

A multi-method approach to the detection of fabricated symptoms

Citation for published version (APA):

Boskovic, I. (2019). *A multi-method approach to the detection of fabricated symptoms*. [Doctoral Thesis, Maastricht University, University of Portsmouth]. Maastricht University.
<https://doi.org/10.26481/dis.20190129ib>

Document status and date:

Published: 01/01/2019

DOI:

[10.26481/dis.20190129ib](https://doi.org/10.26481/dis.20190129ib)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

A Multi-Method Approach to the Detection of Fabricated Symptoms

Irena Boskovic

A Multi-Method Approach to the Detection of Fabricated Symptoms

Dissertation

To obtain the degree of Doctor of Philosophy from
The University of Portsmouth and the degree of Doctor at Maastricht University, on the
authority of Rector Magnificus Prof.dr. Rianne M. Letschert according to the decision
of the Board of Deans
to be defended in public on
Tuesday 29th of January 2019 at 14.00 hours

By

IRENA BOSKOVIC



Supervisors

Prof. dr. H.L.G.J. Merckelbach

Prof. dr. M. Jelicic

Prof. dr. L. Hope

Co-supervisors:

Dr. J. Ost

Doctoral Thesis Committee

Prof. dr. P.E.H.M. Muris, Maastricht University, the Netherlands (Chair)

Dr. L. Akehurst, University of Portsmouth, UK

Prof. M. Kopelman, King's College London, UK

Prof. dr. P. van Koppen, Vrije Universiteit Amsterdam & Maastricht University, the Netherlands

The research presented in this dissertation was conducted as a Doctor of Philosophy Program with the House of Legal Psychology in collaboration between the University of Portsmouth and Maastricht University.

CONTENTS

CHAPTER 1

| | |
|--------------------------------------------------------------|----------|
| GENERAL INTRODUCTION..... | 9 |
| Introduction..... | 11 |
| Defining and Determining Malingering..... | 14 |
| DSM on Malingering and Other Misconceptions..... | 16 |
| Detection of Malingering..... | 18 |
| Verbal Credibility Assessment..... | 18 |
| Task Performance..... | 21 |
| Symptom Validity Tests..... | 22 |
| Cross-Cultural Influence on Symptom Validity Assessment..... | 24 |
| Aim and Outline of the Present Thesis..... | 26 |

CHAPTER 2

| | |
|---------------------------------------------------------------|-----------|
| THE VERIFIABILITY APPROACH TO FABRICATED SYMPTOMS..... | 30 |
| Introduction | 32 |
| Study 1..... | 34 |
| Method..... | 34 |
| Results..... | 37 |
| Study 2..... | 38 |
| Method..... | 38 |
| Results..... | 39 |
| Discussion..... | 41 |
| Study 3: Verifiability on the Run: An Experimental Study..... | 38 |
| Introduction: | 46 |
| Method..... | 49 |
| Results..... | 51 |
| Discussion..... | 54 |

CHAPTER 3

| | |
|---------------------------------------------------------------------------------------------------|-----------|
| THE MODIFIED STROOP TASK AND THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED SYMPTOMS..... | 57 |
| THE CASE OF TEST ANXIETY | 58 |
| Introduction..... | 60 |
| Methods | 62 |
| Results..... | 65 |
| Discussion..... | 67 |
| THE CASE OF HIGH IMPACT EXPERIENCES..... | 70 |
| Introduction..... | 72 |
| Method..... | 74 |
| Results..... | 78 |
| Discussion..... | 83 |

CHAPTER 4

| | |
|----------------------------------------------------------------------|-----------|
| THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED SYMPTOMS..... | 87 |
| Introduction..... | 89 |
| Method..... | 92 |
| Results..... | 94 |
| Discussion..... | 98 |

CHAPTER 5

| | |
|----------------------------------------------------------------------------------------------------------|------------|
| THE VERIFIABILITY APPROACH AND THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED TRAUMA CLAIMS..... | 103 |
| Introduction..... | 105 |
| Method..... | 109 |
| Results..... | 112 |
| Discussion..... | 116 |

CHAPTER 6

| | |
|------------------------------------------------------------------------------------|-----|
| THE CROSS-CULTURAL INFLUENCE ON PRACTITIONERS' PLAUSIBILITY JUDGMENTS | 119 |
| Introduction..... | 121 |
| Method..... | 123 |
| Results..... | 125 |
| Discussion..... | 127 |

CHAPTER 7

| | |
|-------------------------------------------------------------|-----|
| GENERAL DISCUSSION | 131 |
| Main Findings of This Thesis..... | 134 |
| Limitations..... | 138 |
| Future Avenues in Symptom Validity Assessment Research..... | 141 |
| Practical Implications..... | 143 |
| Conclusion..... | 145 |
| REFERENCES | 147 |
| SUMMARY | 167 |
| ACKNOWLEDGMENTS | 171 |
| DISSEMINATION | 173 |
| CURRICULUM VITAE | 177 |
| VALORIZATION ADDENDUM | 179 |
| APPENDICES | 181 |
| APPENDIX 1 | 183 |
| APPENDIX 2..... | 187 |
| APPENDIX 3..... | 191 |
| APPENDIX 4..... | 193 |
| APPENDIX 5..... | 195 |

CHAPTER 1

GENERAL INTRODUCTION

This chapter draws on the following papers:

Boskovic, I., & Merckelbach, H. (2018). Fake Posttraumatic Stress Disorder (PTSD) costs real money. *The Inquisitive Mind Magazine*, 4/2018. Retrieved from: <http://www.in-mind.org/article/fake-posttraumatic-stress-disorder-ptsd-costs-real-money>

Jelicic, M., Merckelbach, H, & Boskovic, I. (2017). Seven myths about feigning. In H. Otgaar & M.L. Howe (Eds.). *Finding the truth in the courtroom: Handling deception, lies, and memories*. Oxford University Press.

Introduction

In 2015, M.L. (33), known as “Action Man Mark”, climbed Kilimanjaro, won two triathlons, and skied in the Swiss Alps. He accomplished all of these feats while receiving disability benefits for a lower back injury incurred while serving in Afghanistan, which he described as causing him “excruciating pain” and prevented him from walking more than 50 meters. Following his recent, very public physical feats, M.L. was accused of fraud. His attorney dismissed the accusations claiming that M.L., as an ex-serviceman, was simply able to “push through the pain”¹. Needless to say, M.L. was found guilty of fraud (The Telegraph, 2017).

The example presented above might mislead one into believing that fraudsters will sooner or later be exposed. Unfortunately, that is not the case. Fraudsters who put effort into their “disability” act, and keep a low profile, may go undetected for years, if they are detected at all (Guriel & Fremouw, 2003). Due to the subjective nature of mental problems, fraudsters usually prefer to feign psychological rather than physical symptoms (Resnick, 2008). The estimated prevalence of fabricated physical symptoms in forensic, litigation, or compensation settings ranges between 20% and 40% of all filed physical complaints (Bass & Halligan, 2007; Greve, Ord, Bianchini, & Curtis, 2009; Mittenberg, Patton, Canyock, & Condit, 2002; see also Young, 2015). The numbers are even more alarming in cases of psychiatric disorders reports, such as combat-related traumatisation, where the fabrication occurs in approximately 50% of cases (Resnick, 2008; Freeman, Powell, & Kimbrell, 2008; Lees-Haley, 1997). Studies presented in Table 1.1 illustrate the extent of symptom fabrication, defined as a failure on symptom validity tests, for a variety of health complaints, and how the estimates differ depending on samples, detection measures, and settings.

Because there must be a proportion of individuals who successfully fabricate their reports, the exact prevalence of symptom fabrication remains unknown (Resnick, 2008). While undetected, those who fabricate symptoms exhaust financial resources, and misuse legal exemptions or treatments aimed at helping people who genuinely suffer as a result of trauma or injury. For instance, if 25% of all disability benefit-receiving adults in the general public or among veterans were fabricating symptoms, the cost of such fraud would be approximately \$20 billion in US alone (Chafetz & Underhill, 2013; Department of Veterans Affairs Office of the Inspector General, 2005; Resnick, West, & Payne, 2008). In the UK, the estimated annual cost of benefit fraud is around £2 billion (Kitchen, 2003).

¹ Link: <https://www.telegraph.co.uk/news/2017/07/20/benefits-cheat-paratrooper-claimed-weak-walk-50m-caught-climbing/>

General Introduction

Table 1.1.

