on SIMS, SRSI, and IOP-29 scores. We included three conditions: one served as a control condition (Genuine condition); two included instructions to fake depression, information about symptoms of depression (i.e., symptom coaching), and a warning “not to overdo it” (see Appendix). In one of the two feigning conditions (Internet-Coached Feigning condition) but not in the other (Coached Feigning condition), participants additionally received Internet links to websites containing information concerning depression. We hypothesized that (i) participants in the Genuine condition would generate the most credible results on all three measures, meaning that they would not endorse pseudosymptoms or endorse them at a much lower rate, compared with the participants in the

two feigning conditions. For the Internet-Coached Feigning condition, we made two mutually exclusive predictions because of the lack of previous research on this topic. That is, we speculated that the (ii) participants in the Internet-Coached Feigning condition could show more credible results on all three measures (i.e., lower endorsement of pseudosymptoms) compared to participants in the Coached Feigning condition. This is because, by using the Internet, participants who were faking depression could better understand which symptoms are characteristic of real depression and thus avoid endorsing pseudosymptoms. Yet, the available literature suggests that searching for health information online increases health anxiety (i.e., cyberchondria), leading to higher endorsement of pathological items in symptom inventories (e.g., Brown et al., 2020; Jungmann et al., 2020; Starcevic & Berle, 2013). We therefore also hypothesized that (iii) the Internet-Coached Feigning might actually lead to higher pseudosymptom scores due to an increased tendency to exaggerate/overreport.

## Method

### Participants

A total of 197 psychology students of the Erasmus University Rotterdam were recruited as participants, with the inclusion criterion being that they were pursuing their studies in English. Four students failed the check for inattentive responding, which we included in all of the used measures (see below). Thus, the final sample consisted of 193 students (88.6% female,  $M_{age} = 20.77$ ,  $SD = 2.97$ ; age range 18–44 years). The vast majority of participants were undergraduate students (99%); only two participants reported being master’s students. On average, participants rated their English proficiency as very good ( $M = 4.25$ ,  $SD = .79$ ). More specifically, the frequency of English proficiency ratings was as follows: 43.5% “native/excellent” (point 5), 42% “good” (point 4), 11% “average” (3), and 3.5% reported their proficiency as “poor” (2). Nobody indicated “terrible” (1) proficiency. Of the total seven students with poor proficiency, one was given SIMS (feigning), one SRSI (feigning), and five received IOP-29 (genuine group). This study was approved by the standing ethical committee of the Erasmus University Rotterdam, the Netherlands.

According to the research design (see below), participants were randomly assigned to one of three groups: SIMS ( $n = 64$ ); SRSI ( $n = 66$ ), or IOP-29 ( $n = 63$ ). The random allocation of participants was performed automatically by the research platform used to conduct the study, i.e., Qualtrics. Within these groups, each person was randomly assigned to receive either Genuine, Coached Feigning, or Internet-Coached Feigning instructions (see Table 1 for more details). Overall, participants’ ages did not differ between the three SVT groups,  $F(2, 192) = .82$ ,  $p = .440$ , nor did participants differ in their self-reported English proficiency,  $F(2, 192) = 1.24$ ,  $p = .293$ . However, the mental health ratings of participants who received different measures differed significantly,  $F(2, 192) = 5.15$ ,  $p = .002$ . Specifically, participants in both the SIMS and SRSI groups

**Table 1.** Age, English proficiency, and mental health check by group and condition.

Group	Condition	N	Age M (SD)	English proficiency M (SD)	Mental health check M (SD)
SIMS	Genuine	21	20.29 (2.41)	4.29 (.72)	3.48 (.87)
	Coached feigning	25	20.44 (1.89)	4.16 (.85)	3.68 (.85)
	Internet-coached feigning	18	20.50 (1.54)	4.50 (.78)	3.94 (.64)
	Total	64	20.41 (1.97)	4.30 (.79)	3.69 (.81)
	H(2)		1.10	2.28	3.32
	p		.577	.319	.189
SRSI	Genuine	24	20.17 (2.33)	4.38 (.57)	3.50 (1.10)
	Coached Feigning	18	21.39 (5.94)	4.17 (.98)	3.78(.65)
	Internet-Coached Feigning	24	21.04 (2.45)	4.42 (.65)	3.46 (.84)
	Total	66	20.82 (3.68)	4.33 (.73)	3.56 (.89)
	H(2)		2.48	.42	1.54
	p		.288	.810	.464
IOP-29	Genuine	26	20.77 (2.69)	4.00 (1.09)	3.08 (.93)
	Coached Feigning	21	21.29 (3.95)	4.10 (.70)	3.33 (1.06)
	Internet-Coached Feigning	16	21.31 (2.02)	4.38 (.50)	3.00 (.81)
	Total	63	21.08 (3.00)	4.13 (.85)	3.14 (.95)
	H(2)		1.68	1.14	1.07
	p		.431	.564	.585
	$\epsilon^2$		.002	.001	.01

Note: (1) Gray boxes provide differential statistics of all participants who responded to the corresponding measure; (2)  $H$  = Kruskal–Wallis test statistic,  $p$  = probability value,  $\epsilon^2$  = Epsilon squared effect size value; (3) Anchors for English proficiency and Mental Health: 1 = “terrible/extremely bad” to 5 = “excellent/extremely good”.

had higher ratings, indicating better mental health, than participants in the IOP-29 group ( $p = .002$  and  $p = .025$ , respectively). However, participants receiving SIMS and those with SRSI did not differ in their mental health ratings ( $p \approx 1.00$ ). Only one participant rated their mental health as extremely bad, but this person was assigned to the Genuine condition and was administered the SRSI.

## Measures

### Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997)

The SIMS consists of 75 true/false self-report items gauging probable feigning. SIMS includes five categories of psychopathology, each portrayed by 15 items: (1) *Affective Disorders*, (2) *Neurological Impairment*, (3) *Amnesia*, (4) *Low Intelligence*, and (5) *Psychosis*. The total number of endorsed symptoms represents the total SIMS score, which is the only important “credibility” index of the SIMS. Indeed, the test authors have indicated that the subscales of the SIMS are not useful for detecting feigned psychopathology, as they are only used to assess what type of psychopathology the respondent is attempting to feign once it has been determined that the total SIMS score exceeds the cutoff value (Widows & Smith, 2005). Furthermore, a high SIMS score does not indicate feigning per se. Rather, the SIMS is recommended as a screening tool for symptom overreporting (i.e., negative response bias), which may require further assessment in case of positive outcomes. Higher scores indicate unlikely symptomatology, with recommended cutoff scores of  $>14$  or  $>16$  (i.e., scores exceeding 14 or 16). In this study, the cutoff  $>16$  was used because this cutoff point has received the most empirical support (Van Impelen et al., 2017). The internal consistency estimate of reliability (Cronbach’s alpha) of the SIMS was .90. For this study, we included only the total score in our analyses, consistent with the recommendations in the SIMS manual.

