Journal of Behavior Therapy and Experimental Psychiatry 65 (2019) 101483

Contents lists available at ScienceDirect

Journal of Behavior Therapy and Experimental Psychiatry

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

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

Irena Boskovic[a,b,*], Lorraine Hope[b], James Ost[b,1], Robin Orthey[a,b], Harald Merckelbach[a]

[a] Maastricht University, the Netherlands
[b] University of Portsmouth, UK

ARTICLE INFO

Keywords:
Modified stroop task
Feigning
SRSI
Simulators

ABSTRACT

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

Method: We tested whether the MST effect is robust against feigning attempts and compared its absence as an index of feigning with over-reporting tendencies on a symptom questionnaire (i.e., the Self-Reported Symptom Inventory (SRSI)). We included participants (n = 22) who i) reported current high impact of aversive experiences (High scorers), ii) reported current low impact (Low scorers) of aversive experiences (n = 24), and iii) actors (n = 18) with low impact, but instructed to feign current high psychological impact of aversive life events (Simulators). We administered the MST, including impact-related, neutral, and feigning-related words, and the SRSI.

Results: We found no MST effect for impact-related words in the high scorers group, or for feigning-related words in the simulators. Relative to high scorers and low scorers, simulators exhibited significantly longer RTs on all types of words and they also endorsed significantly more bogus symptoms on the SRSI. Thus, the SRSI was a more sensitive measure of feigning than the absence of an MST effect.

Limitation: Some limitations are related to our reliance on a sub-clinical student sample, whereas others reflect the unresolved issues surrounding the MST. Thus, the generalizability of our results is uncertain.

Conclusion: Our findings add to the doubts on the idea that the MST can be used to differentiate between genuine and feigned complaints.

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 (incongruent)-RT (congruent)] (see MacLeod, 1991). The Modified Stroop task (MST), also known as the Emotional Stroop task (e.g., Williams, Mathews, & MacLeod, 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

* Corresponding author. Mastricht University, Forensic Psychology Section, P.O. Box 616, 6200, MD, Maastricht, the Netherlands.
E-mail address: irena.boskovic@maastrichtuniversity.nl (I. Boskovic).

[1] Our dear colleague and friend, James Ost, recently passed away. He will be greatly missed.

https://doi.org/10.1016/j.jbtep.2019.101483
Received 21 September 2018; Received in revised form 10 May 2019; Accepted 12 May 2019
Available online 16 May 2019
0005-7916/

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

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

Three considerations a priori cast doubt on the idea that the absence of an MST effect is diagnostic of feigned PTSD or any other disorder. First, the idea presupposes that MST effects in psychopathological groups – e.g., people with PTSD – are robust, but evidence suggest they are not. For example, an earlier study of Buckley, Blanchard, and Hickling (2002) found that a PTSD group, compared with a panic disorder group and controls, did *not* show the specific MST effect but rather an overall longer reaction time on all three categories (i.e., neutral, panic-related, and PTSD-related) of words (Buckley et al., 2002). In other research, the MST effect in PTSD patients has been difficult to reproduce (e.g., Shipherd & Salters-Pedneault, 2008) and a number of unpublished dissertations have reported small or no MST effects in PTSD patients (e.g., Kimble, Frueh, & Marks, 2009).

Second, and related to the first point, the extant literature is unclear about the reaction time pattern that should be characteristic for feigners: absence of a specific MST effect or general slowing of reaction times? Constans et al. (2014) examined the MST performance of war veterans with and without PTSD and with and without a tendency to over-report symptoms (which conceptually comes close to feigning). Patients with over-reporting tendencies had overall significantly longer response times than PTSD patients without over-reporting style tendencies or controls. These results raise the question as to what an overall delayed response pattern indicates. Does it reflect the cognitive load (Vrij, Fisher, Mann, & Leal, 2008) of feigners who are pre-occupied with how to fabricate their symptoms in a convincing way? If that is the case, general (i.e, non-specific) slowing of responses rather than the specific absence of an MST effect might be diagnostic of feigning. These mixed findings raise a more fundamental question: What does the Modified Stroop task actually measure? Is it the attentional bias or emotional arousal associated with certain words (MacLeod, Mathews, & Tata, 1986), or a preoccupation with a specific topic (Cannon, 2003; Mathews & MacLeod, 1985)? For example, Cannon (2003) found an MST effect for feigning-related words (e.g., lie, fake) among students who were instructed to feign mild brain trauma. This suggests that the MST effect merely reflects participants' current concerns – whether they pertain to their disorder or deception relating to that disorder. However, Thomas and Fremouw (2009) failed to find the MST effect for feigning-related words in a group with PTSD, a group without PTSD, and instructed feigners, which suggests that a specific delay of reaction times even for feigning words in those who are instructed to feign is unreliable. Some authors argue that the traditional SE and the MST effect rely on different mechanisms and that, while the SE actually captures the attentional bias, the MST effect is exhibited due to the threat-driven generic slowdown (Algom et al., 2004).