An overview of studies, investigated symptoms, used samples and measures, and prevalence of failure of symptom validity tests (SVT)(%).

| Authors | Symptoms | Sample | Measure | % of SVT failure |
|-------------------------------------------------|----------------------------|----------------------|-----------------------------------------------------|------------------|
| Ardolf, Denney, & Houston, 2007 | Neurocognitive Dysfunction | Criminal defendants | MMPI-2 TOMM Rey word recognition list etc. | 22% |
| Chafetz, 2008 | Cognitive dysfunctions | Disability claimants | TOMM SVS | 46% |
| Freeman, Powell, & Kimbrell, 2008 | PTSD | Vietnam veterans | SIRS M-FAST SIMS | 53% |
| Frueh, Hamner, Cahill, Gold, & Hamlin, 2000 | PTSD | Veterans | MMPI-2 | 30% |
| Greiffenstein & Baker, 2006 | Mild head injury | Compensation seekers | MMPI-2 TOMM Rey word recognition list | 37% |
| Greve, Ord, Bianchini, & Curtis, 2009 | Chronic pain | Compensation seekers | TOMM Word Memory Test | 20-50% |
| Larrabee, 2003 | Mild head injury | Compensation seekers | MMPI-2 MMPI-2 VFD FT WAIS-R etc. | 40% |
| McDermott & Sokolov, 2009 | Psychiatric symptoms | Inmates | SIRS | 66% |
| Mittenberg, Patton, Canyock, & Condit, 2002 | Chronic fatigue | Experts | Experts' frequency judgment | 35% |
| Pollock, Quigley, Worley, & Bashford, 2007 | Psychiatric symptoms | Inmates | MMPI-2 SIRS | 32% |
| Rissmiller, Steer, Friedman, & Demercurio, 1999 | Suicidal ideas | Suicidal inpatients | MMPI-2 | 12% |
| Wierzbicki & Tyson, 2007 | ADHD/Learning disability | College students | WAIS-R etc. | 44% |

In criminal settings, malingerers exploit legal regulations by reporting psychiatric symptoms as mitigating circumstances, which may significantly influence the outcomes of their trials. For instance, in certain jurisdictions in the US, a person suffering from Posttraumatic Stress Disorder (PTSD) may be able to argue for reduced criminal responsibility, leading to a verdict of not guilty by reason of insanity (NGRI), a reduced jail sentence, or psychiatric treatment rather than imprisonment (Berger, McNiel & Binder, 2012; Cohen & Appelbaum, 2016; Friel, White, & Hull, 2008; see also McDermott, Dualan, & Scott, 2013). In fact, PTSD has been shown to be one of conditions that is relatively often malingered for these reasons (see also Rassin, Boskovic, & Merckelbach, 2018; Resnick, et al., 2008; Young, 2016).

It is not just in legal contexts that malingering can have serious consequences. In the clinical setting, for example, one consequence of malingering can be the provision of treatments, including the prescription of psychoactive medication and psychotherapy. Malingerers often seek therapy so as to gather evidence of their ill health. Thus, they report poor therapy outcomes (Anestis, Gottfried, & Joiner, 2015; Van Egmond & Kummeling, 2002) and prolong the treatment until they receive the benefits they aimed to get, after which the majority drop out (Department of Veterans Affairs Office of the Inspector General, 2005). Unnecessary treatments may be detrimental to the health of the malingerer himself and can, for example, lead to addiction or brain changes in case of medications (e.g., Bernardy & Friedman, 2015). In case of exposure therapy for instance, a healthy person might experience unanticipated consequences that may severely influence their functionality.

In the research context, poor progress and negative evaluations of therapy complicate the search of efficient treatments for genuine patients. Furthermore, misreported or exaggerated symptoms have a distorting effect on research outcomes (McGrath, Mitchell, Kim, & Hough, 2010; Merckelbach, Langeland, de Vries, & Draijer, 2014). Hence, certain established patterns or links between physical and psychological phenomena (e.g., hippocampal volume and memory performance; Rienstra et al., 2013) might be overshadowed because some of the patients involved in clinical trials may be feigning a condition.

Although recently more attention has been drawn to the issue of symptom fabrication by both practitioners and researchers (Jelicic, Merckelbach, & Boskovic, 2018), the general public and governmental institutions are reluctant to discuss the magnitude and impact of fabricated symptom reports (e.g., Jackson et al., 2011). This resistance is especially striking in the case of veterans under suspicion of fabricating their complaints (e.g., Poyner, 2010). Questioning the highly emotional statements of people claiming to be victims, or the claims of people seen as heroes who fought for their country, may seem cold-hearted. However, others have argued that evaluating the validity of

such reports is a necessary ethical action of professionals in order to protect people genuinely suffering from various physical and psychological injuries (Heilbronner, 2005; Poyner, 2010).

As shown above, fabrication of symptoms occurs on a non-trivial scale, at least in some settings, and has a serious social and legal impact (Jelicic et al., 2018). Yet, the significance of this phenomenon is still underestimated by many practitioners and researchers (e.g., Jackson et al., 2011). Hence, research evaluating existing, and developing new methods for the detection of symptom fabrication is important and greatly needed. Moreover, sharing the findings of such research and making them available to practitioners working in criminal and clinical settings, may have the strongest impact on the current prevalence of symptom fabrication.

The purpose of this thesis is to explore the utility of a multi-method approach to symptom validity assessment. However, before introducing the methods that could potentially contribute to the quality of symptom validity assessment, it is important to first elaborate on the definition of malingering and other response styles that could distort symptom reports. It is also necessary to evaluate the current official guidelines in the Diagnostic and Statistical Manual (DSM-5) for detecting symptom fabrication.

Defining and Determining Malingering

The term *deception* refers to any intended distortion of a self-report (Bass & Halligan, 2007). In the context of this thesis, the deceptive self-reports under scrutiny are those concerning an individual's mental and physical health symptoms. This form of symptom deception is known as *malingering*, which is defined as the intentional falsification of pathology, either by fully fabricating or exaggerating complaints, driven by a potential external incentive (American Psychiatric Association, 2013). The most frequent incentives behind malingering are financial, such as insurance compensation and disability pensions, legal, such as reduced criminal responsibility, as well as other benefits, for instance, avoiding military duty or obtaining stimulant medication (Jelicic et al., 2018).

Some other health-related deceptive behaviours come conceptually close to malingering, such as factitious disorder and feigning (see Figure 1.). In case of *factitious disorder*, the intentional fabrication of symptoms is more strictly related to internal motivation, prominently taking on a “sick role”, while *feigning* does not specify the underlying type of incentive (Rogers & Bender, 2003). Although this distinction in motivation clearly separates factitious disorder from the externally driven symptom deception, the terms malingering, feigning, and even over-reporting (i.e., exaggeration of pathology) are often used interchangeably (Rogers & Bender, 2018)². Still, in clinical assessment, as long as the presence of clear external incentives has not been established, the term malingering should

² Although some authors (e.g., Rogers, 2018) argue that there are important differences between these constructs (e.g., known vs. not-known motivation), these distinctions are not pertinent to the contexts examined in this thesis and will thus be used interchangeably.

be avoided. What we actually observe in patients is their response style, and this can vary depending on many different factors, not just because of the intentional fabrication of symptoms (Rogers & Bender, 2018). For instance, in a nonclinical sample, those who carelessly respond to a clinical symptom checklist, due to the boredom or tiredness, might inflate the number and characteristics of their complaints (Mayer, Faust, Faust, Baker, & Cook, 2013). Furthermore, some people may exhibit response tendencies, such as acquiescence (i.e., tendency to agree to items) and extreme point responding (Ziegler, 2015). As a result, such individuals may erroneously appear to be clinical cases. In patient samples, personality traits may also have an influence on how symptoms are endorsed. For example, those who have difficulties in identifying and describing their internal state, particularly emotions (i.e., alexithymia; Bagby, Parker & Taylor, 1994), show an inclination towards overestimation of symptoms (Merckelbach, Prins, Boskovic, Niesten, & Campo, 2018). A similar response style was found among individuals with dissociative symptoms, but the exact origin of this link has not yet been clarified (Merckelbach, Boskovic, Pesy, Dalsklev, & Lynn, 2017; Merckelbach, Muris, Horselenberg, & Stougie, 2000). One possible explanation of the overlap between dissociative symptoms and over-reporting might be the vividness of imagination, also known as fantasy proneness (Wilson & Barber, 1983; Merckelbach, Horselenberg, & Muris, 2001). Specifically, it was shown that richness of imagination shares some features with dissociative symptoms (Merckelbach, Campo, Hardy, & Giesbrecht, 2005), but also with the overestimation of complaints (Peace & Masliuk, 2011). Thus, additional factors, such as evaluatee's attentiveness and personality traits, should also be considered in the clinical evaluation of symptom reports.

Whereas the response styles discussed so far amount to over-reporting of symptoms (i.e., negative response bias), it is important to note that people can also show positive response bias in assessment. Thus, people might deny their pathology by minimizing their symptoms and exaggerating positive features (i.e., faking good). The reasons might be social desirability (Paulhus, 2011) or supernormality, a form of deception in which a patient wants to present a healthy facade to examiners in order to, for instance, leave the psychiatric institution (Cima et al., 2003). Another reason for a positive response bias might be to secure employment - an extreme positive manner of responding is highly prevalent in personnel selection (Levashina, Weekley, Roulin, Hauck, 2014).