Following recommendations to include checks for inattentive responding in survey measures (e.g., Meade, & Craig, 2012; Ziegler, 2015), we chose to include one check for approximately 20 items. Therefore, three additional items (e.g., *Please respond to this question with “T”*) were randomly inserted into the SIMS. Only two participants failed this check, and they were therefore excluded from the final sample.

### Self-Report Symptom Inventory (SRSI; Merten et al., 2016)

The SRSI consists of 100 items divided into two main scales: the genuine symptoms scale and pseudosymptoms scale. Genuine symptom subscale domains are (1) *cognitive* problems, (2) *depression*, (3) *pain*, (4) *nonspecific somatic hardship*, and (5) *PTSD/anxiety*. Correspondingly, (1) *cognitive/memory*, (2) *neurological motoric ailment*, (3) *neurological sensory problems*, (4) *pain*, and (5) *anxiety/depression* (incl. PTSD) subscales relate to pseudosymptoms. Each subscale includes 10 items. The pseudosymptoms scale items assess symptom over-endorsement. For diagnostic purposes, a cutoff score of  $>9$  pseudosymptom items is recommended, whereas for screening purposes, a score of  $>6$  is applicable. In this study, the Cronbach’s alphas for the genuine symptom scale and pseudosymptoms scale were .94 and .93, respectively. For the whole SRSI, the Cronbach’s alpha was .96.

In addition to the items of the main scales, the SRSI also contains seven items related to cooperativeness and consistency. For the consistency items (five in total), careless responding would be evident if individuals affirmed them in a way that contradicted their symptom report (Boskovic et al., 2020). In this sample, participants in the Genuine condition ( $M = 1.92$ ,  $SD = 1.91$ ) endorsed significantly higher consistency items than participants in the Coached Feigning condition ( $M = .28$ ,  $SD = .67$ ) and Internet-

Coached Feigning condition ( $M = .54$ ,  $SD = .88$ ) (Welch's  $F(2, 40.48) = 7.49$ ,  $p = .002$ ,  $\eta_p^2 = .24$ ). In contrast, no significant differences were observed between the two feigning conditions,  $p = .519$ . Thus, participants in this study consistently responded according to their instructions. For the purposes of this study, we included only the total scores of the two main scales in our analyses.

Five items that screen for random responding were included randomly throughout the scale. Only one participant failed the checks and was removed from the final sample.

### Inventory of Problems-29 (IOP-29; Viglione et al., 2017)

The IOP-29 contains 29 items, 26 of which use an SVT-like format, and three of which use a PVT-like format. Each of the 26 SVT-like items provides three response options: “true,” “false,” and “does not make sense.” The developers of the IOP-29 argue that the option “does not make sense” allows for a more refined understanding of the true *versus* false response selection, by eliminating the dichotomous forced-choice nature of most available SVTs. The IOP-29 generates a False Disorder probability Score (FDS) based on two sets of reference data (i.e., from patients and from experimental feigners) rather than the more standard, T-metric, healthy controls-based normed score. The computation of the FDS is performed via the website <https://www.iop-test.com>. The FDS is to be interpreted as a likelihood value with higher scores reflecting more implausible presentations of a disorder and lower scores nearing bona fide presentations. This format arguably facilitates decision making in real-life settings (Giromini et al., 2018). A cutoff score of  $FDS = .50$  allows classifying a given IOP-29 as mainly non-credible (if  $\geq .50$ ) or mainly credible (if  $< .50$ ). As noted above, even though the IOP-29 is a relatively new SVT, several international studies have already demonstrated that it has the good psychometric properties for use in forensic assessments (Young et al., 2020).

A check for inattentive responses was also added to this scale (“To this item respond with T”). One subject failed the check and was removed from the data.

### Procedure

The study was conducted online. Participants signed up for the study via the Research Participation System (SONA and ERPS), where they received a Qualtrics link that directed them to the study. After providing general information and consent, participants answered demographic questions. In addition, given that this study was conducted in The Netherlands, we asked participants to rate their English proficiency and current mental health on a 5-point scale (1 for terrible proficiency/extremely poor mental health; 5 for excellent English proficiency/extremely good mental health). Participants were then randomly assigned to either the *Genuine* (i.e., honest participants), *Coached Feigning*, or *Internet-Coached Feigning* condition. While the genuine participants were asked to respond honestly, for the participants

in the two feigning conditions we used a *simulation* research paradigm. That is, coached feigning participants received a vignette in which they were instructed to *simulate* a particular condition and respond as if they were the protagonist of a scenario (see Rogers, 2018c). Specifically, they received instructions to feign depression (see Appendix; adapted from Giromini, Lettieri, et al., 2019; see also Pignolo et al., 2021), which also included information about the key symptoms of depression (i.e., symptom coaching) and a warning not to “overdo it.” Participants in the Internet-Coached Feigning condition were also given two links of websites found in online searches for depression (i.e., searches conducted by the authors at planning stage of study) and asked to carefully review their content (see Appendix). These participants were not allowed to continue with the study unless they clicked on the presented links. Participants in the three conditions were then randomly assigned to one of three groups, which would complete one of the following three inventories each: the SIMS, the SRSI, or the IOP-29. After the main task, participants were presented with several exit questions (e.g., difficulty, motivation, clarity of instructions) and a debriefing form. Permission to use and store data anonymously was obtained. Participants were also asked to provide a randomly generated personal identification code in case they wished to withdraw their responses in the future. All participants who completed the study were granted one research participation credit as compensation.

### Analysis

Due to the small sample sizes, for the main analyses, we opted for the non-parametric version of analyses of variance (ANOVA). Hence, within each group (i.e., for each SVT), we conducted *Kruskal–Wallis* tests to assess score differences across the three research conditions (Genuine, Coached Feigning, and Internet-Coached Feigning). When the main effect was statistically significant, pairwise contrasts were examined using Bonferroni correction. Epsilon squared ( $\epsilon^2 = H/(n^2 - 1)/(n + 1)$ ; Tomczak & Tomczak, 2014) was used for the effect size computations. To facilitate future meta-analytic research, we also present means and standard deviations. Furthermore, classification accuracy was tested, too. Data and the outputs are available at *Open Science Framework* (OSF) platform (<https://osf.io/a9km3/>).

### Results

#### *Preliminary analyses: motivation, clarity of instructions, and task difficulty*

Ratings of *motivation to participate* were obtained on a five-point scale (1 = “Extremely unmotivated”; 5 = “Extremely motivated”). The mean value for the whole sample ( $N = 193$ ) was  $M = 4.07$  ( $SD = .65$ ; range 2–5<sup>1</sup>). We also

<sup>1</sup>Four participants reported motivation below the mid-point (3), one received SIMS and three participants received IOP-29. We re-ran the main analyses without these participants, but the results did not differ, hence, we kept them in the dataset.