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

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

1. Method

1.1. 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; *nimpact groups* = 29 and *nactors* = 21), and attrition of participants who passed the pre-screening (*nimpact groups* = 17 and *nactors* = 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), while 44.5% did not have any official training, but were currently practising acting. On average, Actors had been practising acting for two years (range = 1–108 months).

1.2. Measures

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

Impact of Event Scale – Revised (IES-R; Weiss, 2004; *Cronbach's alpha* current study = 0.93). This 22-item scale assesses reactions to aversive events. Participants were asked to report about their feelings towards a selected event in the previous seven days. The IES-R includes three subscales: Intrusion, Avoidance, and Hyperarousal. Typical item is "Any reminder brought back feelings about it". The responses are given using 5-point Likert scale (anchors: 0 = Not at all; 4 = Extremely), and mean scores are calculated for each subscale.

PTSD Checklist for DSM 5 (PCL-5; Weathers et al., 2013; *Cronbach's alpha* current study = 0.92). The PCL-5 includes 20 symptom items and participants are asked whether they experienced any of these symptoms in the previous month. The PCL-5 correlates highly with the IES-R (Creamer, Bell, & Failla, 2003). Responses are given on 5-point Likert scale (anchors: 0 = Never; 4 = Extremely). Typical item is "Feeling very upset when something reminded you of the stressful experience". We excluded participants who scored 3 ("often") or higher on any of the 20 items. In total, we excluded four people.

Modified Stroop Task (MST**)**. This task was created using an E-prime application, version 2.0.10.353 (see *pstnet. com*). The words (font Calibri, 80 points) in different colors were presented on a computer screen (41.1 by 40.2 cm),[2] 1000 ms after a fixation cross had appeared in the center of the screen, and participants had unlimited time to provide a response. The response was given by clicking on a particular letter on the keyboard that corresponded with one of three word colors (key "D" was painted in red, "H" in green, and "L" in cyan) on the screen. The reaction time was measured in ms (measurement error = 1 msec). The task included three types of words: Impact-related (e.g., Nightmare, Cry, Threat; Buckley et al., 2002; Moradi, Taghavi, Doost, Yule, & Dalgleish, 1999); Feigning-related (e.g., Liar, Fake, Scam; Cannon, 2003), and Neutral words (e.g., Chair, Wall, Pencil; Becker, Rinck, Margraf, & Roth, 2001). The three word categories did not differ in their average length (F (2, 35) = 0.63, p = .54). Each word was presented three times, in a different, randomized order, and a different color. Participants were instructed to react as fast as possible to the colors of the word and to ignore its content. In total, there were 108 trials (12 words × 3 word groups x 3 colors). Prior to the experimental trials, participants were presented with 15 (5 words × 3 colors) practice trials with neutral words (e.g., Belt, Map, Bottle). To test for MST effects, we subtracted latency for neutral words from latency for Impact-related words and latency for neutral words from that of feigning-related words.

[2] The distance between the participants' eyes and the screen was approximately 55 cm. The visual angle was about 15°. 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.

Self-Report Symptom Inventory (SRSI; Merten et al., 2016). The SRSI includes two superordinate scales: One that lists genuine, plausible symptoms (e.g., depression, PTSD, anxiety, pain; *Cronbach's alpha* = .92) and one that pertains to pseudosymptoms (*Cronbach's alpha* = .84). An example of a genuine symptom is: "In the mornings, I wake up earlier than usual". An example of a pseudosymptom is; "I can't remember what happened to me, but I constantly dream about it". For each symptom, participants indicated whether or not they suffer from it (False/True) and the total number of genuine and pseudosymptoms are then summed (ranges 0–50). To identify feigners, Merten et al. (2016) recommended a cutoff of 9 pseudosymptoms. At this cut point, sensitivity is .89 and specificity is .81.