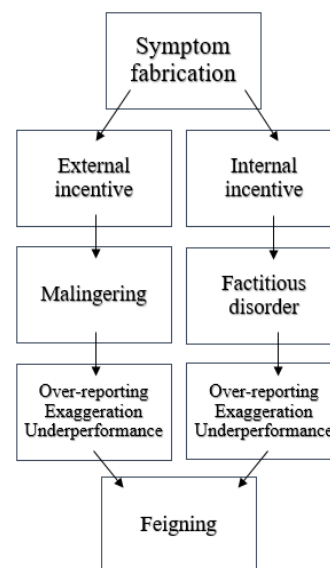


Figure 1. Concepts within the symptom fabrication.

DSM on Malingering and Other Misconceptions

Malingering is not a psychiatric disorder, thus, it cannot be diagnosed. Rather, it should be excluded before determining a psychiatric condition (Tracy & Rix, 2017). The official guidelines in the DSM-5 state that malingering should be suspected if the following are present: i) medico-legal context of assessment; ii) discrepancy between self-reported complaints and objective findings; iii) lack of cooperation of evaluatee, and iv) presence of antisocial personality disorder (DSM-5; American Psychiatric Association, 2013). Although these DSM guidelines might seem straightforward and helpful, they actually misreflect the reality of symptom validity assessment and disregard the research recommendations (Berry & Nelson, 2010). Below, we discuss each of the four DSM indicators of malingering in regard to the current research findings.

One study showed that when the DSM guidelines are used to flag malingering, the correct detection rate of malingerers was only 20% (Rogers, 1990). Foremost, malingering in DSM is presented as a “yes or no” phenomenon, although it is, in fact, dimensional in nature (Rogers, 2008; Slick, Sherman, & Iverson, 1999; Walters et al., 2008). Thus, the evaluatee’s symptom report should be assessed in terms of *likelihood* of malingering (Rosenfeld, Sands, & Van Gorp, 2000; van Impelen, Merckelbach, Jelicic, Niesten, & Campo, 2017). Furthermore, malingering can have different forms ranging from full symptom fabrication, partially over-reported complaints, to misattributing existing symptoms to a false cause (Resnick, 1997). Additionally, simply specifying the medico-legal context of assessment because it entails the presence of incentive is not informative enough. A practitioner should be aware of the varying base rates of malingering depending on different legal context of assessment. Symptom fabrication is more frequent and more difficult to detect in civil (e.g., personal injury cases) than in criminal contexts (Sullivan, Lange, & Dawes, 2007; Merckelbach, Smeets, & Jelicic, 2009; Mittenberg et al., 2002). Moreover, the idea that malingering only occurs within the medico-legal context can lead to overestimation of malingering in this setting and underestimation in others, such as the clinical context (McCarter, Walton, Brooks, & Powell, 2009).

The discrepancy between subjective symptom reports and objective findings, for example a person claiming “excruciating pain” who then publicly climbs Kilimanjaro, as the Action Man Mark did, seems to be a good indicator of deception. In symptom validity assessment, however, there are often no other sources of information available, except evaluatee’s self-reports and practitioner’s subjective judgment. Besides the fact that self-reports of patients are unreliable (e.g., Resnick, 2018), practitioners’ judgments are yet another cause for concern (Zimmerman, 2003). For example, the majority of practitioners believe that deliberate symptom fabrication is uncommon and are therefore reluctant to use tested methods for validating symptom (Allcott et al., 2014; see also Jelicic et al., 2018). Another worrying finding is that only four per cent of practitioners are confident in their ability

to detect malingering (Cohen & Appelbaum, 2016; see Rassin et al., 2018). Moreover, even if “objective” findings are present, one should be aware that military documents, as well as medical reports can be forged (Burkett & Whitley, 1998).

The DSM guidelines suggest that a person who malingers will be uncooperative during the examination. For example, a person may refuse to talk or take tests (Rogers, 2008). However, in the case of war veterans, for instance, it is uncommon that a person is willing to discuss the details of a traumatizing experience (Resnick et al., 2008). Also, research suggests that people telling the truth often have a “transparency illusion”, meaning that they believe that their truthfulness is obvious to others and needs no proof, resulting in brief reports (Savitsky & Gilovich, 2003). In contrast, malingerers present their complaint as a serious handicap by elaborating on their injury in as much detail as possible (Purisch & Sbordone, 1997). Providing detailed reports seems to be a good malingering strategy because of the widespread belief that liars are reluctant to talk (e.g., Rogers, 2008).

The final DSM indicator for malingering, the presence of anti-social traits, has been a subject of discussion for some years. Although some studies reported a connection between anti-social (psychopathic) traits and malingering (e.g., Gacono, Meloy, Sheppard, Speth, & Roske, 1995), others failed to observe such a link (e.g., MacNeil & Holden, 2005). The most recent meta-analysis confirmed a loose connection between malingering and psychopathy. Specifically, Niesten and her colleagues (2015) examined seven studies on this relationship and found a link between malingering and anti-social traits that was, at best, weak (Niesten, Nentjes, Merckelbach, & Bernstein, 2015). Despite the indications that people with anti-social traits are more deceptive (e.g., Hare, 2003), they are not necessarily engaging more in symptom fabrication, nor are they better at it (Van Impelen et al., 2017). Furthermore, anti-social traits and genuine pathology can coexist, meaning that the presence of anti-social traits alone cannot be used as a red flag for dishonest symptom reports (Rogers & Bender, 2018).

Considering the broad and vague nature of the DSM guidelines, researchers have recently argued that the DSM indicators of malingering should be avoided in practice because they may lead to sizeable errors in clinical decision making (Rogers & Schuman, 2005; Rogers & Bender, 2018). The DSM guidelines may not only mislead practitioners, but they may also reinforce misconceptions regarding malingering already held by mental health experts. Practitioners often believe that malingerers cannot play their role for a prolonged period of time (Jelicic et al., 2018) – this leads to the belief that identifying malingerers simply requires time and close observation³. Confessions or

³ Rudolf Hess is one of the examples that disprove this premise. This Nazi official successfully feigned amnesia for almost two years during the Nuremberg trials, before confessing that he had feigned his memory loss (Douglas-Hamilton, 2016; see Jelicic et al., 2018).

strong forensic evidence (e.g., CCTV footage) are rarely available in this field (Rogers, 2008). Furthermore, even when present, the CCTV footage will not necessarily reveal presence of, for instance, reported mental health problems (e.g., trauma, depression)⁴. It is the lack of such evidence that causes practitioners to believe in yet another myth - that malingering rarely occurs (Allcott et al., 2014; e.g., Jackson et al., 2011). Along with the underestimation of the prevalence of malingering, possibly the most dangerous misconception is that malingering is easily detected, and that a clinical interview is sufficient for a practitioner to recognize such deception (Jelicic et al., 2018). Many influential studies showed that interviews are not diagnostically sufficient, with detection success ranging from 0 to 25% (Rosen & Phillips, 2004; see also Jelicic et al., 2018), and that non-verbal cues to deception have no diagnostic value (DePaulo et al., 2003). The danger of this misconception is that it makes practitioners less sensitive towards applying evidence-based methods in symptom validity assessment (e.g., Allcott et al., 2014).

Detection of Malingering

One way of countering these misconceptions is to apply a multi-method approach to the detection of fabricated symptoms. By using different methods combined with collateral information, experts may enhance the quality of their assessment and by doing so decrease the likelihood of errors in their decision making (Weiss & Van, 2017). A multi-method approach involves the simultaneous use of methods that tap into different, but related, aspects of deceptive behaviour, such as i) Verbal Credibility Assessment, ii) Task performance, and iii) Symptom Validity Tests (SVTs). In the next sections, we will elaborate on one specific method from each of the three domains, notably the *Verifiability Approach*, the *Modified Stroop task*, and the *Self-Report Symptom Inventory*, and indicate their possible utility in symptom validity assessment. The reasons behind investigating these three methods particularly were to 1) contribute to the field of symptom validity assessment by introducing and testing a new approach that has not been previously applied in the symptom validity research (the Verifiability Approach); 2) systematically examine an already existing and controversial task (i.e., the Modified Stroop task); and, 3) provide an initial investigation of a newly developed measure for detection of over-reporting (i.e., the Self-Report Symptom Inventory).

