

# Accepted Manuscript

Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task

Irena Boskovic, Lorraine Hope, James Ost, Robin Orthey, Harald Merckelbach



PII: S0005-7916(18)30266-0

DOI: <https://doi.org/10.1016/j.jbtep.2019.101483>

Article Number: 101483

Reference: BTEP 101483

To appear in: *Journal of Behavior Therapy and Experimental Psychiatry*

Received Date: 21 September 2018

Revised Date: 10 May 2019

Accepted Date: 12 May 2019

Please cite this article as: Boskovic, I., Hope, L., Ost, J., Orthey, R., Merckelbach, H., Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task, *Journal of Behavior Therapy and Experimental Psychiatry* (2019), doi: <https://doi.org/10.1016/j.jbtep.2019.101483>.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task**

Irena Boskovic<sup>12</sup>, Lorraine Hope<sup>2</sup>, James Ost<sup>2\*</sup>, Robin Orthey<sup>12</sup> & Harald Merckelbach<sup>1</sup>

<sup>1</sup>Maastricht University, the Netherlands

<sup>2</sup>University of Portsmouth, UK

\*Requests for reprints should be addressed to Irena Boskovic, Maastricht University, Forensic Psychology Section, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: irena.boskovic@maastrichtuniversity.nl).

*Acknowledgement:* This work was supported by the House of Legal Psychology / Erasmus Mundus Joint Doctorate Program. Framework Partnership Agreement (FPA) 2013-0036. Cohort 2015 with Specific Grant Agreement (SGA) 2015-1610.

---

\* Our dear colleague and friend, James Ost, recently passed away. He will be greatly missed.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13

**Detecting feigned high impact experiences: A symptom over-report questionnaire  
outperforms the emotional Stroop task**

ACCEPTED MANUSCRIPT

14

## Abstract

15 *Background and Objectives:* The Modified Stroop Task (MST) effect refers to a prolonged  
16 reaction time (RT) in color-naming words related to an individual's disorder. Some authors  
17 argue that its absence in people who claim symptoms might be an indication of feigning.

18 *Method:* We tested whether the MST effect is robust against feigning attempts and compared  
19 its absence as an index of feigning with over-reporting tendencies on a symptom  
20 questionnaire (i.e., the Self-Reported Symptom Inventory (SRSI)). We included participants  
21 ( $n = 22$ ) who i) reported current high impact of aversive experiences (High scorers), ii)  
22 reported current low impact (Low scorers) of aversive experiences ( $n = 24$ ), and iii) actors ( $n =$   
23 18) with low impact, but instructed to feign current high psychological impact of aversive life  
24 events (Simulators). We administered the MST, including impact-related, neutral, and  
25 feigning-related words, and the SRSI. *Results:* We found no MST effect for impact-related  
26 words in the high scorers group, or for feigning-related words in the simulators. Relative to  
27 high scorers and low scorers, simulators exhibited significantly longer RTs on all types of  
28 words and they also endorsed significantly more bogus symptoms on the SRSI. Thus, the  
29 SRSI was a more sensitive measure of feigning than the absence of an MST effect.

30 *Limitation:* Some limitations are related to our reliance on a sub-clinical student sample,  
31 whereas others reflect the unresolved issues surrounding the MST. Thus, the generalizability  
32 of our results is uncertain. *Conclusion:* Our findings add to the doubts on the idea that the  
33 MST can be used to differentiate between genuine and feigned complaints.

34 *Key words:* Modified Stroop task, Feigning, SRSI, Simulators.

35

**Detecting feigned high impact experiences: A symptom over-report questionnaire  
outperforms the emotional Stroop task**

The original Stroop task (Stroop, 1935) constitutes of color words presented in different colors, of which the participants should name the color and disregard the content. The stimuli are divided into congruent (word “red” presented in red color) and incongruent (word “red” presented in green color) categories. In order to explore participants’ attentional bias, the reaction time to the two types of stimuli is compared, and the difference is known as the *Stroop Effect* [ $SE=RT(\text{incongruent})-RT(\text{congruent})$ ] (see MacLeod, 1991). The Modified Stroop task (MST), also known as the Emotional Stroop task (e.g., Williams et al., 1996), although sharing a similar name with the original task, crucially differs from it (see Algom, Chajut, Lev, 2004). To begin with, the MST lacks the (in)congruency of stimuli, and instead includes disorder-related and neutral words of which the colors have to be named as quickly as possible while ignoring the meaning of the words (e.g., Buckley, Blanchard, & Neill, 2000). The reaction times of participants to the different word categories are compared, and, if a difference emerges, it is labeled as the *MST effect*. This delayed response is considered to be a sign of attentional bias among genuine patients (Elsesser, Sartory & Tackenberg, 2004), although some authors state that, in the case of the MST, other, not attention-related mechanisms, might rather be driving this effect (Algom et al., 2004; see below).

Many authors report that people with certain psychological disorders exhibit delays in naming the color of words that specifically refer to their disorder (e.g., McNally, English, & Lipke, 1993). For example, several studies have found that PTSD patients take longer to color-name trauma-relevant than trauma-irrelevant words (Buckley, Galovski, Blanchard, & Hickling, 2003). In this vein, these authors stated that the presence of the MST effect for trauma-related words reflects genuine trauma symptoms, whereas the absence of the effect in those who claim trauma symptoms might be a sign of intentional feigning. Thus, Buckley at

61 el. (2003) proposed that the MST could serve as a screening tool for differentiating between  
62 feigners and people truly suffering from PTSD. This argument was supported by the  
63 assumption that the reaction-time measures are less vulnerable to manipulation than patients'  
64 self-reports (Constans et al., 2014; Beck, Freeman, Shipherd, Hamblen, & Lackner, 2001;  
65 Kaspi, McNally, & Amir, 1995).

66 To test this hypothesis, Buckley et al. (2003) used the MST to investigate the  
67 sensitivity of the MST effect to feigning. Six PTSD patients, six actors who were instructed to  
68 feign PTSD, and six non-anxious participants (control group) took part in the study. Actors  
69 were taught about PTSD, but not about the typical response pattern of PTSD patients on the  
70 MST. All participants completed an MST that involved PTSD-related and neutral words.  
71 Overall, PTSD patients and actors had significantly slower reaction times than controls for all  
72 words presented. However, the response latencies of actors were similar for PTSD-related  
73 words and neutral words, while participants in the PTSD group exhibited the typical MST  
74 effect (i.e., slower color-naming for PTSD-related words only). These results led the  
75 researchers to conclude that absence of an MST effect might indicate feigning of PTSD:  
76 "...reaction times to Stroop tasks may be useful adjuncts to clinicians who are attempting to  
77 discern psychopathology from malingering" (Buckley et al., 2003, p. 65). They also argued  
78 that future work should focus on the "magnitude of difference" between latencies for color-  
79 naming trauma-related words and latencies for neutral words so as to determine a cutoff point  
80 that might enhance diagnostic decision making (Buckley et al., 2003).