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

1.3. Procedure

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

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

Group Assignment. Participants who passed the pre-screening were assigned either to High Scorers group, Low Scorers group, or Actors group. Participants were assigned to the High Scorers if they had a mean score above 1.5 on the total IES-R scale, but also on its three separate subscales (total range from 0 to 4). Low Scorers and Actors needed to have the scores below the 1.5. This score reflects the cutoff point with the best diagnostic accuracy of PTSD cases (Creamer et al., 2003). The Low scorers and Actors needed to have a PCL-5 mean score under 1.5, while High Scorers should have not exceed 3 on this scale. Testing was performed in groups of up to 4–5 people. All participants were first given the CEQ. High and Low Scorers then received the MST and the SRSI. Actors, after filling out the CEQ, were asked to watch a video of people talking about the consequences of their traumatic experience (https://www.youtube.com/watch?v=PFW4hYsYF). Then, they received a vignette, presenting a case of Alex, who went through a rough break-up with a threatening and abusive partner. Alex files an official complaint and wants the court to formulate a restraining order for the ex-partner. Still, Alex needs to convince the authorities that s/he is under a high psychological impact (i.e., traumatization) from the experiences with his/her ex-partner. Actors were asked to imagine that they were Alex and that the experiment was their official assessment. With this in mind, they performed the MST and filled out the SRSI. Finally, all participants received an exit questionnaire (e.g., "How

Table 1
Pre-screening means across groups on IES-R total and subscales (0–4), and PCL-5 score (0–4).

	Groups			F (2, 61)	η_p^2
	High Scorers	Low Scorers	Actors		
	M (SD)	*M (SD)*	*M (SD)*		
IES-R	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
PLC-5	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.

convincing/educative/stressful was the vignette/video?" etc.) to which they could respond on a 5-point scale (1 = Not at all; 5 = Extremely). Actors were asked whether they used a particular strategy to feign high psychological impact and if so to describe it. At the end of the session, all participants received the debriefing form and were invited to ask questions.

1.4. Statistical analysis

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

2. Results

2.1. Pre-screening measures

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

We ran a Multivariate Analysis of Variance (MANOVA) with five pre-screening scores (IES-R total scale and subscales, and the PLC-5 scale) as dependent variables and groups (High scorers, Low scorers, and Actors) as independent variables (Table 1). Bonferroni post-hoc tests indicated significant differences between High and Low scorers (all *ps* $< .01$), and between High scorers and Actors (all *ps* $< .01$). There were no significant differences between Low scorers and Actors (all *ps* $> .05$).

2.2. Exit questions

Actors rated the video as moderately educational ($M = 3.84$, $SD = 0.71$), highly helpful ($M = 4.00$, $SD = 0.69$), and as somewhat stressful ($M = 2.22$, $SD = 1.35$). The vignette was rated as highly understandable ($M = 4.17$, $SD = 0.92$), convincing ($M = 4.22$, $SD = 0.87$), and somewhat stressful ($M = 2.50$, $SD = 1.04$). Overall, Actors reported high motivation to present themselves as Alex ($M = 4.22$, $SD = 1.06$). 61% of them reported having a strategy in order to perform better on the task, which was mainly evoking their own memories or memories of their close ones going through a similar experience to the one presented in the vignette. Also, they tried to visualize the situation and to analyze the emotional response that should follow.

2.3. Fantasy proneness

We ran a one-way Analysis of Variance (ANOVA) to test differences between three groups in their fantasy proneness scores. High scorers ($M = 11.63$; $SD = 4.58$), Low scorers ($M = 8.25$, $SD = 3.39$), and Actors ($M = 13.28$, $SD = 4.00$) had significantly different scores on the CEQ (F (2, 61) = 8.80, $p < .001$; $\eta_p^2 = 0.22$). Bonferroni follow-up tests indicated significant differences between the High and Low scorers ($p = .02$), and between the Low scorers and Actors ($p < .001$). High scorers and Actors did not significantly differ ($p = .60$).