**Table 2.** Motivation, clarity of instructions, and difficulty scores by group and condition.

Group	Condition	N	Motivation M (SD)	Clarity M (SD)	Difficulty M (SD)
SIMS	Genuine	21	4.14 (.65)	4.90 (.30)	1.86 (.72)
	Coached Feigning	25	3.88 (.67)	4.28 (.79)	2.44 (1.19)
	Internet-Coached Feigning	18	4.28 (.57)	4.56 (.51)	2.44 (1.25)
	Total	64	4.08 (.65)	4.56 (.64)	2.25 (1.10)
	H(2)		3.93	11.38	2.93
	p		.140	.003	.232
	$\epsilon^2$		.06	.18	.04
SRSI	Genuine	24	4.21 (.51)	4.92 (.28)	1.67 (.87)
	Coached Feigning	18	4.33 (.48)	4.61 (.50)	2.56 (1.04)
	Internet-Coached Feigning	24	4.17 (.48)	4.42 (.78)	2.25 (.95)
	Total	66	4.23 (.49)	4.65 (.59)	2.12 (1.00)
	H(2)		1.18	8.99	9.10
	p		.555	.011	.011
	$\epsilon^2$		.018	.14	.14
IOP-29	Genuine	26	4.00 (.63)	4.27 (1.01)	2.67 (.91)
	Coached Feigning	21	3.86 (.79)	4.33 (.79)	2.56 (.81)
	Internet-Coached Feigning	16	3.75 (.85)	3.88 (1.02)	2.56 (.81)
	Total	63	3.89 (.74)	4.19 (.95)	2.46 (.91)
	H(2)		.56	2.65	3.38
	p		.755	.266	.185
	$\epsilon^2$		.009	.04	.05

Note: (1) Gray boxes provide differential statistics of all participants who responded to the corresponding measure; (2)  $H$  = Kruskal–Wallis test statistic,  $p$  = probability value,  $\epsilon^2$  = Epsilon squared effect size value; (3) Anchors for Motivation and Clarity: 1 = “extremely unmotivated/unclear” to 5 = “extremely motivated/clear”; (4) Anchors for Difficulty: 1 = “extremely easy” to 5 = “extremely difficult”.

asked the participants to rate the *clarity of instructions* on a five-point scale (1 = “Extremely unclear”; 5 = “Extremely clear”), obtaining a mean value of  $M = 4.47$  ( $SD = .76$ ;  $N = 193$ ). Lastly, participants rated the *difficulty in filling out the inventories* (1 = “Extremely easy”; 5 = “Extremely difficult”), and the mean value for the whole sample ( $N = 193$ ) was  $M = 2.27$  ( $SD = 1.01$ ).

We then inspected whether, within each SVT group, participants assigned to the three different conditions differed in terms of motivation, clarity of instructions, and difficulty of the task (for descriptive statistics, see Table 2). Genuine, coached feigning, and Internet-coached feigning participants who received SIMS did not significantly differ from each other in terms of motivation ( $H(2) = 3.93$ ,  $p = .140$ ,  $\epsilon^2 = .06$ ), nor in the ratings of difficulty of the task ( $H(2) = 2.92$ ,  $p = .232$ ,  $\epsilon^2 = .04$ ). However, they did provide different ratings of clarity of instructions ( $H(2) = 11.38$ ,  $p = .003$ ,  $\epsilon^2 = .18$ ). Pair-wise comparisons using Bonferroni correction indicated that participants in the Genuine condition rated the clarity of the instructions significantly higher than those in the Coached Feigning condition ( $p = .003$ ), whereas those in the Genuine condition did not differ in their ratings from those in the Internet-Coached Feigning condition ( $p = .108$ ), nor did the two feigning conditions differ from each other ( $p = .939$ ).

A similar trend was observed among the participants who filled out SRSI. Namely, the main effects were statistically significant for both ratings of clarity ( $H(2) = 8.99$ ,  $p = .011$ ,  $\epsilon^2 = .14$ ) and ratings of difficulty ( $H(2) = 9.10$ ,  $p = .011$ ,  $\epsilon^2 = .14$ ). Participants in the Genuine condition rated the clarity of instructions higher than both those in the Coached Feigning condition ( $p = .041$ ) and those in the Internet-Coached Feigning condition ( $p = .004$ ), whereas the two feigning conditions did not significantly differ from each other ( $p = .524$ ). When asked about the difficulty of the

task, participants in the Genuine condition rated it significantly lower than both coached feigners ( $p = .004$ ) and Internet-coached feigners ( $p = .035$ ). Again, the two feigning conditions did not significantly differ from each other ( $p = .353$ ).

There were no significant differences between participants in the Genuine, Coached Feigning, and Internet-Coached Feigning conditions who received the IOP-29 in terms of motivation ( $H(2) = .56$ ,  $p = .755$ ,  $\epsilon^2 = .009$ ), clarity of instructions ( $H(2) = 2.65$ ,  $p = .266$ ,  $\epsilon^2 = .04$ ), or difficulty of the task ( $H(2) = 3.38$ ,  $p = .185$ ,  $\epsilon^2 = .05$ ).

### **SIMS: comparison between genuine, coached feigning, and internet-coached feigning**

To compare whether conditions differed in their endorsement of SIMS items, we ran a *Kruskal–Wallis* test with condition (Genuine vs. Coached Feigning vs. Internet-Coached Feigning) as independent variable and SIMS total as dependent variable. Results showed an overall significant difference between participants assigned to the three conditions,  $H(2) = 20.81$ ,  $p < .001$ ,  $\epsilon^2 = .33$ . More specifically, participants in the Genuine condition ( $M = 11.95$ ,  $SD = 7.41$ ) endorsed fewer items than those in the Coached Feigning ( $M = 24.96$ ,  $SD = 7.85$ );  $p < .001$ ) and Internet-coached Feigning ( $M = 23.78$ ,  $SD = 9.82$ ;  $p = .001$ ) conditions. The two feigning conditions did not significantly differ from each other in their SIMS scores,  $p = .590$ .

### **Detection rate**

We employed the recommended >16 items cut off (see van Impelen et al., 2014) in order to assess the detection accuracy of the SIMS (for the Area Under the Curve (AUC) results for all three measures, see Supplemental Table 1). As

shown in Table 3, 88% of participants in the coached feigning and 78% of participants in the Internet-coached feigning were correctly detected as over-endorsers. Because the two feigning conditions did not yield statistically significantly different results, we also computed the detection rate for *feigners* overall (i.e., combined), which was 84%. Results also showed that 29% of the genuine participants were falsely classified as over-endorsers (i.e., false positives).