Verbal Credibility Assessment. Symptom assessment usually begins with an interview. Many assessment protocols stipulate the clinical interview as an important part of diagnostic decision making. However, the extant literature casts doubt on the utility of unstructured interviews in the context of deception detection (Hauch, Sporer, Michael, & Meissner, 2016; Jelicic et al., 2018). Instead, recent research shows support for alternative methods for examining evaluatees' verbal

⁴ An exception might be videomaterial testifying to the damaging actions of patients with Munchausen or Munchausen by proxy (see Bauer, 2004)

statements that focus on language use and the details they provide in their statements. It is already widely accepted in the lie detection field that evaluating linguistic features of an evaluatee's statement, rather than observing an evaluatee's behaviour, increases the accuracy of veracity judgment (DePaulo et al., 2003; Vrij, 2008). The two most frequently studied methods for coding verbal statements are the Reality Monitoring paradigm (RM; Johnson & Raye, 1981) and Criteria Based Content Analysis (CBCA; Köhnken, 2004; Steller, 1989).

The premise behind the RM paradigm is that information, depending on its external (experienced) or internal (imagined) origin, differs in how it is described. The external experiences include more contextual, sensory, and semantic information, and the imagined ones involve more cognitive details related to imagination and thinking (Masip, Sporer, Garrido, & Herrero, 2007). These differences should help in distinguishing information received from the outside world, and internally generated information (Simons, Garrison, & Johnson, 2017). Generating lies qualifies as an internal process, therefore, it is expected that intentionally deceptive statements would include fewer contextual, sensory or semantic details than truthful statements. The RM approach includes four types of details (i.e., criteria) that should be present in a truthful statement: sensory details, contextual details, semantic, and internal/idiosyncratic details. However, a meta-analysis on the utility of RM in detecting deception did not yield promising findings. This might be due to differing operationalizations of the RM criteria across studies (Masip, Sporer, Garrido, & Herrero, 2005; Oberlander et al., 2016).

The CBCA approach includes 19 different verbal content criteria relating to general characteristics, specific content, peculiarities of contents, motivation-related contents, and offense-specific elements within a statement (Steller & Köhnken, 1989). According to this model, the presence of each CBCA criterion in a statement is indicative of truthfulness (Hauch, Masip, Sporer, & Blandon-Gitlin, 2017). The CBCA was first applied in the credibility assessment of child sexual abuse reports, a context in which it received strong support over the years (Kohnken, 2004). A recent meta-analysis on the CBCA provided overall favourable results for the reliability of its criteria in credibility assessment. However, using the CBCA summary scores alone in decision making is strongly discouraged (Hauch et al., 2017).

Oberlander et al. (2016) showed that RM and CBCA do not differ in their effectiveness. Meanwhile, Bogaard et al. (2013) provided important findings suggesting that both the RM and the CBCA approaches are vulnerable to contextual bias. In other words, providing any domain-irrelevant information to coders significantly misdirects their interpretation of verbal material and scoring of both RM and CBCA criteria. Therefore, the reliability of both methods must be taken with caution (Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2013).

Recently, researchers developed a new verbal credibility assessment method, the Verifiability Approach (VA; Nahari, Vrij, & Fisher, 2014a), in order to facilitate the detection of fabricated reports. Unlike the two previously discussed methods, the use of the VA does not rely on the assumption that the quantity of reported details alone can serve as an indicator of veracity. These authors argue that the quality of reported information, specifically verifiability of details in evaluatees' reports, should be inspected, rather than the amount of provided information *per se* (Nahari, Vrij, & Fisher, 2014). In order to label a detail as verifiable, it needs to describe activities that can potentially be checked, such as documented actions, witnessed or performed with another identifiable person, recorded actions, or actions that include the use of technology (Nahari & Vrij, 2014). The assumption behind the VA is that truth tellers include more information that can be verified, (such as "I am a Ph.D. student at Maastricht University"), whereas liars avoid mentioning such details and opt to compensate with information that will not be possible for an interviewer to verify (e.g., "I study very hard") (Nahari, 2018). The VA has been tested in the assessment of alibis using mock crime paradigms (Nahari & Vrij, 2014) and in insurance claims (Nahari, Leal, Vrij, Warmelink, & Vernham, 2014). Applying the VA to alibi reports led to a high detection rate for liars. However, these findings were not replicated in the insurance claim study (Nahari et al., 2014).

The reason behind this discrepancy in findings might be the different degrees of freedom liars have in generating alibis in mock crime paradigms compared with insurance claims. That is, in mock crime paradigms the interviewer is familiar with the ground truth. Thus, liars do not have many available resources to fabricate information without being caught out. In an insurance claims setting, however, liars have more freedom in creating their accounts due to the lack of interviewer's knowledge about the reported incident (Vrij, Nahari, Isitt, & Leal, 2016). With this in mind, Vrij et al. (2016) included what they termed an "Information Protocol", in which participants were informed that the details of their statements could be subsequently checked by the interviewer. Introducing the Information Protocol actually improved detection rates in the insurance settings (Harvey, Vrij, Nahari & Ludwig, 2017; Nahari, Vrij, & Fisher, 2014b; Vrij, Nahari, Isitt, & Leal, 2016). It appeared that knowing the criteria one needs to fulfil so as to appear truthful increased the differences in reported verifiable details between truth tellers and liars, because liars were unable to provide verifiable information without risking being exposed as deceptive (Nahari, Vrij, & Fisher, 2014b). This shift in findings suggests that the VA, unlike other verbal credibility assessment methods, may be less sensitive to countermeasures (Nahari, 2018). Moreover, as the VA approach requires the coding of two categories of details only (verifiable and non-verifiable), with clear instructions for each, practitioners applying the VA may be less vulnerable to contextual bias, than when using the RM and/or CBCA.

In a pioneering study by Akehurst et al. (2017), researchers combined the RM and the CBCA criteria in order to detect exaggerated statements following the exposure to a cold pressor test. Overall, applying these methods led to higher detection rates of participants who over-reported their complaints than when evaluators employed self-generated criteria of veracity. The most valuable criteria were extracted from the CBCA and involved ‘accounts of own mental state’ and inclusion of ‘unusual details’. However, the exaggerated statements were not found not to include fewer details than truthful ones (Akehurst et al., 2017). These findings counter the notion that fabricated statements will include fewer details as predicted by both the RM and the CBCA method. However, they testify to the need to consider alternatives, specifically, the need to examine the utility of the VA method. The VA seems particularly suitable for checking the veracity of physical symptom, which are often followed by collateral information (e.g., prescriptions, doctor visits, etc.). However, to date there have not been any attempts to test the utility of the VA in symptom reports.

Task Performance. Many researchers argue that, in symptom validity assessment, self-reports should be replaced by more objective methods, like those measuring reaction time (e.g., Mangiulli et al., 2018). The rationale behind this claim is that using tasks that rely on reaction time should be less vulnerable to manipulation than self-report. Hence, they should provide more objective and independent data (Constans et al., 2014; Shipherd & Salters-Pedneault, 2008). One of the most frequently used reaction-time tasks is the Modified Stroop task (MST). During the MST, an evaluatee is presented with disorder-related and neutral words in different colors, and instructed to name the word colors as quickly as possible disregarding the meaning of the word. It is assumed that an evaluatee with a certain disorder will exhibit a prolonged reaction time (RT) once words related to that disorder (e.g., ‘exam’, for test anxiety complaints) are presented, compared to when neutral words (e.g., chair) are displayed. It has been argued that this difference in response latency (i.e., the MST effect) captures attentional bias typical for genuine patients (Williams, Watts, MacLeod, & Mathews, 1997).

The MST effect has been tested in people suffering from panic disorder, obsessive-compulsive disorder, social and other forms of phobias, and PTSD (Becker, Rinck, Margraf, & Roth, 2001). Results suggested that the MST effect is highly specific, meaning that a person suffering from, for example, social phobia will not show the prolonged reaction time when color-naming PTSD-related words, and vice versa (Mathews & Macleod, 1985). Some researchers (e.g., Buckley et al., 2003; Constans et al., 2014) applied the MST in the assessment of malingering of PTSD. These researchers claimed that the use of the MST was an important contribution to the detection of malingerers. Buckley et al. (2003) stated that “...reaction times to Stroop tasks may be useful adjuncts to clinicians who are attempting to discern psychopathology from malingering” (p. 65). However, this interpretation might be overstating the success of the MST in detecting fabricated symptoms by

ignoring the inconsistency of MST results. Specifically, in a previous study, the MST effect did not occur among patients with PTSD (Buckley, Blanchard, & Hickling, 2002). Despite this null finding, the researchers nevertheless claimed that, in the absence of a robust MST effect, even a prolonged reaction time could be indicative of feigning (Buckley et al., 2002).

Constants et al. (2014) administered the MST to war veterans with and without PTSD diagnosis and with and without a tendency to over-report symptoms. Patients with over-reporting tendencies had overall significantly longer response times than PTSD patients without over-reporting style tendencies or controls. These results cast doubt on Buckley et al.'s (2002) statement, and raises the question of what overall prolonged response latencies in fact indicate. One explanation might be related to the cognitive load (Vrij, Fisher, Mann, & Leal, 2008) of feigners who are pre-occupied with how to fabricate their symptoms in a convincing way. The extra cognitive load associated with feigning then leads them to be slower in naming the colors of words related to the "object" of their pre-occupation, such as symptoms or the act of feigning itself. Cannon (2003) found the MST effect among malingerers once the presented words were changed so that their meaning was related to the act of feigning. However, other researchers have not succeeded in replicating this "feigning" MST effect (e.g., Thomas & Fremouw, 2009).