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
n	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)
Interference score			
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)

2.4. Modified Stroop Task

A two-way mixed ANOVA was calculated, with Groups as a between-subjects factor (High scorers vs. Low scorers vs. Actors), Word type as a within-subject factor (Neutral vs. Impact-related vs. Feigning-related), and reaction time as dependent variable. Our results showed a significant effect of Group, F (2, 61) = 13.58, $p < .001$, $\eta_p^2 = 0.30$. However, there was no significant effect of Word type, F (2, 122) = 1.02, $p = .316$, $\eta_p^2 = 0.02$, and no significant interaction between the Groups and Word type, F (4, 120) = 1.69, $p = .157$, $\eta_p^2 = 0.05$. Looking into the Bonferroni post-hoc tests, the High scorers and Low scorers did not significantly differ from each other (ps $> .05$), whereas Actors scored significantly higher than both High (p $< .001$) and Low scorers (p $< .001$) on all three word types. Means and standard deviations are given in Table 2[3]

We also ran a MANOVA with Impact-related and Feigning-related *MST effect* as dependent variables (Impact-related words RT – Neutral words RT and Feigning-related words RT – Neutral words RT) and group as an independent variable. Overall, groups did not significantly differ with respect to these interference scores ($\lambda = .90$, F (2, 61) = 1.69, $p = .16$, $\eta_p^2 = 0.05$). Yet, Actors were the only group with a positive scores which indicated overall longer response latencies on Impact-related and Feigning-related words compared with the neutral words (see Table 2).[4]

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

To test the sensitivity of the MST in detecting over-reporting, we compared the High scorers, Low scorers, and the Actors group, and

[3] 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 ($M = 2.28$, $SD = 1.71$).

[4] 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) = 0.65, $p = .626$, $\eta_p^2 = 0.02$).

Table 3
Paired t-test for differences in reaction times (RT) between word categories in groups.

Word type pairs	Group		
	High scorers (df = 21)	Low scorers (df = 23)	Actors (df = 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.

Table 4
Detection rates of MST and SRSI in High Impact, Low Impact and Actors group.

Group	MST Pass (≥0.10 ms)	MST Fail (≤ −0.10 ms)	SRSI Pass (≤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.

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)	η_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 PTSD/Anx/Dep	4.55 (2.41)	3.71 (1.99)	9.34 (1.54)	43.64*	.59
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).

employed for the MST effect a liberal criterion of ≥0.10 ms (positive latency for Impact-related interference score (Impact-related RT - Neutral words RT)). Anyone above this cutoff was considered to manifest a standard MST effect. A score below ≤ - 0.10 ms was taken to be indicative of the absence of a standard MST effect, which would imply feigning according to Buckley et al. (2003). The correct classification for the High scorers and Actors was below the chance level (41% and 34%, respectively), and somewhat above it for the Low scorers group (58%; see Table 4).

2.5. Self-Report Symptom Inventory

Table 5 shows means and standard deviations of the three groups on the main SRSI scales and relevant subscales. A MANOVA showed that overall, groups significantly differed on the two SRSI scales (λ = 0.66, F (2, 61) = 14.88, p < .001, η_p^2 = 0.33). There was a significant difference between groups on both plausible and pseudo-symptom scales of the SRSI (F (2, 61) = 45.62, p < .001, η_p^2 = 0.60; F (2, 61) = 46.09, p < .001, η_p^2 = 0.60, respectively). Bonferroni post-hoc tests indicated that Actors endorsed significantly more symptoms than High and Low scorers (ps < .01), which did not mutually differ in their endorsement of either plausible (p = .14) or pseudo-symptoms (p = .53).

Similar was found when we looked into groups' endorsement of plausible PTSD/Anxiety/Depression symptoms and pseudo-PTSD symptoms. Overall, there was a significant difference between groups (λ = 0.33, F (2, 61) = 21.81, p < .001, η_p^2 = 0.42). Actors scored significantly higher on both subscales than the High and Low scorers, (ps < .001; η_p^2's > .59), which did not significantly differ from each other (ps = .50).

We used the standard cutoff point of nine pseudosymptoms (Merten et al., 2016). Table 4 shows the number and percentages of correct classifications and false positives (i.e., High scorers and Low scorers misclassified as feigners) and false negatives (i.e., Actors misclassified as High scorers). As can be seen, the SRSI was more effective than the absence of standard MST effect in detecting feigning in the Actors group. Still, 6 participants (27%) in the High scorers group, and 3 (12.5%) from the Low scorers group were misclassified by the SRSI as feigners.