### **SRSI: comparison between genuine, coached feigning, and internet-coached feigning**

As with SIMS, we also conducted Kruskal–Wallis tests to investigate whether the three conditions (Genuine vs. Coached Feigning vs. Internet-Coached Feigning) yielded different SRSI's genuine symptoms and pseudosymptoms scores. Participants assigned to the three conditions significantly differed on genuine symptoms scale,  $H(2) = 18.17$ ,  $p < .001$ ,  $\varepsilon^2 = .28$ , and pseudosymptoms scale,  $H(2) = 8.92$ ,  $p = .012$ ,  $\varepsilon^2 = .14$ . Pairwise comparisons with Bonferroni correction indicated that participants in the Genuine condition endorsed significantly fewer items on the genuine symptoms scale ( $M = 17.46$ ,  $SD = 9.70$ ) than those in the Coached Feigning ( $M = 30.11$ ,  $SD = 9.49$ ,  $p = .001$ ) and Internet-Coached Feigning conditions ( $M = 30.38$ ,  $SD = 10.97$ ,  $p < .001$ ). Participants in the Coached Feigning and Internet-Coached Feigning conditions did not exhibit significantly different scores on the genuine symptoms scale,  $ps = 1.00$ . For the pseudosymptoms scale, pairwise comparisons indicated a significant difference between genuine participants ( $M = 4.46$ ,  $SD = 4.04$ ) and coached feigners ( $M = 10.33$ ,  $SD = 7.85$ ;  $p = .018$ ), and a non-significant difference between scores of Genuine condition and Internet-Coached feigners ( $M = 9.75$ ,  $SD = 10.36$ ,  $p = .061$ ). The two feigning conditions obtained similar pseudosymptoms scores ( $p = 1.00$ ).

### **Detection rate SRSI**

Employing the screening cutoff score of  $>6$  pseudosymptoms items, 60% of the feigners were detected. More

specifically, 67% of the participants in the Coached Feigning condition and 54% of those in the Internet-Coached Feigning condition were detected correctly as over-endorsers. However, 25% of genuine participants also endorsed more than 6 pseudosymptoms. Using the recommended standard ( $>9$ ) cutoff point, 41% of the participants in the combined feigning conditions were detected correctly. Specifically, 39% of over-endorsers in the Coached Feigning condition and 42% of those in the Internet-Coached Feigning condition were detected. Also, 17% of genuine participants endorsed more than nine pseudo-items. The sensitivity and specificity rates at these cutoff points are reported in Table 4.

### **IOP-29: comparison between genuine, coached feigning, and internet-coached feigning**

As it was done with the other two measures, we ran the Kruskal–Wallis test with FDS as dependent and condition as independent variables. Overall, the three conditions (Genuine vs. Coached Feigning vs. Internet-Coached Feigning) yielded significantly different scores,  $H(2) = 41.38$ ,  $p < .001$ ,  $\varepsilon^2 = .67$ . Pairwise comparisons with Bonferroni correction indicated that participants in the Genuine condition ( $M = .14$ ,  $SD = .11$ ) obtained a lower FDS score than those in the Coached Feigning ( $M = .77$ ,  $SD = .20$ ,  $p < .001$ ) and Internet-Coached Feigning ( $M = .69$ ,  $SD = .21$ ,  $p < .001$ ) conditions. The two feigning conditions did not differ from each other ( $p = .851$ ).

### **Detection rate**

We tested the detection accuracy of the proposed cutoff score of the IOP-29, i.e.,  $FDS \geq .50$  (Viglione & Giromini, 2020). Employing this cutoff led to the correct classification of 96% of genuine participants, with one participant being classified as non-credible. Further, the employment of the proposed cutoff led to the detection of 86% of participants in the Coached Feigning condition and 88% of those in the Internet-Coached Feigning condition (for more details, see Table 5).

### **Scores on SIMS, SRSI, and IOP-29, and mental health check and english proficiency**

Lastly, we inspected how well the mental health self-reports of participants in the *Genuine* condition associated with their scores on all three measures. Such information could inform on the extent to which genuine psychopathology

**Table 3.** Minimum and maximum score on SIMS across conditions, and sensitivity and specificity for  $>16$  cutoff SIMS.

Condition	n	SIMS score		SIMS $> 16$	
		Minimum	Maximum	Sensitivity	Specificity
Genuine vs.	21	4	27	–	.71
Feigning conditions combined	43	6	40	.84	–
Coached feigning	25	15	42	.88	–
Internet-coached feigning	18	6	40	.78	–

**Table 4.** Minimum and maximum number of endorsed pseudosymptoms across conditions, and sensitivity and specificity for  $>6$  and  $>9$  cutoff SRSI.

Condition	n	SRSI Pseudosymptoms scale					
		Minimum	Maximum	$>6$		$>9$	
				Sensitivity	Specificity	Sensitivity	Specificity
Genuine vs.	24	0	13	–	.75	–	.83
Feigning conditions combined	42	1	50	.60	–	.42	–
Coached feigning	18	1	27	.67	–	.39	–
Internet-coached feigning	24	1	50	.54	–	.42	–

**Table 5.** Minimum and maximum of FDS scores for all conditions and sensitivity and specificity for FDS >.50.

Condition	n	FDS score		FDS = .50	
		Min	Max	Sensitivity	Specificity
Genuine vs.	26	.02	.54	–	.96
Feigning conditions combined	37	.09	.97	.86	–
Coached feigning	21	.18	.97	.86	–
Internet-coached feigning	16	.09	.89	.88	–

Note: FDS = False Disorder Probability Score.

could generate false positive results. The correlation between mental health report (lower number indicating poorer mental health) and total score on SIMS was significant and in a negative direction, *Spearman's rho* =  $-.580$ ,  $p = .006$  ( $n = 21$ ). A similar result was observed for the total genuine score of the SRSI, *Spearman's rho* =  $-.560$ ,  $p = .004$  ( $n = 24$ ), whereas the correlation between mental health ratings and SRSI pseudosymptoms scale scores was not significant, *Spearman's rho* =  $-.367$ ,  $p = .078$ . Finally, there was no significant correlation between mental health reports and IOP-29 FDS scores, *Spearman's rho* =  $-.227$ ,  $p = .264$  ( $n = 26$ ).

We also investigated whether there is a relationship between *genuine* participants' self-reported English proficiency and their scores on all three measures. The results showed that English proficiency of our participants did not significantly correlate with the total score on SIMS, *Spearman's rho* =  $-.060$ ,  $p = .797$ , nor with scores on SRSI's genuine symptom scale, *Spearman's rho* =  $.145$ ,  $p = .500$ , and pseudosymptoms scale, *Spearman's rho* =  $.007$ ,  $p = .972$ . However, the scores of the IOP-29 FDS significantly (positively) correlated with participants' proficiency in English, *Spearman's rho* =  $.414$ ,  $p = .036$ . Yet, once we excluded genuine participants in the IOP-29 condition who scored below 3 on their English proficiency ( $n = 5$ ), the correlation was no longer significant, *Spearman's rho* =  $.238$ ,  $p = .299$  ( $n = 21$ ).

## Discussion

In this simulation study, we aimed to test the robustness of three symptom validity tests (SVTs), namely the SIMS, the SRSI, and the IOP-29, to symptom coaching of psychology students who were asked to feign depression. In addition to information about symptoms of depression, participants who feigned depression also received clear instructions not to exaggerate their symptoms. Further, a group of experimental feigners also received Internet links for additional information about depression. Overall, our preliminary results suggest that all three measures performed decently, but that there is a need to continue this line of research and find ways to further improve the robustness of SVTs to coaching.