A meta-analysis of findings concerning the utility of the MST effect in PTSD by Kimble and his colleagues (2009) yielded important findings. The researchers included published papers ($n = 18$) as well as (unpublished) doctorate theses ($n = 12$) that administered the MST to patients with PTSD and compared them with non-PTSD samples. They compared the two sources of data and concluded that published papers considerably more often included favourable MST ($\chi^2(2) = 4.35, p < .05$) results than unpublished doctoral theses (Kimble, Frueh, & Marks, 2009; see also Rosenthal, 1979).

Applying the MST to the malingering field might have its benefits, not the least because it is a low-cost and time-saving procedure. However, it also carries the potential risk of erroneously categorizing genuine patients as malingerers (false positives) and vice versa (false negatives). Yet, the MST is still prominent not only in research (e.g., Bielecki, Popiel, Zawadzski, & Sedek, 2017), but also among practitioners. A recent study showed that the Stroop task is among the ten most frequently used methods in clinical practice both in Canada and US (Rabin, Paolillo, & Barr, 2016).

Symptom Validity Tests. In addition to verbal credibility assessment methods and the reaction-time tasks, there are a number of instruments that can help to detect feigning. These instruments are collectively known as Symptom Validity Tests (SVTs; see for review Bianchini, Mathias, & Greve, 2001). The SVTs are a strongly supported method for symptom validity assessment by a majority of practitioners (see Bush, Heilbrunner, & Ruff, 2014; Chafetz et al., 2015; Dandachi-FitzGerald, Ponds, & Merten, 2013; McCarter, Walton, Brooks, & Powell, 2009). The

SVTs can be grouped into two categories: the Performance Validity Tests (PVTs) and Self-Report Validity Tests (SRVTs; Merten, Merckelbach, Giger, & Stevens, 2016). The first category, PVTs, consists of simple memory, perception, or reasoning tasks that are combined with a two-alternative, forced-choice procedure. In the Morel Emotional Numbing task, for example, an evaluatee is presented with pictures of faces depicting emotional expressions (e.g., anger). After each picture, the evaluatee is asked to indicate which of two emotions the depicted person expressed (e.g., anger vs happiness). Even people with serious neurological problems can do this task very well. Just by guessing, a person can obtain a correct score of 50%. Therefore, if an individual performs under chance level, it is highly likely that a person is deliberately underperforming (Morel & Marshman, 2008).

Alternatively, SRVTs are aimed at detecting over-reporting of symptoms by presenting evaluatees with absurd, implausible symptoms that are unlikely to be endorsed by genuine patients. The most frequently used SRVT is the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; Dandachi-FitzGerald, Ponds, & Merten, 2013; Van Impelen, Merckelbach, Jelicic, & Merten, 2014). The SIMS contains 75 bizarre items pertaining to feigned memory impairments, psychosis, and depression. A typical item is “Sometimes when writing a phone number, I notice that the numbers come out backwards even though I don't mean to do it”. People who endorse such claims are exaggerating their condition, therefore, their reports are likely to be invalid (Merten & Merckelbach, 2013). The utility of the SIMS in the correct detection of feigners exceeds 90%, with 7% of false positive classifications (Van Impelen et al., 2014). Still, bizarre items might be obvious to the evaluatees, a shortcoming that diminishes the face validity of SIMS. Furthermore, the SIMS items pertain to amnesia, psychosis, and low intelligence, which occur mostly in criminal but not civil proceedings (e.g., crime-related amnesia).

Taking the flaws of the SIMS into account, a group of researchers recently developed a new SRVT named the Self-Report Symptom Inventory (SRSI; Merten et al., 2016). The SRSI includes both genuine and pseudosymptoms. A typical item from the genuine symptom scale is “I have nightmares about things that happened to me”, whereas a typical example from the pseudosymptom scale is “I can't remember what happened to me, but I constantly dream about it.” Because both types of items are distributed within the questionnaire, it is more difficult for an examinee to realize the real aim of the assessment. Furthermore, the SRSI consists of genuine symptoms related to 1) cognitive complaints, 2) depression, 3) pain, 4) somatic problems, and; 5) anxiety/posttraumatic stress disorder; and pseudosymptoms that tap into 1) cognitive/memory complaints, 2) neurological motor issues, 3) neurological sensory, 4) pain, and 5) anxiety/depression. Thus, each of the two main (genuine symptoms and pseudosymptoms) scales includes five subscales describing the most prevalent complaints within civil medico-legal context (Merten et al., 2016). Because the SRSI includes

genuine symptoms as well, its utility is not only restricted to screening of atypical symptoms, but it can also be administered to screen for actual psychopathology. The authors of SRSI suggest using two cutoff points. A screening cutoff score is more than six endorsed pseudosymptoms, while a standard cutoff point is the endorsement of more than nine pseudosymptoms. With these two scales, an evaluator can also calculate the ratio (number of pseudosymptoms / number of genuine symptoms) and determine the ratio between the genuine claims and exaggerated symptoms. To date, studies investigating the utility of the SRSI in detecting over-reporting have produced favourable results, such as a correct detection rate of over-reporting above 80% (Merten et al., 2016). However, further systematic investigation of the scale is needed.

Cross-Cultural Influence on Symptom Validity Assessment

Although the procedure for symptom validity assessment may seem relatively consistent due to the well-known general guidelines (e.g., Slick et al., 1999), certain factors, such as culture, have to be additionally considered. The majority of tests used in the detection of symptom fabrication have been constructed in the English language, and their validation has been based on Western European and American samples (Correa, 2018; Correa & Rogers, 2010). One has to be aware of differences that may arise in the manifestation and reporting of symptoms, as well in assessing these reports, depending on the cultural background of both evaluatee and evaluator. Studies show that when the cultural background of the evaluator and evaluatee are different, this may undermine the effectiveness of forensic assessment (Weiss & Rosenfeld, 2012). A number of research lines seem to support this observation.

First, the prevalence of feigning might differ across countries, depending on motivation and evaluation contexts. For instance, in countries in which monetary compensation is common, malingering may be more lucrative, thus more frequent (Nijdam-Jones & Rosenfeld, 2017). This might explain the statistics showing that the top five countries with the highest prevalence rate of lifetime PTSD are Canada, the Netherlands, Australia, the United States, and New Zealand, which are the most developed countries and considered to be the least vulnerable to traumatization (Duckers, Alisic, & Brewin, 2016). However, it could also mean that standards for diagnosing PTSD differ across countries. This cross-cultural difference has already been recognized in depression diagnosis (e.g., Kessler & Bromet, 2013). Physical symptoms of depression are considered to be more diagnostically valuable in non-Western countries, whereas psychological ailments that follow depression bring more weight to the diagnosis in Western culture (Kessler & Bromet, 2013). The dominance of physical symptoms among non-Western cultures might reflect the stigma among non-Western countries towards psychological problems (Stefanovics et al., 2016).

Second, differences in culture can lead to different understanding and responding to the test material. According to some researchers, members of “typical western”, individualistic cultures are shown to be overall less prone to respond with “Yes” to items, compared with collectivistic countries (Smith, 2004; Johnson, Kulesa, Cho, & Shavitt, 2005), which are known to adopt more conformist behaviour (Hofstede, 2001). Also, native English speakers show a tendency towards a more extreme response styles than individuals for whom English is a second language (Harzing, 2006). This might also be a specific cultural trait considering that Greeks, Italians, and Spanish lean towards extreme responding more often than English, German, or French responders (Van Herk, Poortinga, & Verhallen, 2004).

Third, cultural backgrounds may not only affect how patients express psychological or medical complaints, but also how healthcare professionals evaluate the plausibility of these complaints (e.g., Kleinman & Cohen, 1997). Although cross-cultural research is scarce regarding discrepancies between health experts, the extant literature indicates that practitioners from different cultural backgrounds have a lower agreement in the evaluation of psychological and neuropsychological symptoms than practitioners from similar cultures (Ferrari, Constantoyannis, & Papadakis, 2001; Giosan, Glovsky, & Haslam, 2001).