2.6. Explorative analysis

To explore whether fantasy proneness was related to standard and feigning MST effects and the genuine and pseudosymptoms scales of the SRSI, we calculated *Pearson product-moment correlations* including only the High and Low scorers (honest participants, n = 46). The CEQ score did not significantly correlate with the standard MST effect (r = - 0.14; p = .35) or the feigning MST effect (r = 0.03; p = .84). On the other hand, the CEQ was correlated significantly and positively with the genuine and pseudosymptoms of the SRSI (Pearson r's being .58 and .50, p's < .001, respectively).

We performed a Welch's t-test to investigate possible differences in fantasy proneness between participants from the High and Low scorers who were misclassified by SRSI as feigners (24% of the High scorers and 12.5% of the Low scorers, n = 9) and the rest of the participants in these groups (n = 37). Results indicated significant differences. Participants who were misclassified manifested significantly higher

levels of fantasy proneness ($M = 13.78$, $SD = 2.82$; Welch's t-test (17.29) = 4.20, $p < .001$, $r = 0.45$), than participants who were correctly classified ($M = 8.92$, $SD = 4.09$).[5]

3. Discussion

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 et al., 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.

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).

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

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

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

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

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

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

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.

Conflicts 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.

Acknowledgement

This work was supported by the House of Legal Psychology / Erasmus Mundus Joint Doctorate Program. Framework Partnership

[5] The data and analysis are available on Open Science Framework (https://osf.io/v3uhq/).

Agreement (FPA) 2013-0036. Cohort 2015 with Specific Grant Agreement (SGA) 2015-1610.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbtep.2019.101483.

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional stroop phenomenon: A generic slowdown, not a stroop effect. *Journal of Experimental Psychology: General, 133*(3), 323.

Becker, E. S., Rinck, M., Margraf, J., & Roth, W. T. (2001). The emotional Stroop effect in anxiety disorders: General emotionality or disorder specificity? *Journal of Anxiety Disorders, 15*, 147–159.

Beck, J. G., Freeman, J. B., Shipherd, J. C., Hamblen, J. L., & Lackner, J. M. (2001). Specificity of Stroop interference in patients with pain and PTSD. *Journal of Abnormal Psychology, 110*, 536–543.

Boskovic, I., Biermans, A., Merten, T., Jelicic, M., Hope, L., & Merckelbach, H. (2018). The modified stroop task is susceptible to feigning: Stroop performance and symptom over-endorsement in feigned test anxiety. *Frontiers in Psychology, 9*, 1195. https://doi.org/10.3389/fpsyg.2018.01195.

Buckley, T. C., Blanchard, E. B., & Hickling, E. J. (2002). Automatic and strategic processing of threat stimuli: A comparison between PTSD, panic disorder, and non-anxiety controls. *Cognitive Therapy and Research, 26*, 97–115.

Buckley, T. C., Blanchard, E. B., & Neill, W. T. (2000). Information processing and PTSD: A review of the empirical literature. *Clinical Psychology Review, 20*, 1041–1065.

Buckley, T. C., Galovski, T., Blanchard, E. B., & Hickling, E. J. (2003). Is the emotional stroop paradigm sensitive to malingering? A between-groups study with professional actors and actual trauma survivors. *Journal of Traumatic Stress, 16*, 59–66.

Cannon, B. J. (2003). An emotional stroop effect to malingering related words. *Perceptual & Motor Skills, 96*, 827–834.

Constans, J. I., Kimbrell, T. A., Nanney, J. T., Marx, B. P., Jegley, S., & Pyne, J. M. (2014). Over-reporting bias and the modified Stroop effect in Operation Enduring and Iraqi Freedom veterans with and without PTSD. *Journal of Abnormal Psychology, 123*, 81–90.

Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the impact of event scale—revised. *Behaviour Research and Therapy, 41*, 1489–1496.

Elal, G., & Slade, P. (2005). Traumatic exposure severity scale (TESS): A measure of exposure to major disasters. *Journal of Traumatic Stress, 18*, 213–220.