81 Three considerations a priori cast doubt on the idea that the absence of an MST effect  
82 is diagnostic of feigned PTSD or any other disorder. First, the idea presupposes that MST  
83 effects in psychopathological groups – e.g., people with PTSD – are robust, but evidence  
84 suggest they are not. For example, an earlier study of Buckley and colleagues (2002) found  
85 that a PTSD group, compared with a panic disorder group and controls, did *not* show the

86 specific MST effect but rather an overall longer reaction time on all three categories (i.e.,  
87 neutral, panic-related, and PTSD-related) of words (Buckley, Blanchard, & Hickling, 2002).  
88 In other research, the MST effect in PTSD patients has been difficult to reproduce (e.g.,  
89 Shipherd & Salters-Pedneault, 2008) and a number of unpublished dissertations have reported  
90 small or no MST effects in PTSD patients (e.g., Kimble, Frueh, & Marks, 2009).

91         Second, and related to the first point, the extant literature is unclear about the reaction  
92 time pattern that should be characteristic for feigners: absence of a specific MST effect or  
93 general slowing of reaction times? Constans et al. (2014) examined the MST performance of  
94 war veterans with and without PTSD and with and without a tendency to over-report  
95 symptoms (which conceptually comes close to feigning). Patients with over-reporting  
96 tendencies had overall significantly longer response times than PTSD patients without over-  
97 reporting style tendencies or controls. These results raise the question as to what an overall  
98 delayed response pattern indicates. Does it reflect the cognitive load (Vrij, Fisher, Mann, &  
99 Leal, 2008) of feigners who are pre-occupied with how to fabricate their symptoms in a  
100 convincing way? If that is the case, general (i.e., non-specific) slowing of responses rather than  
101 the specific absence of an MST effect might be diagnostic of feigning. These mixed findings  
102 raise a more fundamental question: What does the Modified Stroop task actually measure? Is  
103 it the attentional bias or emotional arousal associated with certain words (MacLeod, Mathews,  
104 & Tata, 1986), or a preoccupation with a specific topic (Cannon, 2003; Mathews & MacLeod,  
105 1985)? For example, Cannon (2003) found an MST effect for feigning-related words (e.g., lie,  
106 fake) among students who were instructed to feign mild brain trauma. This suggests that the  
107 MST effect merely reflects participants' current concerns – whether they pertain to their  
108 disorder or deception relating to that disorder. However, Thomas and Fremouw (2009) failed  
109 to find the MST effect for feigning-related words in a group with PTSD, a group without  
110 PTSD, and instructed feigners, which suggests that a specific delay of reaction times even for

111 feigning words in those who are instructed to feign is unreliable. Some authors argue that the  
112 traditional SE and the MST effect rely on different mechanisms and that, while the SE  
113 actually captures the attentional bias, the MST effect is exhibited due to the threat-driven  
114 generic slowdown (Algom et al., 2004).

115 A third issue concerns the diagnostic power of the MST effect, and specifically the  
116 purported likelihood of the absence of the effect in feigners. Even if it is assumed that the  
117 MST effect is robust in the sense that it occurs far more often in genuine patients than in  
118 feigners, the question remains whether its absence allows for better identification of feigners  
119 than self-report instruments that screen for symptom over-reporting. An example of such self-  
120 report instrument is the Self-Report Symptom Inventory (SRSI; Merten, Merckelbach, Giger,  
121 & Stevens, 2016), which includes genuine and pseudosymptoms. Our previous study  
122 (Boskovic, Biermans, Merten, Jelicic, Hope, & Merckelbach, 2018) focused on feigned test-  
123 anxiety in students and showed that the MST effect is easy to fabricate and that the SRSI was  
124 more effective in detecting feigners than the absence of an MST effect.

125 In this study, we wanted to compare the MST with the SRSI to examine the extent to  
126 which both methods discriminate between feigned and genuine PTSD-like symptoms, using a  
127 similar design to the study by Buckley et al. (2003). However, the standing ethical committee  
128 restrained us from including students with full-blown PTSD in our sample. Instead, we were  
129 allowed to include a subclinical student sample comprising (i) individuals who reported  
130 currently experiencing a high psychological impact of previous aversive life events (High  
131 scorers); (ii) individuals who reported a low impact of previous aversive life events (Low  
132 scorers); and (iii) Actors who were instructed to feign PTSD symptoms. The three groups  
133 were administered an MST involving impact-related, feigning-related, and neutral words.  
134 Additionally, we administered the SRSI. We expected that actors would show the typical  
135 MST effect for both trauma-related and feigning-related words (Buckley et al., 2003; Cannon,



2003). We also anticipated that Actors would endorse significantly more genuine and pseudosymptoms on the SRSI than the High or Low scorers, and that by using this measure it would be possible to detect over 70% of Actors as simulators, as in our previous study (Boskovic et al., 2018). In order to have a closer look at participants' traits that might have an influence on their responding style, we also screened participants for their fantasy proneness (rich imagery; Wilson & Barber, 1983). Fantasy proneness was shown to be related with over-endorsement of atypical items on symptoms measures (Merckelbach, 2004), thus, we wanted to investigate whether this trait plays any role in their performance.

## Method

### Sample

In total, we recruited 138 participants, 92 psychology students for the two impact scores groups and 46 actors, students from Faculty of Performing Arts and working actors from local city theatres. An a priori *G-power* analysis based on the effect (Cohen's  $d = 0.92$ ) found in Buckley et al. (2003) revealed an ideal sample size of 51 participants. Due to applying elimination criteria (see Procedure section;  $n_{\text{impact groups}} = 29$  and  $n_{\text{actors}} = 21$ ), and attrition of participants who passed the pre-screening ( $n_{\text{impact groups}} = 17$  and  $n_{\text{actors}} = 7$ ), our final sample included 64 participants (81.2% female). Participants were assigned to: 1) a High scorers group ( $n = 22$ ); 2) a Low scorers group ( $n = 24$ ), or 3) Actors (with low impact scores) ( $n = 18$ ). Participant ages ranged from 18 to 35 years ( $M = 19.83$ ,  $SD = 3.61$  years). Participants from the High scorers and Low scorers groups received two credit points (0.5 for pre-screening and 1.5 for the experiment) for participation, while Actors received the compensation (£10) only after finishing the whole study.