Finally, knowing that the most frequently used SRVTs (e.g., the SIMS) consist of atypical symptoms, it is important to know whether what is considered to be an atypical symptom may depend on the cultural background of patients and evaluators (e.g., Weiss & Rosenfeld, 2012). For practitioners in certain cultures, symptoms that are qualified as “rare” or “atypical” elsewhere might be more acceptable or even expected (Charles, Gafni, Whelan, & O’Brien, 2006). These issues directly concern the cross-cultural utility of SVTs. Merten and Rogers (2017, p. 106) observed that “...assuming that any feigning measure is universally applicable across languages and diverse cultures is categorically unacceptable”. However, in their recent meta-analysis on cross-cultural feigning assessment, Nijdam-Jones and Rosenfeld (2017) found that the SIMS had the highest overall classification accuracy, indicating the lowest level of variability across cultures and languages. Unfortunately, no research has examined the cultural background of professionals who make decisions about the plausibility of various symptoms. Thus, the extent to which cultural differences in terms of, for instance, professionals’ interpretation of atypical symptoms may complicate the detection of feigning is an issue that deserves systematic research.

Aim and Outline of the Present Thesis

The main objective of this thesis was to investigate the utility of a multi-method approach to detecting fabricated symptoms. We investigated the merits of three potential routes to detection: verbal credibility, reaction times, and the Self-Report Validity Tests (SRVTs). Specifically, we investigated the methods corresponding to the three domains, the Verifiability Approach (VA), the Modified Stroop task (MST), and the Self-Report Symptom Inventory (SRSI). The VA, as a recently developed lie-detection tool, has not yet been applied in the field of symptom validity assessment. In contrast, the MST has been administered in the detection of symptom fabrication. However, research results to date have been inconsistent. The SRSI is a newly developed tool for detection of over-reporting, thus, its utility is yet to be investigated. Lastly, we also wanted to investigate the influence of practitioners' cultural background on their decision-making concerning the plausibility of certain symptoms. Therefore, the goals of this thesis were to: 1) examine whether applying a new tool, the VA, could assist in the detection of fabricated symptom reports, and if so, to what extent; 2) investigate the contribution of the MST to the symptom validity assessment; 3) initialise the assessment of the SRSI as a symptom validity screening tool, and 4) assess the influence of culture on practitioners' symptom plausibility judgment. The overarching purpose of our research was to evaluate which of these methods would enhance the quality and reliability of the symptom validity assessment the most, and ideally, help practitioners' decision making.

In our studies, we mainly relied on an experimental simulation design with non-clinical subjects. Thus, students were asked to either genuinely report about their (subclinical) symptoms or to fabricate non-existing complaints. This is important to note because, as such, the simulation design cannot fully emulate clinical assessment in a real-world setting (Niesten et al., 2017; Rogers, 2018). Consequently, research with actual patients, especially those exaggerating their symptoms (partial malingers c.f. Resnick, 1997), might yield different findings.

In Chapter 2, we addressed the utility of the VA in reports of malingered physical symptoms. Considering that physical symptoms are often followed by collateral information that could be verified (e.g., prescriptions, doctor visits), we explored the applicability of the VA to the detection of fabricated symptom reports. In Study 1, we asked participants with genuine physical symptoms (truth tellers) and students without any complaints but instructed to feign (malingerers), to elaborate on their symptoms. In Study 2, a similar design was used, except that we also administered the Information Protocol, meaning that we informed our participants beforehand about what kind of details we would check their reports for. The goals of these studies were to examine whether the VA might help detecting feigners' reporting strategies. Additionally, in Study 3, we used the VA to assess participants' symptom reports, after experimentally inducing physical symptoms in one groups of

participants, while instructing two other groups to malingering following different instructions. One group of malingerers was only informed about the task truth tellers had performed and asked to report as if they had been involved in it. The second group was told to use embedded lies, meaning that they should incorporate truthful experiences within their reports. In all three studies, we anticipated that truthful symptom reports would include more verifiable information than malingerers' accounts, and that fabricated reports would mostly incorporate the non-verifiable details.

Moving on from the verbal credibility assessment, in Chapter 3, we compared the utility of a reaction-time task, the MST, with the newly developed SVT, the SRSI, in detecting malingered symptoms. We tested which of the two methods provided more diagnostically valuable information for detecting fabricated symptoms. Specifically, we were interested in whether the MST effect (the prolonged reaction time in color-naming the disorder-related words) would be sensitive to malingering, and if the SRSI scores would reveal malingerers' tendency towards endorsing pseudosymptoms. In Study 4, we investigated the utility of the MST and the SRSI in the detection of malingered test-anxiety claims, by using a within-subject design. In the first session, only participants without higher levels of test anxiety were eligible to perform the MST and to fill out the SRSI. Seven days later (the second session), the same participants were invited, and were given the MST and the SRSI again. However, prior to the second session, they received instructions to malingering suffering from severe test anxiety. In the Study 5, we administered the MST and the SRSI to the three groups of participants: i) a group with a high impact of aversive life experiences, ii) a group with a low impact of such experiences, and iii) actors, without a current history of highly aversive experiences, but who received instructions to malingering the impact of such an event. In both of the studies, we anticipated that malingerers would be able to produce the MST effect, but that their SRSI genuine symptoms and pseudosymptoms scores would be significantly higher than in honest comparison condition or groups.

In Chapter 4, we focused on evaluating the utility of the SRSI as a stand-alone detection method. The SRSI includes subscales that explain various complaints, ranging from neurological, cognitive symptoms to pain and anxiety issues. Therefore, in this study, we evaluated and compared the effectiveness of the SRSI in detecting malingered psychological versus physical symptoms. We applied the SRSI to three groups of participants: i) healthy comparison group 2) healthy groups instructed to malingering anxiety symptoms, and iii) healthy group instructed to fabricate pain complaints. We anticipated that both groups of malingerers were score significantly higher on the SRSI than healthy comparison group, both on genuine symptom scale and on pseudosymptoms.

Next, in Chapter 5, we decided to focus on PTSD-like reports, because the self-reports of traumatic experiences and its consequences are the among the most prevalent, yet the most

challenging to detect complaints in symptom validity assessment. We tested whether applying the VA to the exposure narratives and administering the SRSI for symptom reports could help in revealing malingerers' strategies in reporting PTSD-related complains. One group of participants witnessed an aversive event via Virtual Reality (VR), while the other group was instructed to mangle that experience. Both groups were asked to provide the narrative about the scene they witnessed and to report the symptoms as consequences of such exposure. We anticipated that truth tellers would provide more verifiable information about the exposure scene and endorse less pseudosymptoms on the SRSI than malingerers.

Culture may have a significant influence on practitioners' decisions in symptom validity assessment, especially when it comes to the question whether or not symptoms are exaggerated. The best known SVTs, such as the SIMS, rely on atypical symptoms as a detection strategy of exaggeration. Therefore, it is worth investigating whether practitioners across cultures have a similar view on what is considered as an atypical symptom. In Chapter 6, we tested the cross-cultural stability of implausible symptoms from the perspective of practitioners. Specifically, we included practitioners from Western and non-Western countries, and investigated their plausibility judgments of items included in the SIMS, which we mixed with symptoms included in standard measures of genuine pathology. This way, we wanted to examine whether practitioners will make a distinction between them. We anticipated that Western practitioners will give lower plausibility ratings to SIMS symptoms than to standard symptoms, compared with non-Western professionals.

The main results from Chapters 2-6⁵, together with methodological issues and future directions of research in the field of malingering, are addressed in the general discussion in Chapter 7. The practical implications and contributions of this thesis to the symptom validity assessment literature are also examine.

⁵ The studies within each chapter are presented in article format because they are either published, or currently under revision at scientific journals.

CHAPTER 2

THE VERIFIABILITY APPROACH TO FABRICATED SYMPTOMS

This chapter draws on the following papers:

Boskovic, I., Bogaard, G., Merckelbach, H., Vrij, A., & Hope, L. (2017). The Verifiability Approach to detection of malingered physical symptoms. *Psychology, Crime & Law*, 23, 717-729.

Boskovic, I., Gallardo, C., Vrij, A., Hope, L., & Merckelbach, H. (in press). Verifiability on the Run: An Experimental Study on the Verifiability Approach to Malingered Symptoms. *Psychiatry, Psychology, and Law*.

THE VERIFIABILITY APPROACH TO FABRICATED SYMPTOMS

STUDY 1 & STUDY 2

Abstract

Inspired by recent research showing that liars are reluctant to include verifiable details in their accounts, we explored in two studies ($N = 125$; $N = 105$) whether participants who report fabricated symptoms ('malingerers') present fewer verifiable details than participants who report genuine ill-health symptoms. In Study 1, participants were instructed to describe a typical day on which they had experienced a genuine or malingered symptom. Truth tellers' statements included significantly higher proportions of verifiable details concerning the reported symptoms than malingerers' statements. Compared with truth tellers, malingerers generated longer statements with more unverifiable details. In Study 2, we informed participants that their statements may be assessed for verifiable or checkable details. Malingerers often mentioned 'false' witnesses to provide checkable information and differences between malingerers and truth tellers in statement length, and checkable and uncheckable details were no longer significant. The utility and implications of the Verifiability Approach to detection of malingering are discussed.

Key words: Verifiability Approach, malingering, detection of deception, physical symptoms, symptoms report.