First, the results of our study showed that the total scores of all three measures significantly differed by condition, with participants in the Genuine condition generating the most credible symptom presentations (i.e., scored lowest on all instruments) and those in the feigning conditions generating the least credible ones (i.e., scored highest on all

instruments), supporting our primary hypothesis. Differences between participants in the genuine versus feigning conditions were significant on the total SIMS score, on the main genuine symptoms and pseudosymptoms scales on SRSI, and on the "false disorder probability" (FDS) score of the IOP-29.

The SIMS scores of our feigning conditions were relatively similar to those reported in previous SIMS studies. Conversely, our genuine participants endorsed more symptoms than a control group in a prior similar project (Clegg et al., 2009). A similar trend was also noted on the SRSI, with our participants endorsing more items than in previous research on depression, regardless of condition membership (Stevens et al., 2018). This tendency toward excessive over-endorsing, even among genuine participants, may explain their non-significant differences from Internet-coached feigners' scores on the SRSI's pseudosymptoms scale. However, our participants in the Genuine condition scored similar to the low anxiety control group of a previous study on anxiety-related problems (Boskovic et al., 2019). Moreover, Zahid et al. (2022) recently reported a false positive rate of 40% when administering the SIMS on an undergraduate student sample and using the standard cutscore of  $\geq 16$ . As such, albeit somewhat puzzling, our results concerning specificity are not completely in contrast with recent research.

As for the IOP-29, the average scores obtained in the Genuine condition are in line with previous research on the utility of the IOP-29 in the detection of fabricated depression (Giromini, Lettieri, et al., 2019; Ilgunaite et al., 2020; Šömen et al., 2021). In this study, thus, the IOP-29 performed slightly better than the SIMS, which in turn performed slightly better than the SRSI. However, when considering the differences in response style between the three measures, it is important to also note the differences in their length. The IOP-29 is the shortest, followed by the SIMS, and then the SRSI, which contains more than one hundred items. It is therefore possible that in the Genuine conditions our participants' attention waned over time, resulting in higher symptom overendorsement on lengthier measures such as SIMS and SRSI.

Second, the addition of information available on the Internet did not result in statistically significant differences between the two feigning conditions. On the one hand, it did not jeopardize the effectiveness of our studied SVTs; on the other hand, it did not even facilitate their task, as one might expect given that searching for symptoms on the Internet is known to increase symptom endorsement in symptom inventories (Brown et al., 2020; Jungmann et al., 2020; Starcevic & Berle, 2013). Because both conditions received a detailed description of depression symptoms, it is likely that adding information available on the Internet had little additional value. All in all, however, looking at all mean values, Internet-coached participants scored slightly lower than coached feigners on all three measures, suggesting that information available via the Internet may have made participants somewhat more familiar with the

symptoms of genuine depression, and therefore more cautious in item endorsement.

Third, and related to the previous point, the detection rates of SIMS and SRSI suggest that the Internet-coached feigners may have been more reluctant to endorse symptoms than participants in the Coached Feigning condition who had not received additional Internet information. Indeed, participants with Internet coaching were less likely to be detected than coached feigners without Internet information (SIMS: 78% vs 88%; SRSI 54% vs 67%, respectively). With respect to SIMS, these results are somewhat consistent with those previously reported by Merten et al. (2010) and Jelcic et al. (2011). Nevertheless, SRSI detection rates are lower than in previous studies examining symptom coaching (e.g., Boskovic et al., 2019). This discrepancy might be related to our instruction *not to overdo* the symptom presentation, which was intended as a protection against acquiescence response bias (Ray, 1983), and is unlikely to resemble real-life situations. Still, this type of instruction could be considered strategy-coaching, which has been shown to be more successful in reducing symptom endorsement than symptom coaching (Jelcic et al., 2011). Because SIMS and SRSI heavily rely on the evaluatees' tendency to overendorse symptoms, instructing participants about the symptoms of depression while also counterbalancing their overreporting by telling them not to "overdo it" potentially diminished their effectiveness. In particular, it is possible that the SRSI was more affected by such instructions because it contains both genuine symptoms and pseudosymptoms. Thus, it may have been easier for participants who were informed about the symptoms of depression to compare the items presented and to limit their endorsement to genuine items only. This assumption is supported by the evidently higher scores generated by the two feigning conditions on the genuine symptom scale as well, with participants coached via the Internet endorsing more genuine items. The decrease in detection of Internet coached participants compared to feigners without additional coaching was the least apparent on the IOP-29 (86% vs 88%). Our results are very similar to those previously reported on the effects of coaching and performance on the IOP-29 (Gegner et al., 2021). However, it is important to note that the coaching instructions of both conditions of feigners were taken from a previously published paper using the IOP-29 (Giromini et al., 2019). Thus, further research with other instructions is needed to more fairly compare the effectiveness of the three SVTs included in this study, as it may be that the IOP-29 was at an advantage in this study, especially considering that not all of its items follow the usual SVT rationale.

Finally, we also examined whether scores on SIMS, SRSI, and IOP-29 correlated with the mental health ratings that *genuine* participants had provided before the study began. The results showed that the better the mental health of the participants, the lower the scores on SIMS, and a similar trend was observed for SRSI genuine scale, too. There was no significant correlation between mental health self-report ratings and IOP-29 scores. We also checked whether the self-reported English proficiency of our genuine participants

was related to their scores on all three measures, and only the score on the IOP-29 was positively related to language proficiency. That is, the better the language proficiency, the higher the FDS. However, when the few outlier participants who self-rated their language skills as "poor" were excluded, this correlation was no longer significant. This finding is nevertheless surprising, as previous research has shown that speakers with lower proficiency tend to score higher on SVTs (van der Heide et al., 2020). Since the IOP-29 also contains tasks similar to those of the PVTs, this result needs further investigation.

It is important to reflect on the limitations of this study. First, despite the ambitious goal of including SIMS, SRSI, and IOP-29 in this project, we did not administer the tests together with the same participants. Thus, while we attempted to examine the comparative validity of the SIMS, SRSI, and IOP-29, we did not do so in a sufficiently exhaustive manner and future research should focus on this. Second, our sample size may have been too small to detect small to medium sized differences, not only between genuine participants and feigners, but specifically between the two feigning conditions. More importantly, our study revealed an unexpectedly high number of false positive results on SIMS and SRSI compared to previous studies conducted in person using the standard paper-and-pencil format. It should be noted that we administered the tests online, which, according to the test developers, represents a significant deviation from standard conditions of use (Merten et al., 2021). On the other hand, it is also worth noting that our study was conducted during the Covid-19 lock-down period, and that it is likely that our participants experienced higher levels of general distress than samples used in the past. Third, and relatedly, our sample consisted of undergraduates, particularly female psychology students, which limits the ecological validity of our findings (Rogers, 2018a). We also suggest that future investigations include checks for prior knowledge about the symptoms participants are asked to fabricate. In this regard, psychology students may not be a representative sample. Considering that we did not have a control condition consisting of patients with depression, our results can only be considered in terms of sensitivity and potential indications for future studies. In addition, it is possible that our genuine participants did not fully follow the instructions and perhaps tended to exaggerate their psychological problems. Fourth, because our study was conducted in The Netherlands, our participants were most likely not native English speakers, and their language proficiency was self-reported. Although they receive education in English, there is still a possibility that some of them did not fully understand the items they were responding to. Another important omission is that we did not specifically ask about the nationality of the participants, so we cannot determine the percentage of students who were native English speakers. Finally, this study was conducted online using Qualtrics. Inattentive responding is known to be more common in studies conducted in this format, so we included inattentiveness checks in all three measures. However, we could not ensure, for example, that participants in the Internet-

Coached Feigning condition actually read/listened to the content of the Internet links, because the program could only check that they clicked on the links to continue with the study.