In order for actors to be eligible to participate, they needed to practice acting, with or without official training; 22.2% of the participants identified as professional actors, 33.3% identified as amateur actors (i.e., had received acting training but work in other professions),

161 while 44.5% did not have any official training, but were currently practising acting. On  
162 average, Actors had been practising acting for two years (range = 1-108 months).

### 163 **Measures**

164 **Aversive Events Lists.** By combining items from the Inventory of College Students'  
165 Recent Life Experience (Kohn, Lafreniere, & Gurevich, 1990), the Negative Event (hassle)  
166 Scale for Middle Aged Adults (Maybery, 2013), and the shorter form of the List of Recent  
167 Experiences (Henderson et al., 1981), we created a list of 18 aversive events (e.g., end of  
168 intimate relationship, death of family member/close friend; see Supplemental table 1). The  
169 events were selected to fit prevalent experiences among university students (Kohn et al.,  
170 1990). Participants also had the opportunity to add other aversive events if not already listed.  
171 Participants were asked to choose/add the most aversive event from the list that had happened  
172 to them in the previous six months.

173 **Impact of Event Scale – Revised (IES-R;** Weiss, 2004; *Cronbach's alpha* current  
174 study = .93). This 22-item scale assesses reactions to aversive events. Participants were asked  
175 to report about their feelings towards a selected event in the previous seven days. The IES-R  
176 includes three subscales: Intrusion, Avoidance, and Hyperarousal. Typical item is "Any  
177 reminder brought back feelings about it". The responses are given using 5-point Likert scale  
178 (anchors: 0 = Not at all; 4 = Extremely), and mean scores are calculated for each subscale.

179 **PTSD Checklist for DSM 5 (PCL-5;** Weathers et al., 2013; *Cronbach's alpha* current  
180 study = .92). The PCL-5 includes 20 symptom items and participants are asked whether they  
181 experienced any of these symptoms in the previous month. The PCL-5 correlates highly with  
182 the IES-R (Creamer, Bell, & Failla, 2003). Responses are given on 5-point Likert scale  
183 (anchors: 0 = Never; 4 = Extremely). Typical item is "Feeling very upset when something  
184 reminded you of the stressful experience". We excluded participants who scored 3 ("often")  
185 or higher on any of the 20 items. In total, we excluded four people.

186           **Modified Stroop Task (MST)**. This task was created using an E-prime application,  
187 version 2.0.10.353 (see *pstnet.com*). The words (font Calibri, 80 points) in different colors  
188 were presented on a computer screen (41.1 by 40.2 cm)<sup>1</sup>, 1000 ms after a fixation cross had  
189 appeared in the center of the screen, and participants had unlimited time to provide a  
190 response. The response was given by clicking on a particular letter on the keyboard that  
191 corresponded with one of three word colors (key “D” was painted in red, “H” in green, and  
192 “L” in cyan) on the screen. The reaction time was measured in ms (measurement error = 1  
193 msec). The task included three types of words: Impact-related (e.g., Nightmare, Cry, Threat;  
194 Buckley et al., 2002; Moradi et al., 1999); Feigning-related (e.g., Liar, Fake, Scam; Cannon,  
195 2003), and Neutral words (e.g., Chair, Wall, Pencil; Becker, Rinck, Margraf, & Roth, 2001).  
196 The three word categories did not differ in their average length ( $F(2, 35) = .63, p = .54$ ). Each  
197 word was presented three times, in a different, randomized order, and a different color.  
198 Participants were instructed to react as fast as possible to the colors of the word and to ignore  
199 its content. In total, there were 108 trials (12 words x 3 word groups x 3 colors). Prior to the  
200 experimental trials, participants were presented with 15 (5 words x 3 colors) practice trials  
201 with neutral words (e.g., Belt, Map, Bottle). To test for MST effects, we subtracted latency for  
202 neutral words from latency for Impact-related words and latency for neutral words from that  
203 of feigning-related words.

204           **Self-Report Symptom Inventory (SRSI)**; Merten et al., 2016). The SRSI includes two  
205 superordinate scales: One that lists genuine, plausible symptoms (e.g., depression, PTSD,  
206 anxiety, pain; *Cronbach's alpha* = .92) and one that pertains to pseudosymptoms (*Cronbach's*  
207 *alpha* = .84). An example of a genuine symptom is: “In the mornings, I wake up earlier than  
208 usual”. An example of a pseudosymptom is; “I can't remember what happened to me, but I

---

<sup>1</sup>The distance between the participants' eyes and the screen was approximately 55 cm. The visual angle was about 15 degrees. Ink colors are standard RGB 24-bit Eprime/Windows colors. Precisely, red: RGB values are 255, 0, 0; green: RBG = 0, 128, 0; cyan: RGB = 0, 255, 255.

209 constantly dream about it”. For each symptom, participants indicated whether or not they  
210 suffer from it (False/True) and the total number of genuine and pseudosymptoms are then  
211 summed (ranges 0-50). To identify feigners, Merten et al. (2016) recommended a cutoff of 9  
212 pseudosymptoms. At this cut point, sensitivity is .89 and specificity is .81.

213 **Creative Experience Scale** (CEQ; Merckelbach, Horselenberg, & Muris, 2001;  
214 *Cronbach's alpha* = .76). The CEQ contains 25 dichotomous (i.e., Yes/No) items assessing  
215 the extent to which a person is fantasy prone. Typical items is: “I am never bored because I  
216 start fantasizing when things get boring”. “Yes” answers are summed, with higher scores  
217 indicating higher levels of fantasy proneness.

### 218 **Procedure**

219 The study included two parts: A pre-screening phase and the main experiment. In both  
220 parts, the participants were asked to provide informed consent. The pre-screening was  
221 conducted online, using the online platform Qualtrics. Participants were initially informed that  
222 the focus of the study was about processing of emotional information. Participants were asked  
223 to select or add one event that had happened to them in previous six months from the  
224 Aversive Events List, after which they completed the IES –R and the PCL-5.