The Verifiability Approach to Detect Malingering of Physical Symptoms

Fabrication of physical symptoms in a medico-legal context burdens the health care system and ultimately may harm the care that genuine patients deserve (Bianchini, Greve, & Glynn, 2005). Thus, it is important to develop tools and strategies that can help in identifying people who fabricate ('malingers') symptoms of ill health. Malingering is defined as the intentional production of false or grossly exaggerated symptoms motivated by external incentives. Incentives may consist of financial rewards gained through personal injury litigation or workers' compensation procedures (McDermott & Feldman, 2007), or reduced criminal responsibility (American Psychiatric Association, 2000, p. 739). It is difficult to determine on what scale malingering occurs, because 'successful' malingerers remain undetected (Resnick, West, & Payne, 2008). However, a conservative estimate is that, for example, 20% of chronic pain patients exaggerate their symptoms (Greve, Ord, Bianchini, & Curtis, 2009), while in cases of mild head injury and chronic fatigue prevalence rates of malingering are an estimated 35% (Mittenberg, Patton, Canyock, & Condit, 2002).

The most frequently used methods for the detection of malingering involve examining intentional underperformance on simple memory tasks (e.g., Iverson & Binder, 2000) or examining over-endorsement of physical or psychological symptoms on self-report tests (Merten, Merckelbach, Giger, & Stevens, 2016). Such tests are called Symptom Validity Tests (SVTs) and have shown to be useful in forensic settings which involve malingering (see Sleep, Petty, & Wygant, 2015; Bianchini, Mathias, & Greve, 2001), but are less often applied in clinical (diagnostic) settings (Schoenberg, Dorr, & Morgan, 2003; Roger, Sewell, & Salekin, 1994).

In general, malingering of physical symptoms such as pain have been under-researched, or investigated mainly through psychometric perspectives (see Crighton et al., 2014). Pain is a reliable concomitant of many physical symptoms, but research so far has failed to design specific methods to detect malingering of pain (Greve, Bianchini, & Brewer, 2013; Fishbain, Cutler, Rosomoff, & Rosomoff, 1999). One difficulty in detecting fabrication of physical symptoms such as pain is that genuine symptoms do fluctuate over time in intensity and durability (Fishbain et al., 1999). Malingerers can therefore report about their genuine 'bad' moments from the past, as if they are still ongoing. Another difficulty is that it is impossible to quantify pain with methods that are independent of patients' self-reports (McDermott & Feldman, 2007). Finally, almost everyone is familiar with pain as a symptom and therefore most malingerers are likely to know what kind of sensations should be reported to appear convincing, which impedes the detection of malingering in this domain (Hamilton & Feldman, 2001).

Given these considerations, there is a need for novel malingering detection methods that do not just focus on memory functioning and/or psychopathology, but on the verbal details of patients'

symptom reports. One recent study that addressed this issue in a systematic fashion is that of Akehurst, Easton, Fuller, Drane, Kuzmin, and Litchfield (2017). These researchers employed a combination of criteria of different verbal lie detection methods, such as Criteria Based Content Analysis (CBCA; see Blandon-Gitlin, Pezdek, Lindsay, & Hagen, 2009; Steller & Kohnken, 1989) and Reality Monitoring (RM; Johnson & Raye, 1981; see Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2013) to identify exaggerated symptoms after exposure to an experimental stressor. Evaluators who used these methods were better in discriminating between truth tellers and malingerers than evaluators who did not use these methods. However, the mere quantity of details in symptoms reports could not serve as robust indicators of veracity (Akehurst et al., 2017). This suggests that the richness of details in symptom reports, the main idea behind CBCA method, is not diagnostic of honesty.

One potentially promising avenue is a newly devised verbal lie detection method: The Verifiability Approach (Nahari, Vrij, & Fisher, 2012). The Verifiability Approach is based on two aspects of deceptive strategies. First, liars tend to provide statements that are rich in details, because they want to make a convincing impression and believe that detailed stories sound convincing. Second, liars tend to avoid mentioning details that could be checked by investigators. As a solution to these conflicting strategies, liars, compared with truth tellers, typically provide fewer details that can be verified and more details that cannot be verified (Nahari, Vrij, & Fisher, 2014a, b; Nahari & Vrij, 2014).

The Verifiability Approach is a promising lie detection approach, but so far, it has not been applied in the context of malingering. Research to date suggests that the efficacy of the Verifiability Approach depends on the context in which it is used (Vrij, Nahari, Isitt, & Leal, 2016; Nahari et al., 2014). For example, in mock crimes scenarios, where an interviewer knows all the details of the “crime”, liars have difficulty in providing verifiable details. Liars find it difficult to demonstrate that they were at a different location than the crime scene during the time the crime occurred. In contrast, in insurance claim cases, someone could falsely claim to have lost his phone while running, but could then truthfully describe his run. This type of situation provides liars with more degrees of freedom to generate false verifiable details (Vrij et al., 2016; Nahari et al., 2014). This might also explain why the Verifiability Approach was not an effective strategy for discriminating between true and false insurance claims (Nahari, Leal, Vrij, Warmelink, & Vernham, 2014). In the case of symptom malingering, similar problems might occur, such as the unknown ground truth and the option to incorporate genuine experiences or symptoms in the fabrication of lies.

Using an “Information Protocol”, several studies (Harvey, Vrij, Nahari, & Ludwig, 2017; Nahari et al., 2014b; Vrij et al., 2016) have examined the effect of informing participants, that the

details of their statements could be subsequently checked by the interviewer. Across studies, this warning has resulted in an increased number of verifiable details being reported by truth tellers but not by liars, strengthening the efficiency of the Verifiability Approach. As part of the Verifiability Approach, this warning has also facilitated discrimination between truths and lies in recent insurance claims study (Harvey et al., 2016; Vrij et al., 2016). Thus, the Information Protocol instructions seem to motivate truth tellers to search their memory for additional verifiable details - something that is not possible for liars to do.

Our two studies are a first attempt to explore the usefulness of the Verifiability Approach to the detection of malingering. We asked participants to write a statement reporting real or fabricated common physical symptoms, such as a headache or stomach ache (Petrie, Faasse, Crichton, & Grey, 2014). In the first study, participants were given the task of writing a symptom report, while in the second study, they were informed that their statements may be checked by a medical professional. We predicted that in both studies, truth tellers would provide significantly more verifiable details about their symptoms than malingerers, whereas malingerers would include more non-verifiable details in their statements than truth tellers (Hypothesis 1). We also predicted that the proportion of verifiable details (verifiable details / total details) would be higher for truth tellers than for malingerers (Hypothesis 2).

Study 1

Method

Participants

We conducted an online study that included 125 undergraduate psychology students. Participants were 17 - 38 years of age, with an average age of 20 years ($SD = 2.48$). The majority were women (86%).

From the total number of participants, 41 reported having real physical symptoms of different medical conditions (see Procedure), whereas 84 did not report any symptoms of physical ill-health. On the basis of these initial symptom reports, participants were allocated into two groups: truth tellers (with real symptoms) and malingerers (without symptoms).

Procedure

After participants signed up for the study, they were directed to an online link to start the survey in Qualtrics. After answering demographic questions, participants were asked to report any physical symptoms of ill-health they were experiencing ("Do you currently or did you in the last week suffer from any physical symptoms, such as a headache, stomach ache, fatigue etc.?"). Participants who answered in the affirmative were considered as 'truth tellers', whereas participants who

responded negatively to the question were next instructed to malingers. Table 2.1 shows the frequencies of selected symptoms for truth tellers and malingerers.

Table 2.1.

Frequencies and percentage of selected symptoms in truth tellers and instructed malingerers in both studies.

| Symptoms | Study 1 | | Study 2 | |
|--------------------------------------------|-------------------------------|---------------------------|-------------------------------|---------------------------|
| | Truth tellers <i>n</i> (%) | Malingers <i>n</i> (%) | Truth tellers <i>n</i> (%) | Malingers <i>n</i> (%) |
| Back or neck pain | 13 (31.7) | 10 (11.9) | 12 (31.6) | 8 (11.9) |
| Headache | 10 (24.4) | 35 (41.7) | 9 (23.7) | 23 (34.3) |
| Fatigue or loss of energy | 8 (19.5) | 19 (22.6) | 6 (15.8) | 15 (22.6) |
| Upset stomach or indigestion | 5 (12.2) | 5 (5.9) | 5 (13.2) | 7 (10.4) |
| Insomnia or sleeping problems | 2 (4.9) | 10 (11.9) | 1 (2.6) | 5 (7.5) |
| Congested or runny nose | 1 (2.4) | 2 (2.4) | 0 | 1 (1.5) |
| Joint pain or stiffness | 1 (2.4) | 0 | 1 (2.6) | 0 |
| Cough | 1 (2.4) | 0 | 1 (2.6) | 4 (6.0) |
| Muscle pain | 0 | 3 (3.6) | 0 | 1 (1.5) |
| Low blood pressure or circulation problems | 0 | 0 | 0 | 2 (3.0) |
| Added symptoms: | | | | |
| - Vertigo | 0 | 0 | 0 | 1 (1.5) |
| - Intestine pain | 0 | 0 | 1 (2.6) | 0 |
| - Knee pain | 0 | 0 | 1 (2.6) | 0 |
| - Sore throat | 0 | 0 | 1 (2.6) | 0 |
| Total | 41 (100) | 84 (100) | 38(100) | 67(100) |