Elsesser, K., Sartory, G., & Tackenberg, A. (2004). Attention, heart rate, and startle response during exposure to trauma-relevant pictures: A comparison of recent trauma victims and patients with posttraumatic stress disorder. *Journal of Abnormal Psychology, 113*, 289–301.

Henderson, A. S., Byrne, D. G., & Duncan-Jones, P. (1981). *Neurosis and the social environment.* New York: Academic Press.

Kaspi, S. P., McNally, R. J., & Amir, N. (1995). Cognitive processing of emotional information in posttraumatic stress disorder. *Cognitive Therapy and Research, 19*, 433–444.

Kimble, M. O., Frueh, B. C., & Marks, L. (2009). Does the modified stroop effect exist in PTSD? Evidence from dissertation abstracts and the peer reviewed literature. *Journal of Anxiety Disorders, 23*, 650–655.

Kohn, P. M., Lafreniere, K., & Gurevich, M. (1990). The inventory of college students' recent life experiences: A decontaminated hassles scale for a special population. *Journal of Behavioral Medicine, 13*, 619–630.

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*, 15–20.

Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy, 23*, 563–569.

Maybery, D. J. (2013). Negative event (hassle) scale for Middle aged Adults (frequency and severity). Measurement instrument database for the social science. Retrieved from www.midss.ie.

McGuire, B. E. (2002). Malingered post-traumatic stress symptoms on the Impact of Event Scale. *Legal and Criminological Psychology, 7*, 165–171.

McNally, R. J., English, G. E., & Lipke, H. J. (1993). Assessment of intrusive cognition in PTSD: Use of the modified Stroop paradigm. *Journal of Traumatic Stress, 6*, 33–41.

Merckelbach, H. (2004). Telling a good story: Fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences, 37*, 1371–1382.

Merckelbach, H., Horselenberg, R., & Muris, P. (2001). The creative experiences questionnaire (CEQ): A brief self-report measure of fantasy proneness. *Personality and Individual Differences, 31*, 987–995.

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.

Moradi, A., Taghavi, M., Doost, N. H., Yule, W., & Dalgleish, T. (1999). Performance of children and adolescents with PTSD on the Stroop colour-naming task. *Psychological Medicine, 29*, 415–419.

Peace, K. A., & Masliuk, K. A. (2011). Do motivations for malingering matter? Symptoms of malingered PTSD as a function of motivation and trauma type. *Psychological Injury and Law, 4*, 44–55.

Shipherd, J. C., & Salters-Pedneault, K. (2008). Attention, memory, intrusive thoughts, and acceptance in PTSD: An update on the empirical literature for clinicians. *Cognitive and Behavioral Practice, 15*, 349–363.

Smyth, J. M., Hockemeyer, J. R., Heron, K. E., Wonderlich, S. A., & Pennebaker, J. W. (2008). Prevalence, type, disclosure, and severity of adverse life events in college students. *Journal of American College Health, 57*, 69–76.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643–662.

Thomas, T. A., & Fremouw, W. J. (2009). The Stroop interference effect and spontaneous recall in true and malingered motor-vehicle accident-related PTSD. *Journal of Forensic Psychiatry and Psychology, 20*, 936–949.

Thomson, P., Keehn, E. B., & Gumpel, T. P. (2009). Generators and interpretors in a performing arts population: Dissociation, trauma, fantasy proneness, and affective states. *Creativity Research Journal, 21*, 72–91.

Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior, 32*, 253–265.

Weathers, F. W., Litz, B. T., Keane, T. M., Palmieri, P. A., Marx, B. P., & Schnurr, P. P. (2013). The PTSD checklist for DSM-5 (PCL-5). Scale available from: the National Center for PTSD at www.ptsd.va.gov.

Weiss, D. S. (2004). The impact of event scale-revised. In J. P. Wilson, & T. M. Keane (Eds.). *Assessing psychological trauma and PTSD: A practitioner's handbook* (pp. 168–189). (2nd ed.). New York: Guilford Press.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*, 3.

Wilson, S. C., & Barber, T. X. (1983). Fantasy-prone personality: Implications for understanding imagery, hypnosis, and parapsychological phenomena. In A. A. Sheikh (Ed.). *Imagery: Current theory, research, and application* (pp. 340–387). New York: Wiley.