225 **Exclusion of Participants.** In order to adhere to the ethical restrictions, participants  
226 were *excluded* from further participation if: 1) the event they reported was traumatic and there  
227 was a concern that including them in the study would contribute to their distress (e.g.,  
228 “suicidal ideas”; “severe illness”; “jumped on and beaten while going back home at night”;  $n$   
229 = 3); 2) their score on PCL-5 score was above three (“often”;  $n$  = 3); and 3) if their scores on  
230 the IES-R subscales (Intrusion; Avoidance, and Hyperarousal) were not in accordance with  
231 the group assignment criteria (either the scores on all three subscales above 1.5 or below 1.5;  
232  $n$  = 43). Participants were not informed of the exact reason for not including them in the  
233 second phase of the study. However, information about different resources and sources of

234 information about mental health concerns were supplied to all participants (see Supplemental  
235 Table 2).

236         **Group Assignment.** Participants who passed the pre-screening were assigned either to  
237 High Scorers group, Low Scorers group, or Actors group. Participants were assigned to the  
238 High Scorers if they had a mean score above 1.5 on the total IES-R scale, but also on its three  
239 separate subscales (total range from 0 to 4). Low Scorers and Actors needed to have the  
240 scores below the 1.5. This score reflects the cutoff point with the best diagnostic accuracy of  
241 PTSD cases (Creamer et al., 2003). The Low scorers and Actors needed to have a PCL-5  
242 mean score under 1.5, while High Scorers should have not exceed 3 on this scale. Testing was  
243 performed in groups of up to 4-5 people. All participants were first given the CEQ. High and  
244 Low Scorers then received the MST and the SRSI. Actors, after filling out the CEQ, were  
245 asked to watch a video of people talking about the consequences of their traumatic experience  
246 (<https://www.youtube.com/watch?v=PFW4hYsYF>). Then, they received a vignette,  
247 presenting a case of Alex, who went through a rough break-up with a threatening and abusive  
248 partner. Alex files an official complaint and wants the court to formulate a restraining order  
249 for the ex-partner. Still, Alex needs to convince the authorities that s/he is under a high  
250 psychological impact (i.e., traumatization) from the experiences with his/her ex-partner.  
251 Actors were asked to imagine that they were Alex and that the experiment was their official  
252 assessment. With this in mind, they performed the MST and filled out the SRSI. Finally, all  
253 participants received an exit questionnaire (e.g., “How convincing/educative/stressful was the  
254 vignette/video?” etc.) to which they could respond on a 5-point scale (1 = Not at all; 5 =  
255 Extremely). Actors were asked whether they used a particular strategy to feign high  
256 psychological impact and if so to describe it. At the end of the session, all participants  
257 received the debriefing form and were invited to ask questions.

258 **Statistical analysis**

259 Our data were analysed using the two-way mixed model Analyses of Variance  
 260 (ANOVA), Multivariate Analysis of Variance (*MANOVA*), *t*-test and *pair t*-tests. Partial eta  
 261 squared ( $\eta_p^2$ ) and *Cohen's ds* were used for the effect size.

## 262 Results

### 263 Pre-screening Measures

264 The three groups did not differ in the frequency with which they selected certain  
 265 events from the Aversive Events List (*Fisher-Freeman Halton* test = 27.4,  $p = .21$ ; see  
 266 Supplemental table 1).

267 We ran a Multivariate Analysis of Variance (*MANOVA*) with five pre-screening  
 268 scores (*IES-R* total scale and subscales, and the *PLC-5* scale) as dependent variables and  
 269 groups (High scorers, Low scorers, and Actors) as independent variables (Table 1).

270 Bonferroni post-hoc tests indicated significant differences between High and Low scorers (all  
 271  $ps < .01$ ), and between High scorers and Actors (all  $ps < .01$ ). There were no significant  
 272 differences between Low scorers and Actors (all  $ps > .05$ ).

Table 1.

*Pre-screening means across groups on IES-R total and subscales (0-4), and PCL-5 score (0-4).*

	Groups			<i>F</i> (2, 61)	$\eta_p^2$
	High Scorers	Low Scorers	Actors		
	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )	<i>M</i> ( <i>SD</i> )		
<i>IES-R</i>	2.05 (.30)	.53 (.28)	.75 (.38)	143.40*	.82
• Intrusion	2.29 (.58)	.60 (.36)	.83 (.35)	81.30*	.75
• Avoidance	2.13 (.64)	.58 (.38)	.81 (.45)	62.22*	.67
• Hyperarousal	1.63 (.54)	.35 (.34)	.48 (.38)	61.00*	.65
<i>PLC-5</i>	1.61 (.69)	.43 (.37)	.65 (.43)	32.18*	.51

*Note:* \*  $p < .01$ ; Bonferroni post-hoc tests indicated significant differences between High and Low scorers ( $ps < .01$ ) and between High scorers and Actors (all  $ps < .01$ ) on all types of measures. However, Low scorers and Actors did not differ significantly (all  $ps > .05$ ) on any of the measures.

273

274 **Exit questions**

275           Actors rated the video as moderately educational ( $M = 3.84, SD = .71$ ), highly helpful  
276 ( $M = 4.00, SD = .69$ ), and as somewhat stressful ( $M = 2.22, SD = 1.35$ ). The vignette was  
277 rated as highly understandable ( $M = 4.17, SD = .92$ ), convincing ( $M = 4.22, SD = .87$ ), and  
278 somewhat stressful ( $M = 2.50, SD = 1.04$ ). Overall, Actors reported high motivation to present  
279 themselves as Alex ( $M = 4.22, SD = 1.06$ ). 61% of them reported having a strategy in order to  
280 perform better on the task, which was mainly evoking their own memories or memories of  
281 their close ones going through a similar experience to the one presented in the vignette. Also,  
282 they tried to visualize the situation and to analyze the emotional response that should follow.

283 **Fantasy proneness**

284           We ran a one-way Analysis of Variance (ANOVA) to test differences between three  
285 groups in their fantasy proneness scores. High scorers ( $M = 11.63; SD = 4.58$ ), Low scorers  
286 ( $M = 8.25, SD = 3.39$ ), and Actors ( $M = 13.28, SD = 4.00$ ) had significantly different scores  
287 on the CEQ ( $F(2, 61) = 8.80, p < .001; \eta_p^2 = .22$ ). Bonferroni follow-up tests indicated  
288 significant differences between the High and Low scorers ( $p = .02$ ), and between the Low  
289 scorers and Actors ( $p < .001$ ). High scorers and Actors did not significantly differ ( $p = .60$ ).