Taken together, our preliminary findings suggest that all three SVTs included in the study, i.e., SIMS, SRSI, and IOP-29, are sensitive to feigned depression even in the presence of symptom coaching, both with and without additional internet-based information. However, additional research using a clinical comparison simulation or criterion groups study design in which all three SVTs are administered to the same individuals, as well as the inclusion of different or alternative instructions, patient samples, and objective tests of language proficiency, would be beneficial.

## Ethical approval

This study was conducted in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

## Disclosure statement

The first and second authors declare no conflict of interest. The third author declares that he owns a share in the corporate (LLC) that possesses the rights to Inventory of Problems instruments.

## Funding

The author(s) reported there is no funding associated with the work featured in this article.

## ORCID

Irena Boskovic  <http://orcid.org/0000-0003-1675-1004>  
Ali Yunus Emre Akca  <http://orcid.org/0000-0002-3450-7059>  
Luciano Giromini  <http://orcid.org/0000-0002-9540-4803>

## References

- Akca, Y. E. A., Sloommaekers, L., & Boskovic, I. (2020). Verifiability and symptom endorsement in genuine, exaggerated, and malingered pain. *Psychological Injury and Law*, 13, 235–245. <https://doi.org/10.1007/s12207-020-09375-w>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.
- Banovic, I., Filippi, F., Viglione, D. J., Scrima, F., Zennaro, A., Zappalà, A., & Giromini, L. (2021). Detecting coached feigning of schizophrenia with the Inventory of Problems – 29 (IOP-29) and its Memory Module (IOP-M): A simulation study on a French community sample. *International Journal of Forensic Mental Health*, 21(1), 37–53. <https://doi.org/10.1080/14999013.2021.1906798>
- Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology*, 19, 96. <https://doi.org/10.1136/practneurol-2018-001950>
- Bauer, L., & McCaffrey, R. J. (2006). Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? *Archives of Clinical Neuropsychology*, 21(1), 121–126. <https://doi.org/10.1016/j.acn.2005.06.010>
- Boskovic, I. (2019). *A multi-method approach to the detection of fabricated symptoms* [Doctoral Thesis Maastricht University]. <https://doi.org/10.26481/dis.20190129ib>
- Boskovic, I., Hope, L., Ost, J., Orthey, R., & Merckelbach, H. (2019). Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task. *Journal of Behavior Therapy and Experimental Psychiatry*, 65, 101483. <https://doi.org/10.1016/j.jbtep.2019.101483>
- Boskovic, I., Merckelbach, H., Merten, T., Hope, L., & Jelicic, M. (2020). The Self-Report Symptom Inventory as an instrument for detecting symptom over-reporting: An exploratory study with instructed simulators. *European Journal of Psychological Assessment*, 36, 730–739. <https://doi.org/10.1027/1015-5759/a000547>
- Boskovic, I., van der Heide, D., Hope, L., Merckelbach, H., & Jelicic, M. (2017). Plausibility judgments of atypical symptoms across cultures: An explorative study among western and non-western experts. *Psychological Injury and Law*, 10(3), 274–281. <https://doi.org/10.1007/s12207-017-9294-6>
- Brown, R. J., Skelly, N., & Chew-Graham, C. A. (2020). Online health research and health anxiety: A systematic review and conceptual integration. *Clinical Psychology: Science and Practice*, 27, e12299. <https://doi.org/10.1111/cpsp.12299>
- Bush, S. S., Ruff, R. M., Troster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20(4), 419–426. <https://doi.org/10.1016/j.acn.2005.02.002>
- Carvalho, L., Reis, A., Colombaroli, M. S., Pasian, S. R., Miguel, F. K., Erdodi, L. A., Viglione, D. J., & Giromini, L. (2021). Discriminating feigned from credible PTSD symptoms: a validation of a Brazilian version of the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law*, 14, 58–70. <https://doi.org/10.1007/s12207-021-09403-3>
- Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American academy of clinical neuropsychology social security administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29(6), 723–740. <https://doi.org/10.1080/13854046.2015.1099738>
- Clegg, C., Fremouw, W., & Mogge, N. (2009). Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among outpatients seeking to claim disability. *The Journal of Forensic Psychiatry & Psychology*, 20, 239–254.
- Crîșan, I., Maricuțoiu, L. P., & Sava, F. A. (2021). Strategies to detect invalid performance in cognitive testing: An updated and extended meta-analysis. *Current Psychology*. Advance online publication. <https://doi.org/10.1007/s12144-021-01659-x>
- Dandachi-FitzGerald, B., Merckelbach, H., Bošković, I., & Jelicic, M. (2020). Do you know people who feign? Proxy respondents about feigned symptoms. *Psychological Injury and Law*, 13, 225–234. <https://doi.org/10.1007/s12207-020-09387-6>
- Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: a survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28(8), 771–783. <https://doi.org/10.1093/arclin/act073>
- Essig, S. M., Mittenberg, W., Petersen, R. S., Strauman, S., & Cooper, J. T. (2001). Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology*, 16(3), 271–291.
- Fazio, R. L., Sanders, J. F., & Denney, R. L. (2015). Comparison of performance of the Test of Memory Malingering and Word Memory Test in a criminal forensic sample. *Archives of Clinical Neuropsychology*, 30(4), 293–301. <https://doi.org/10.1093/arclin/acv024>
- Gegner, J., Erdodi, L. A., Giromini, L., Viglione, D. J., Bosi, J., & Brusadelli, E. (2021). An Australian study on feigned mTBI using the Inventory of Problems – 29 (IOP-29), its Memory Module (IOP-M), and the Rey Fifteen Item Test (FIT). *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1864375>