290 **Modified Stroop Task**

291           A two-way mixed ANOVA was calculated, with Groups as a between-subjects factor  
292 (High scorers vs. Low scorers vs. Actors), Word type as a within-subject factor (Neutral vs.  
293 Impact-related vs. Feigning-related), and reaction time as dependent variable. Our results  
294 showed a significant effect of Group,  $F(2, 61) = 13.58, p < .001, \eta_p^2 = .30$ . However, there  
295 was no significant effect of Word type,  $F(2, 122) = 1.02, p = .316, \eta_p^2 = .02$ , and no  
296 significant interaction between the Groups and Word type,  $F(4, 120) = 1.69, p = .157, \eta_p^2$   
297  $= .05$ . Looking into the Bonferroni post-hoc tests, the High scorers and Low scorers did not

298 significantly differ from each other ( $p > .05$ ), whereas Actors scored significantly higher than  
 299 both High ( $p < .001$ ) and Low scorers ( $p < .001$ ) on all three word types. Means and standard  
 300 deviations are given in Table 2<sup>2</sup>.

301 We also ran a MANOVA with Impact-related and Feigning-related *MST effect* as  
 302 dependent variables (Impact-related words RT – Neutral words RT and Feigning-related  
 303 words RT – Neutral words RT) and group as an independent variable. Overall, groups did not  
 304 significantly differ with respect to these interference scores ( $\lambda = .90$ ,  $F(2, 61) = 1.69$ ,  $p = .16$ ,  
 305  $\eta_p^2 = .05$ ). Yet, Actors were the only group with a positive scores which indicated overall  
 306 longer response latencies on Impact-related and Feigning-related words compared with the  
 307 neutral words (see Table 2)<sup>3</sup>.

Table 2.

*Groups' means and standard deviations for reaction times (RT in ms) of groups on the MST word types and the Interference scores.*

Word type	Group		
	High Scorers	Low Scorers	Actors
<i>n</i>	22	24	18
Neutral words	587.74 (101.30)	588.99 (97.84)	788.10 (222.83)
Impact-related words	574.24 (93.97)	586.15 (78.11)	833.68 (307.93)
Feigning-related words	570.62 (82.06)	585.16 (84.45)	803.72 (263.84)
<b>Interference score</b>			
Impact-related	-13.50 (35.30)	-2.85 (37.86)	45.58 (132.35)
Feigning-related	-17.12 (46.73)	-3.83 (34.57)	15.61 (82.72)

<sup>2</sup> The number of errors during the MST task ranged from 0 to 13 ( $M = 3.03$ ,  $SD = 2.54$ ) across participants. The errors in the Low scorers condition ranged from 0 to 13 ( $M = 3.38$ ,  $SD = 2.96$ ), and from 0 to 10 ( $M = 3.27$ ,  $SD = 2.59$ ) in the High scorers condition. The errors in the Actors group ranged from 0 to 5 ( $M = 2.28$ ,  $SD = 1.71$ ).

<sup>3</sup> After close inspection of the data, we noticed that 4 participants (1 from High scorers, 1 from Low scorers, and 2 from Actors group) had significantly higher scores than the rest of the group members. We performed all the analyses without these participants, but the results did not differ ( $\lambda = .95$ ,  $F(2, 61) = .65$ ,  $p = .626$ ,  $\eta_p^2 = .02$ ).



308 Furthermore, we investigated the differences between reaction times to Impact-related,  
 309 Feigning-related, and Neutral words *within* each group, and the differences in response  
 310 latencies did not attain significance in any of the groups. Thus, we did not obtain a significant  
 311 standard MST effect in the High scorers or a significant feigning MST effect in the Actors  
 312 (see Table 3).

Table 3.

*Paired t-test for differences in reaction times (RT) between word categories in groups.*

Word type pairs	Group		
	High scorers ( <i>df</i> = 21)	Low scorers ( <i>df</i> = 23)	Actors ( <i>df</i> = 17)
Impact-related words – Neutral words	-1.80	-.37	1.46
Feigning-related words – Neutral words	-1.72	-.54	.80
Impact-related words – Feigning-related words	.44	.12	1.52

*Note: all  $ps > .05$ ; The difference between Impact-related words and Neutral words refers to the so called standard MST effect; The difference between feigning-related words and Neutral words refers to the so called feigning MST effect.*

313 To test the sensitivity of the MST in detecting over-reporting, we compared the High  
 314 scorers, Low scorers, and the Actors group, and employed for the MST effect a liberal  
 315 criterion of  $\geq 0.10$  ms (positive latency for Impact-related interference score (Impact-related  
 316 RT - Neutral words RT)). Anyone above this cutoff was considered to manifest a standard  
 317 MST effect. A score below  $\leq -0.10$  ms was taken to be indicative of the absence of a standard  
 318 MST effect, which would imply feigning according to Buckley et al. (2003). The correct  
 319 classification for the High scorers and Actors was below the chance level (41% and 34%,  
 320 respectively), and somewhat above it for the Low scorers group (58%; see Table 4).

Table 4.

*Detection rates of MST and SRSI in High Impact, Low Impact and Actors group.*

Group	MST Pass ( $\geq 0.10$ ms)	MST Fail ( $\leq -0.10$ ms)	SRSI Pass ( $\leq 9$ )	SRSI Fail ( $> 9$ )	Correct classification	
					MST	SRSI
High scorers	9 (41%)	13 (59%)	16 (73%)	6 (27%)	41%	73%
Low scorers	10 (41%)	14 (58%)	21 (87.5%)	3 (12.5%)	58%	87.5%
Actors	12 (66%)	6 (34%)	2 (11%)	16 (89%)	34%	89%

*Note:* MST “pass” refers to a positive interference score indicating presence of high impact, “fail” is taken to be indicative of the absence of high psychological impact; SRSI “pass” indicates normal symptom endorsement tendencies, while “fail” is indicative of over-reporting.

321

### 322 Self-Report Symptom Inventory

323 Table 5 shows means and standard deviations of the three groups on the main SRSI  
 324 scales and relevant subscales. A MANOVA showed that overall, groups significantly differed  
 325 on the two SRSI scales ( $\lambda = .66$ ,  $F(2, 61) = 14.88$ ,  $p < .001$ ,  $\eta_p^2 = .33$ ). There was a significant  
 326 difference between groups on both plausible and pseudo-symptom scales of the SRSI ( $F(2,$   
 327  $61) = 45.62$ ,  $p < .001$ ,  $\eta_p^2 = .60$ ;  $F(2, 61) = 46.09$ ,  $p < .001$ ,  $\eta_p^2 = .60$ , respectively).

328 Bonferroni post-hoc tests indicated that Actors endorsed significantly more symptoms than  
 329 High and Low scorers ( $ps < .01$ ), which did not mutually differ in their endorsement of either  
 330 plausible ( $p = .14$ ) or pseudo-symptoms ( $p = .53$ ).

331 Similar was found when we looked into groups’ endorsement of plausible  
 332 PTSD/Anxiety/Depression symptoms and pseudo-PTSD symptoms. Overall, there was a  
 333 significant difference between groups ( $\lambda = .33$ ,  $F(2, 61) = 21.81$ ,  $p < .001$ ,  $\eta_p^2 = .42$ ). Actors  
 334 scored significantly higher on both subscales than the High and Low scorers, ( $ps < .001$ ;  $\eta_p^2$ ’s  
 335  $> .59$ ), which did not significantly differ from each other ( $ps = .50$ ).

Table 5.

*Means and standard deviations of all groups on the Self-Report Symptom Inventory’s plausible and pseudo-symptom scales.*

SRSI	Groups			F test (2, 61)	$\eta_p^2$
	High Impact	Low Impact	Actors		
	M (SD)	M (SD)	M (SD)		
Plausible symptoms	21.00 (9.97)	16.08 (7.06)	39.84 (7.34)	45.62*	.59
Pseudo-symptoms	6.96 (6.45)	3.96 (3.11)	25.00 (11.54)	46.09*	.60
Plausible	4.55 (2.41)	3.71 (1.99)	9.34 (1.54)	43.64*	.59

PTSD/Anx/Dep

Pseudo PTSD/Anx/Dep	1.86 (2.15)	1.04 (1.30)	7.00 (2.50)	51.57*	.63
------------------------	-------------	-------------	-------------	--------	-----

Note: \*  $p < .01$ ; Bonferroni indicated significant differences between Actors and both High and Low scorers ( $p < .01$ ), but no significant differences between the High and Low scorers ( $p > .05$ ).

336 We used the standard cutoff point of nine pseudosymptoms (Merten et al., 2016).

337 Table 4 shows the number and percentages of correct classifications and false positives (i.e.,

338 High scorers and Low scorers misclassified as feigners) and false negatives (i.e., Actors

339 misclassified as High scorers). As can be seen, the SRSI was more effective than the absence

340 of standard MST effect in detecting feigning in the Actors group. Still, 6 participants (27%) in

341 the High scorers group, and 3 (12.5%) from the Low scorers group were misclassified by the

342 SRSI as feigners.

### 343 **Explorative analysis**

344 To explore whether fantasy proneness was related to standard and feigning MST

345 effects and the genuine and pseudosymptoms scales of the SRSI, we calculated *Pearson*

346 *product-moment correlations* including only the High and Low scorers (honest participants,  $n$

347 = 46). The CEQ score did not significantly correlate with the standard MST effect ( $r = -0.14$ ;

348  $p = .35$ ) or the feigning MST effect ( $r = .03$ ;  $p = .84$ ). On the other hand, the CEQ was

349 correlated significantly and positively with the genuine and pseudosymptoms of the SRSI

350 (Pearson  $r$ 's being .58 and .50,  $p$ 's  $< .001$ , respectively).

351 We performed a Welch's  $t$ -test to investigate possible differences in fantasy proneness

352 between participants from the High and Low scorers who were misclassified by SRSI as

353 feigners (24% of the High scorers and 12.5% of the Low scorers,  $n = 9$ ) and the rest of the

354 participants in these groups ( $n = 37$ ). Results indicated significant differences. Participants

355 who were misclassified manifested significantly higher levels of fantasy proneness ( $M =$

356  $13.78, SD = 2.82$ ; *Welch's t-test*( $17.29$ ) =  $4.20, p < .001, r = .45$ ), than participants who were  
357 correctly classified ( $M = 8.92, SD = 4.09$ )<sup>4</sup>.

358

359

## Discussion

360

361

362

363

364

365

We compared the efficacy of the Modified Stroop Task and a symptom over-report questionnaire (SRSI) in detecting feigned high impact experiences. Our results can be summarized as follows: First, in keeping with what others have noted about its fragile nature (e.g., Kimble, Frueh, & Marks, 2009), we did not observe the standard MST effect in our study. That is, the group with High (impact) scores did not manifest longer reaction times for impact-related words than for other categories of words.

366

367

368

369

370

371

372

373

Second, in line with Constans et al. (2014) who found patients with over-reporting tendencies to exhibit longer response times than PTSD patients without over-reporting style, we observed that Actors exhibited “overall slowing down” on all types of words (impact-related, feigning-related, and neutral). However, Actors did not show the significantly different interference score for Impact-related words (Impact-related words RT – Neutral words RT) compared with High and Low scorers. The lack of differences between groups with regard to the impact-related MST effect contradicts the assumption of its high sensitivity to real symptomatology and its utility in the detection of malingering (Buckley et al., 2003).

374

375

376

377

378

379

Third, Actors failed to exhibit also a feigning MST effect. If we consider preoccupation with a certain topic (i.e., cognitive load, see Vrij et al., 2008), as a cause of the feigning MST effect (Cannon, 2003; Mathews & MacLeod, 1985), one could argue that the absence of this effect in our study has to do with our reliance on actors. Perhaps, then, Actors were not triggered by words such as “liar” or “fake”, precisely because they did not interpret their participation in our study as feigning but rather as role-playing. However, there is no

---

<sup>4</sup> The data and analysis are available on Open Science Framework (<https://osf.io/v3uhq/>).

380 reason to assume that in real life, dedicated feigners do not understand their feigning as role-  
381 playing that is justified by the circumstances.

382 Fourth, we followed the suggestion of Buckley et al. (2003) and employed cutoff  
383 scores for differences in reaction times between word categories to identify feigners. After  
384 applying a liberal cutoff point to the MST data, the “hits” for both High scorers and Actor  
385 were below chance level, and slightly above for the Low scorers group. Overall, these  
386 findings provide no evidence to justify the use of the MST as a diagnostic tool to detect  
387 feigning.