- Giger, P., & Merten, T. (2019). Equivalence of the German and the French Versions of the Self-Report Symptom Inventory. *Swiss Journal of Psychology*, 78, 5–13. <https://doi.org/10.1024/1421-0185/a000218>
- Giromini, L., Barbosa, F., Coga, G., Azeredo, A., Viglione, D. J., & Zennaro, A. (2019). Using the inventory of problems – 29 (IOP-29) with the Test of Memory Malingering (TOMM) in symptom validity assessment: A study with a Portuguese sample of experimental feigners. *Applied Neuropsychology: Adult*, 27(6), 504–516. <https://doi.org/10.1080/23279095.2019.1570929>
- Giromini, L., Lettieri, S. C., Zizolfi, S., Zizolfi, D., Viglione, D. J., Brusadelli, E., Perfetti, B., di Carlo, D. A., & Zennaro, A. (2019). Beyond rare-symptoms endorsement: A clinical comparison simulation study using the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) with the Inventory of Problems-29 (IOP-29). *Psychological Injury and Law*, 12, 212–224. <https://doi.org/10.1007/s12207-019-09357-7>
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2018). A clinical comparison, simulation study testing the validity of SIMS and IOP-29 with an Italian sample. *Psychological Injury and Law*, 11, 340–350. <https://doi.org/10.1007/s12207-018-9314-1>
- Giromini, L., Viglione, D. J., Pignolo, C., & Zennaro, A. (2020). An Inventory of Problems-29 sensitivity study investigating feigning of four different symptom presentations via malingering experimental paradigm. *Journal of Personality Assessment*, 102(4), 563–572. <https://doi.org/10.1080/00223891.2019.1566914>
- Giromini, L., & Viglione, D. J. (2022). Assessing Negative Response Bias with the Inventory of Problems-29 (IOP-29): A quantitative literature review. *Psychological Injury and Law*. Advance Online Publication. <https://doi.org/10.1007/s12207-021-09437-7>
- Giromini, L., Young, G., & Sellbom, M. (2022). Assessing negative response bias using self-report measures: New articles, new issues. *Psychological Injury and Law*. Advance Online Publication. <https://doi.org/10.1007/s12207-022-09444-2>
- Grant, A. F., Lace, J. W., Teague, C. L., Lowell, K. T., Ruppert, P. D., Garner, A. A., & Gfeller, J. D. (2020). Detecting feigned symptoms of depression, anxiety, and ADHD, in college students with the structured inventory of malingered symptomatology. *Applied Neuropsychology: Adult*. Advance online publication. <https://doi.org/10.1080/23279095.2020.1769097>
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference, P. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23(7), 1093–1129. <https://doi.org/10.1080/13854040903155063>
- Ilgunaite, G., Giromini, L., Bosi, J., Viglione, D. J., & Zennaro, A. (2020). A clinical comparison simulation study using the Inventory of Problems-29 (IOP-29) with the Center for Epidemiologic Studies Depression Scale (CES-D) in Lithuania. *Applied Neuropsychology: Adult*, 29(2), 155–162. <https://doi.org/10.1080/23279095.2020.1725518>
- Jelicic, M., Ceunen, E., Peters, M. J. V., & Merckelbach, H. (2011). Detecting coached feigning using the test of Memory Malingering (TOMM) and the structured inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology*, 67(9), 850–855. <https://doi.org/10.1002/jclp.20805>
- Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malingering tests: A simulation study in naïve and coached malingerers. *The International Journal of Neuroscience*, 117(8), 1185–1192. <https://doi.org/10.1080/00207450600934697>
- Jungmann, S. M., Brand, S., Kolb, J., & Witthöft, M. (2020). Do Dr. Google and health apps have (comparable) side effects? An experimental study. *Clinical Psychological Science*, 8, 306–317. <https://doi.org/10.1177/2167702619894904>
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18(4), 625–630. <https://doi.org/10.1017/s1355617712000240>
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). [Formula: see text] Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29(6), 741–776. Neuropsychologists' <https://doi.org/10.1080/13854046.2015.1087597>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455.
- Merten, T., Dandachi-FitzGerald, B., Boskovic, I., Puente-López, E., & Merckelbach, H. (2021). The Self-Report Symptom Inventory. *Psychological Injury and Law*, 15, 94–103. <https://doi.org/10.1007/s12207-021-09434-w>
- Merten, T., Lorenz, R., & Schlatow, S. (2010). Posttraumatic stress disorder can easily be faked, but faking can be detected in most cases. *German Journal of Psychiatry*, 13, 140–149.
- Merten, T., & Merckelbach, H. (2012). Forced-choice tests as single-case experiments in the differential diagnosis of intentional symptom distortion. *Journal of Experimental Psychopathology*, 4, 20–37. <https://doi.org/10.5127/jep.023711>
- Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, 6, 122–137. <https://doi.org/10.1007/s12207-013-9155-x>
- Merten, T., & Merckelbach, H. (2020). Factitious disorder and malingering. In J. R. Geddes, N. C. Andreasen, & G. M. Goodwin (Eds.), *New Oxford textbook of psychiatry* (3rd ed., pp. 1342–1349). Oxford University Press.
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9, 102–111. <https://doi.org/10.1007/s12207-016-9257-3>
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24(8), 1094–1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, 29(11), 1321–1336. <https://doi.org/10.1037/pas0000438>
- Pignolo, C., Giromini, L., Ales, F., & Zennaro, A. (2021). Detection of feigning of different symptom presentations with the PAI and IOP-29. *Assessment*. Advance Online Publication. <https://doi.org/10.1177/10731911211061282>
- Ray, J. J. (1983). Reviving the problem of acquiescent response bias. *The Journal of Social Psychology*, 121, 81–96. <https://doi.org/10.1080/00224545.1983.9924470>
- Rogers, R. (2018a). Structured interviews and dissimulation. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 422–449). Guilford press.
- Rogers, R. (2018b). An introduction to response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 3–17). Guilford Press.
- Rogers, R. (2018c). Researching response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 592–614). Guilford Press.
- Rogers, R., & Bender, S. D. (Eds.). (2018). *Clinical assessment of malingering and deception* (4th ed.). Guilford Press.
- Roma, P., Giromini, L., Burla, F., Ferracuti, S., Viglione, D. J., & Mazza, C. (2020). Ecological validity of the Inventory of Problems-29 (IOP-29): An Italian study of court-ordered, psychological injury evaluations using the Structured Inventory of Malingered Symptomatology (SIMS) as Criterion Variable. *Psychological Injury and Law*, 13, 57–65. <https://doi.org/10.1007/s12207-019-09368-4>
- Ruiz, M. A., Drake, E. B., Glass, A., Marcotte, D., & van Gorp, W. G. (2002). Trying to beat the system: Misuse of the internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology: Research and Practice*, 33, 294–299. <https://doi.org/10.1037/0735-7028.33.3.294>
- Sherman, E. M. S., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological

- dysfunction criteria. *Archives of Clinical Neuropsychology*, 35(6), 735–764. <https://doi.org/10.1093/arclin/aca019>
- Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545)
- 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, 183–189.
- Šömen, M. M., Lesjak, S., Majaron, T., Lavopa, L., Giromini, L., Viglione, D. J., & Podlessek, A. (2021). Using the Inventory of Problems-29 (IOP-29) with the Inventory of Problems Memory (IOP-M) in malingering-related assessments: a study with a Slovenian sample of experimental feigners. *Psychological Injury and Law*, 14(2), 104–113. <https://doi.org/10.1007/s12207-021-09412-2>
- Starcevic, V., & Berle, D. (2013). Cyberchondria: Towards a better understanding of excessive health-related Internet use. *Expert Review of Neurotherapeutics*, 13(2), 205–213. <https://doi.org/10.1586/ern.12.162>
- Stevens, A., Schmidt, D., & Hautzinger, M. (2018). Major depression – A study on the validity of clinicians' diagnoses in medicolegal assessment. *The Journal of Forensic Psychiatry & Psychology*, 29, 794–809. <https://doi.org/10.1080/14789949.2018.1477974>
- Suhr, J. A., & Gunstad, J. (2007). Coaching and malingering: A review. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 287–311). Oxford University Press.
- Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends in Sport Sciences*, 1, 19–25.
- Van Der Heide, D., Boskovic, I., Van Harten, P., & Merckelbach, H. (2020). Psychosis as a confounder of symptom credibility testing in a transcultural sample. *The Journal of Forensic Psychiatry & Psychology*, 31, 64–75.
- van Helvoort, D., Merckelbach, H., & Merten, T. (2019). The Self-Report Symptom Inventory (SRSI) is sensitive to instructed feigning, but not to genuine psychopathology in male forensic inpatients: An initial study. *The Clinical Neuropsychologist*, 33(6), 1069–1082. <https://doi.org/10.1080/13854046.2018.1559359>
- van Impelen, A., Merckelbach, H., Jellic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, 28(8), 1336–1365. <https://doi.org/10.1080/13854046.2014.984763>
- van Impelen, A., Merckelbach, H., Jellic, M., Niesten, I. J. M., & Campo, J. À. (2017). Differentiating factitious from malingered symptomatology: The development of a psychometric approach. *Psychological Injury and Law*, 10(4), 341–357. <https://doi.org/10.1007/s12207-017-9301-y>
- Viglione, D. J., Giromini, L., & Landis, P. (2017). The development of the Inventory of Problems-29: A brief self-administered measure for discriminating bona fide from feigned psychiatric and cognitive complaints. *Journal of Personality Assessment*, 99(5), 534–544. <https://doi.org/10.1080/00223891.2016.1233882>
- Viglione, D. J., & Giromini, L. (2020). *Inventory of problems-29: Professional manual*. IOP-Test, LLC.
- Widows, M. R., & Smith, G. P. (2005). *Structured inventory of malingered symptomatology professional manual*. Psychological Assessment Resources.
- Winters, C. L., Giromini, L., Crawford, T. J., Ales, F., Viglione, D. J., & Warmelink, L. (2020). An Inventory of Problems-29 (IOP-29) study investigating feigned schizophrenia and random responding in a British community sample. *Psychiatry, Psychology and Law*, 28(2), 235–254. <https://doi.org/10.1080/13218719.2020.1767720>
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law*, 8, 188–199. <https://doi.org/10.1007/s12207-015-9232-4>
- Young, G., Foote, W. E., Kerig, P. K., Mailis, A., Brovko, J., Kohutis, E. A., McCall, S., Hapidou, E. G., Fokas, K. F., & Goodman-Delahunty, J. (2020). Introducing psychological injury and law. *Psychological Injury and Law*, 13, 452–463. <https://doi.org/10.1007/s12207-020-09396-5>
- Zahid, A., Dhillon, S., & Zakzanis, K. K. (2022). Examining base rates of symptom endorsement and the roles of sex and depressive symptoms on the Structured Inventory of Malingered Symptomatology (SIMS) in a non-clinical population. *Psychological Injury and Law*. Advance Online Publication. <https://doi.org/10.1007/s12207-021-09439-5>
- Ziegler, M. (2015). “F\*\*\* you, I won’t do what you told me!” – Response biases as threats to psychological assessment. *European Journal of Psychological Assessment*, 31, 153–158.

## Appendix

### Instructions

#### Feigning condition

##### Dear participant

Soon, we will ask you to fill out a questionnaire. However, we want you to fill out that questionnaire as someone who is diagnosed with major depression disorder. We would like you to put yourself in the shoes of a person who has had an accident at work and is now suffering from mental health problems—namely, **depression**—related to that accident and for which he has requested to be put on disability. To help you provide a credible presentation, please read the following text, and try to pretend that you are the person depicted in this scenario.

You are an administrator at a small, well-established firm. Your boss has been trying to cut expenses by having the cleaning crew work before regular work hours are over, thus getting the job done at a cut rate. You have repeatedly informed him that this is not a safe working condition for the employees, but he has not changed the procedure. One day, near the end of the day, you are leaving to do a special errand for your boss. As you cross a freshly mopped floor, you slip and fall, landing hard on your tailbone. As a result, you have been out of work for 2 weeks on disability and continue to experience a fair amount of pain, particularly when you sit for any length of time. The workers compensation physician insists that he can find nothing to explain the pain and refuses to authorize any more time off or disability payments, stating that you are able to return to work, a job that requires long periods of time sitting at your computer. Before terminating your case, the physician refers you to the staff psychologist for a routine evaluation. You realize that this evaluation is your only opportunity to remain on disability under your employer's obligation. You have no additional coverage and need an income until you are fully recovered. You also feel that your boss is responsible, and that money should come from the company through workers compensation. So, your only choice is to present yourself as having significant depression on the tests that the psychologist is going to give you. You therefore decide to attempt to present yourself as having a major depression as the result of your accident, to remain on disability.<sup>^</sup>

Here the symptoms of the Major Depression Disorder. Keep in mind that depressed patients typically have 5 or more of the following symptoms, but most likely not all of them: 1. Depressed mood most of the day, nearly every day (e.g., feeling sad, empty, hopeless), 2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day, 3. Significant weight loss when not dieting or weight gain, or decrease or increase in appetite nearly every day, 4. Insomnia or hypersomnia nearly every day, 5. Psychomotor agitation or retardation nearly every day, 6. Fatigue or loss of energy nearly every day, 7. Feelings of worthlessness or excessive or inappropriate guilt nearly every day, 8. Diminished ability to think or concentrate, or indecisiveness, nearly every day, 9. Recurrent thoughts of death, recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

When you take the tests and try to pretend you suffer from a Major Depressive Disorder, please keep in mind that if you present your condition in an extremely dramatic way, your performance may not be believable, and the examiner might understand that you do not suffer from depression but are only faking it. So, try to not over-do it. If you will be able to produce test results that are consistent with those

produced by people who really suffer from Major Depression Disorder and you will not look like a feigner, you may win a small prize consisting of a 10€gift card in the lottery!

### *Internet-coached feigning condition*

The same as for the Feigning Condition, but plus:

It is known that people are better at feigning depression if they check the internet for available information. So, we found two links for you and ask you to click on the following links and read the presented information. Please do not close this tab, as the study will stop, but just open the links in a new tab and come back once you are done

reading the information. It is of high importance that you actually check these links as you will not be allowed to proceed with the study if you do not. But please do not continue the study unless you read the provided information.

The information provided on these links is brief and it will not take you a long time to read it:

1. Watch a 1-minute video and read the text:  
<https://www.webmd.com/depression/guide/detecting-depression>
2. Read about psychological, physical, and other symptoms of depression:  
<https://www.nhs.uk/conditions/clinical-depression/symptoms/>