388 Fifth, Actors selected significantly more genuine and pseudosymptoms of the SRSI,  
389 compared with participants in the High and Low (impact) scorers, which did not differ from  
390 each other. These findings are well in line with previously established response patterns  
391 among over-reporters (Merten et al., 2016). Most importantly, the SRSI showed higher  
392 sensitivity to over-reporting than the MST. Scores on the SRSI detected 89% of Actors as  
393 over-reporters, which is even better detection rate than in our previous study (77%; Boskovic  
394 et al., 2018).

395 Sixth, more than a quarter (27%) of the High scorers and 12.5% of the Low scorers  
396 were classified by the SRSI as over-reporters. These might be false positives (meaning that  
397 participants were wrongly categorized as over-reporters), but taking into account the  
398 vulnerability of the IES to exaggeration (McGuire, 2002), one could also argue that these  
399 participants actually over-reported their current impact. The latter might be more possible  
400 considering the similar scores of High and Low scorers on SRSI genuine symptom scale and  
401 PTSD/anxiety/depression subscales. To address this issue, we had a closer look at  
402 participants' fantasy proneness (CEQ) scores, which are related to over-reporting  
403 (Merckelbach, 2004; Peace & Masliuk, 2011). CEQ scores did significantly correlated with  
404 endorsement of both genuine and pseudo SRSI symptoms. Moreover, participants in both

405 impact scores groups who were flagged by the SRSI as over-reporters did manifest  
406 significantly higher levels of fantasy proneness than correctly classified participants. Thus, it  
407 is likely that those participants in the impact scorers groups who exhibited elevated levels of  
408 fantasy proneness, engaged in over-reporting of their symptoms.

409 A few limitations of the current study warrant comment. Some of the limitations were  
410 pre-set by the study design, whereas others reflect the unresolved issues surrounding the  
411 MST. First, due to the external ethical restrictions, we used a *subclinical* student sample and  
412 our results may not be generalizable to people who present with full-blown PTSD symptoms,  
413 although studies have found that college sample might be adequate when investigating  
414 impacts of aversive life events (Smyth, Hockemeyer, Heron, Winderlich, & Pennebaker,  
415 2008). Second, participants were pre-screened based on their self-report without being asked  
416 to provide any collateral information, which could have led to inaccurate group assignment of  
417 participants to groups. Third, taking into consideration the variability of the events  
418 participants reported, we could not select words for the MST that were uniquely associated  
419 with every possible aversive event (e.g., fight with a friend or a family member dying).  
420 Therefore, it is possible that some of the words that we included (taken from Buckley et al.,  
421 2003 and Moradi et al., 1999) did not trigger all participants' current impact. Fourth, our pre-  
422 screening results indicated that actors might present a specific population when it comes to  
423 aversive life experiences. While around 31% of participants reported either a traumatic  
424 experience and/or high-intensity PTSD symptoms, this prevalence reached 45% among  
425 actors. Taking into consideration that actors show higher levels of engagement in fantasy than  
426 non-actors (Merckelbach et al., 2001), it is possible that they tend to distort their complaints  
427 by exaggerating the aversiveness of an event. Another possibility is that this population is  
428 especially vulnerable to high impact experiences and psychological distress (e.g., Elal &  
429 Slade, 2005; Thomson, Keehn, & Gumpel, 2009).

430 In sum then, neither the absence of a standard MST effect, nor a general longer  
431 response latency during the MST reliably distinguished between people who had experienced  
432 a relatively high impact event from those who fabricated such impact. Despite the limitations  
433 of our study (e.g., sub-clinical sample, broad range of stimulus words), our results, along with  
434 those reported earlier (Boskovic et al., 2018), indicate that the MST does not allow accurate  
435 detection of feigned symptoms, making it unsuitable as a diagnostic tool. Arguably, this is  
436 related to the MST lacking a clear and articulated protocol for its design, use, and  
437 interpretation. On a positive note, although not perfect, the SRSI might be a promising  
438 alternative for differentiating between honest and feigned symptom presentation.

439

440 **References**

- 441 Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon:  
442 a generic slowdown, not a stroop effect. *Journal of experimental psychology:*  
443 *General, 133*(3), 323.
- 444 Beck, J. G., Freeman, J. B., Shipherd, J. C., Hamblen, J. L., & Lackner, J. M. (2001).  
445 Specificity of Stroop interference in patients with pain and PTSD. *Journal of Abnormal*  
446 *Psychology, 110*, 536-543.
- 447 Becker, E. S., Rinck, M., Margraf, J., & Roth, W. T. (2001). The emotional Stroop effect in  
448 anxiety disorders: General emotionality or disorder specificity? *Journal of Anxiety*  
449 *Disorders, 15*, 147-159.
- 450 Boskovic, I., Biermans, A., Merten, T., Jelcic, M., Hope, L., & Merckelbach, H. (2018). The  
451 Modified Stroop Task Is Susceptible To Feigning: Stroop Performance and Symptom  
452 Over-endorsement in Feigned Test Anxiety. *Frontiers in Psychology, 9*, 1195. doi:  
453 10.3389/fpsyg.2018.01195.
- 454 Buckley, T. C., Blanchard, E. B., & Hickling, E. J. (2002). Automatic and strategic processing  
455 of threat stimuli: A comparison between PTSD, panic disorder, and nonanxiety controls.  
456 *Cognitive Therapy and Research, 26*, 97-115.
- 457 Buckley, T. C., Blanchard, E. B., & Neill, W. T. (2000). Information processing and PTSD: A  
458 review of the empirical literature. *Clinical Psychology Review, 20*, 1041-1065.
- 459 Buckley, T. C., Galovski, T., Blanchard, E. B., & Hickling, E. J. (2003). Is the emotional  
460 Stroop paradigm sensitive to malingering? A between-groups study with professional  
461 actors and actual trauma survivors. *Journal of Traumatic Stress, 16*, 59-66.



- 462 Cannon, B. J. (2003). An Emotional Stroop Effect to Malingering Related Words. *Perceptual*  
463 *and Motor Skills*, 96, 827-834.
- 464 Constans, J. I., Kimbrell, T. A., Nanney, J. T., Marx, B. P., Jegley, S., & Pyne, J. M. (2014).  
465 Over-reporting bias and the modified Stroop effect in Operation Enduring and Iraqi  
466 Freedom veterans with and without PTSD. *Journal of Abnormal Psychology*, 123, 81-  
467 90.
- 468 Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the impact of event  
469 scale—revised. *Behaviour Research and Therapy*, 41, 1489-1496.
- 470 Elal, G., & Slade, P. (2005). Traumatic exposure severity scale (TESS): A measure of  
471 exposure to major disasters. *Journal of Traumatic Stress*, 18, 213–220.
- 472 Elsesser, K., Sartory, G., & Tackenberg, A. (2004). Attention, heart rate, and startle response  
473 during exposure to trauma-relevant pictures: a comparison of recent trauma victims and  
474 patients with posttraumatic stress disorder. *Journal of Abnormal Psychology*, 113, 289-  
475 301.
- 476 Henderson, A. S., Byrne, D. G. and Duncan-Jones, P. (1981). *Neurosis and the social*  
477 *environment*. New York: Academic Press.
- 478 Kaspi, S. P., McNally, R. J., & Amir, N. (1995). Cognitive processing of emotional  
479 information in posttraumatic stress disorder. *Cognitive Therapy and Research*, 19, 433-  
480 444.
- 481 Kimble, M. O., Frueh, B. C., & Marks, L. (2009). Does the modified Stroop effect exist in  
482 PTSD? Evidence from dissertation abstracts and the peer reviewed literature. *Journal of*  
483 *Anxiety Disorders*, 23, 650-655.

- 484 Kohn, P. M., Lafreniere, K., & Gurevich, M. (1990). The inventory of college students' recent  
485 life experiences: A decontaminated hassles scale for a special population. *Journal of*  
486 *Behavioral Medicine, 13*, 619-630.
- 487 MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative  
488 review. *Psychological bulletin, 109*, 163-203.
- 489 MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders.  
490 *Journal of Abnormal Psychology, 95*, 15-20.
- 491 Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states.  
492 *Behaviour Research and Therapy, 23*, 563-569.
- 493 Maybery, D. J. (2013) . *Negative Event (hassle) Scale for Middle Aged Adults (frequency and*  
494 *severity)*. Measurement Instrument Database for the Social Science. Retrieved from  
495 [www.midss.ie](http://www.midss.ie)
- 496 McGuire, B. E. (2002). Malingered post-traumatic stress symptoms on the Impact of Event  
497 Scale. *Legal and Criminological Psychology, 7*, 165-171.
- 498 McNally, R. J., English, G. E., & Lipke, H. J. (1993). Assessment of intrusive cognition in  
499 PTSD: Use of the modified Stroop paradigm. *Journal of Traumatic Stress, 6*, 33-41.
- 500 Merckelbach, H. (2004). Telling a good story: fantasy proneness and the quality of fabricated  
501 memories. *Personality and Individual Differences, 37*, 1371-1382.
- 502 Merckelbach, H., Horselenberg, R., & Muris, P. (2001). The Creative Experiences  
503 Questionnaire (CEQ): A brief self-report measure of fantasy proneness. *Personality and*  
504 *Individual differences, 31*, 987-995.

- 505 Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom  
506 Inventory (SRSI): A new instrument for the assessment of distorted symptom  
507 endorsement. *Psychological Injury and Law*, 9, 102-111.
- 508 Moradi, A., Taghavi, M., Doost, N. H., Yule, W., Dagleish, T. (1999). Performance of  
509 children and adolescents with PTSD on the Stroop colour-naming task. *Psychological*  
510 *Medicine*, 29, 415-419.
- 511 Peace, K. A., & Masliuk, K. A. (2011). Do motivations for malingering matter? Symptoms of  
512 malingered PTSD as a function of motivation and trauma type. *Psychological Injury*  
513 *and Law*, 4, 44-55.
- 514 Shipherd, J. C., & Salters-Pedneault, K. (2008). Attention, memory, intrusive thoughts, and  
515 acceptance in PTSD: an update on the empirical literature for clinicians. *Cognitive and*  
516 *Behavioral Practice*, 15, 349-363.
- 517 Smyth, J. M., Hockemeyer, J. R., Heron, K. E., Wonderlich, S. A., & Pennebaker, J. W.  
518 (2008). Prevalence, type, disclosure, and severity of adverse life events in college  
519 students. *Journal of American College Health*, 57, 69-76.
- 520 Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental*  
521 *Psychology*, 18, 643-662
- 522 Thomas, T. A., & Fremouw, W. J. (2009). The Stroop interference effect and spontaneous  
523 recall in true and malingered motor-vehicle accident-related PTSD. *The Journal of*  
524 *Forensic Psychiatry & Psychology*, 20, 936-949.
- 525 Thomson, P., Keehn, E. B., & Gumpel, T. P. (2009). Generators and interpreters in a  
526 performing arts population: Dissociation, trauma, fantasy proneness, and affective  
527 states. *Creativity Research Journal*, 21, 72-91.

- 528 Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing  
529 cognitive load to facilitate lie detection: The benefit of recalling an event in reverse  
530 order. *Law and Human Behavior*, 32, 253-265.
- 531 Weathers, F.W., Litz, B.T., Keane, T.M., Palmieri, P.A., Marx, B.P., & Schnurr, P.P. (2013).  
532 *The PTSD Checklist for DSM-5 (PCL-5)*. Scale available from the National Center for  
533 PTSD at [www.ptsd.va.gov](http://www.ptsd.va.gov).
- 534 Weiss, D. S. (2004). The Impact of Event Scale-Revised. In J. P. Wilson, & T. M. Keane  
535 (Eds.), *Assessing psychological trauma and PTSD: A practitioner's handbook* (2<sup>nd</sup> ed.,  
536 pp. 168-189). New York: Guilford Press.
- 537 Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and  
538 psychopathology. *Psychological Bulletin*, 120, 3.
- 539 Wilson, S. C., & Barber, T. X. (1983). Fantasy-prone personality: implications for  
540 understanding imagery, hypnosis, and parapsychological phenomena. In A. A. Sheikh  
541 (Ed.), *Imagery: current theory, research, and application* (pp. 340-387). New York:  
542 Wiley.

543

**Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task**

**Highlights:**

- The Modified Stroop Effect could not indicate feigned/genuine responses
- Feigners endorse more symptoms on the SRSI than truth tellers
- The SRSI outperforms the MSE in detecting feigned aversive experiences

**Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task**

**Compliance with Ethical Standards**

*Ethical approval:* All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

*Conflict of Interest:* The authors declare that they have no conflict of interest.

*Informed Consent:* Informed consent was obtained from all individual participants included in the study.