We presented the participants of both groups with 10 of the most frequent physical symptoms reported in the general population (Petrie et al., 2014). Participants also could add additional symptoms if these were not on the list (no new symptoms were added). Malingerers were instructed to select one of the listed symptoms and to write a statement about the target symptom as though they suffered from it. They were presented with the following instructions: “Imagine that you suffer from this specific symptom and try to imagine all the details of experiencing that symptom. *Consider that you did not attend your exam because of this symptom. Imagine we are the exam committee asking you to provide us with specific details of the day on which you experienced the symptom.* Give us a description of your behaviour during the day you “had the symptom”. Your report should start with the morning in which you noticed the symptom and then proceed through the next hours until you went to bed.” Truth tellers received a similar instruction, except that they were asked to give a chronological account of the last day they suffered from their symptom of ill-health. Both groups wrote reports about their symptoms. No length nor time limitations were imposed.

After truth tellers and malingerers had written their statements, we asked them to evaluate the difficulty of this task on 7-point Likert scale (1 = Very easy; 7= Very difficult). They were then thanked for participation and rewarded with one research credit.

Coding

All statements were coded by one coder, and the second coder scored a randomly selected 20% of all statements. Both coders were blind to the veracity of the statements. Following Nahari and Vrij (2014), all details were coded either as *verifiable* or *non-verifiable*. For a detail to be coded as verifiable, it had to meet one of the following criteria. The activities 1) were documented (appointment with a doctor, prescriptions, receipt etc.) and, therefore, potentially checkable; 2) involved an action carried out together with (an) other identified person(s) rather than alone or with a stranger who could not easily be traced; 3) pertained to something that was witnessed by (an) other identified person(s); 4) were reported as being recorded (e.g., on CCTV) by the interviewee; 5) used technology (use of cash machine, bank cards, phone, tablet, computer); or 6) could potentially be checked by blood analysis and medical tests (taking specific pills).

Within the two main groups of details (*verifiable* and *non-verifiable*) we also coded for the presence of the following details, all derived from the Reality Monitoring literature (Johnson & Raye, 1981): Perceptual (i.e., information about what a person has seen, smelt, heard or felt); Spatial (i.e., information about spatial arrangement of objects or people); Temporal (i.e., information about the time when a behaviour/action happened, an event happened or a sequence of events/behaviours happened), and descriptive (i.e., specific description of action, objects or symptoms) details. Every word describing a symptom (“headache”, “stomach ache”, “pain”, “fatigue”), emotional feeling (“I feel”, “anxiety”, “scared”), internal experience or state (“worried”, “decided”, “I wished/wanted”, “thirsty”, “tired”), or information about what a person saw, heard or tasted (“I saw red dots”, “noise”, “bitter”), was coded as a *perceptual* detail. *Spatial* codes included every detail about where an event happened (“at home”, “in the streets”, “at car”), or about spatial arrangements of people or objects (“upstairs/ downstairs”, “down”, “up”, “in front”). *Temporal* details included information about the time in general (“at noon”, “midnight”, “day”), or about a specific time (“at 13h”), or time sequences of the events (“before”, “after”, “during”, “the next day”, “previously”). We also coded descriptions of actions and objects. Every description of an action (“took an Aspirin 500mg”, “called a doctor”, “talked to my friend”), symptom (“strong”, “sharp”, “coming in waves”), or object (“shiny”) was coded as *descriptive* detail. In total, there were four subtypes (perceptual, spatial, temporal, descriptive) of verifiable and of non-verifiable details.

To examine the inter-rater reliabilities between coders intraclass correlation coefficients (ICC) were calculated. They were excellent for verifiable (ICC = .98) and non-verifiable details (ICC = .94),

as well as for the total sum of details ($ICC = .94$). Regarding the separate categories of detail, except for spatial details ($ICC = .63$), the majority of ICC's indicated almost perfect agreement (all ICC's $> .84$; see Supplemental table 2.1). As we did not formulate specific hypotheses about the different detail categories, we will not discuss them in detail below. However, the number of details per category for the two groups can be found in the Appendix 1 (Supplemental table 2.1., 2.2., and 2.3.).

Results

Difficulty of the Task

To check whether truth tellers might have found the task less difficult than malingerers an independent t -test was conducted. However, truth tellers and malingerers reported similar difficulty levels, means being 4.10 ($SD = 1.39$) and 4.46 ($SD = 1.40$), respectively, $t(123) = 1.36, p = .18$.

Length of the Statements

On average, participants produced 89.26 words per statement. Truth tellers provided significantly shorter statements ($M = 66.71, SD = 48.76$) than malingerers ($M = 100.27, SD = 83.52$), $t(123) = 2.38, p = .02, Cohen's d = .49$.

Number of Verifiable and Non-Verifiable Details

The difference between truth tellers ($M = .93, SD = 2.26$) and malingerers ($M = .45, SD = 1.61$) in the raw number of verifiable details reported was not significant, $t(123) = 1.34, p = .18, Cohen's d = .25$. However, truth tellers reported significantly less non-verifiable details ($M = 18.83, SD = 10.43$) than malingerers ($M = 28.29, SD = 21.32$), $t(123) = 2.69, p = .01; Cohen's d = .56$. These results partially support Hypothesis 1.

Proportions of Verifiable Details

Verifiable details were reported by 16.8% of participants, with 24.4% of truth tellers and 13.1% of malingerers reporting at least one verifiable detail, $\chi^2(1, 125) = 2.51, p = .11$. Of the total number of provided details, verifiable information comprised 2.4%.

As in previous studies on verifiability, and as another way to control for statement length, we calculated the proportions of verifiable details: the ratio between the total number of checkable details and overall number of details (verifiable details / total of details). Truth tellers had significantly higher proportions ($M = .05, SD = .12$) than malingerers ($M = .01, SD = .03$), $t(123) = 2.43, p = .01, Cohen's d = .46$. This result supports Hypothesis 2.

Controlling for the length of the statements, an Analysis of Covariance (ANCOVA) showed that there was a significant effect of group (truth tellers vs. malingerers) on the number of verifiable details, $F(1, 123) = 4.72, p = .03, \eta^2 = .03$, and on proportion of verifiable details, $F(1, 123) = 5.70, p = .02, \eta^2 = .04$. However, the effect of group on the number of non-verifiable details, $F(1, 123) = 1.75, p = .19, \eta^2 = .001$, was non-significant.

Study 2

The results of Study 1 revealed that malingerers generated longer statements that were richer in non-verifiable details. Truth tellers produced higher proportions of verifiable details than malingerers – although the proportion of verifiable details produced was low in both conditions. Thus, our results are in line with previous studies on the Verifiability Approach (Nahari & Vrij, 2014; Nahari et al., 2014a, b) and also suggest that low verifiability reports might be a feature of people who malingering suffering from physical symptoms. In Study 2 we tested whether differences in verifiability between truth tellers and liars would become more pronounced when participants are given additional instructions about verifiable details, as was found in previous studies (see Nahari et al., 2014b; Vrij et al., 2016; Harvey et al., 2016).

Method

Participants

105 undergraduate psychology students were recruited. Participants' age ranged from 18 to 26 years, with an average of 20 years ($SD = 1.48$). The majority were women (74%).

From the total number of participants, 38 reported having physical symptoms, while 67 denied suffering from any physical condition. Therefore, as in the previous study, participants were allocated to two groups: truth tellers (with real symptoms) and malingerers (fabricating an account of symptoms).

Procedure

Study 2 followed a similar procedure as Study 1. Participants followed the link that led them to Qualtrics. Once they agreed to participate and answered the demographic questions, they were presented with the symptom list. Participants had an option to choose one of ten symptoms from the list or to add a new one (see Table 2.1). However, unlike our first study, before starting with writing the statements about their symptoms, the participants were given an 'Information Protocol', which informed them that the details they provide may be checked (as in Harvey et al., 2016). The Information Protocol explicitly outlined what kind of information is considered a verifiable detail: "We know from research that liars prefer to avoid providing details that can be verified whereas truth tellers prefer to provide verifiable details. Therefore, we are going to give your statement to medical professionals and ask them to decide if your statement is truthful, based on the extent to which the details you provide can be verified. Verifiable details are activities that can be documented and therefore verifiable (phone calls, doctor appointment, prescriptions etc., or activities that could be checked through blood analysis and medical documentation), carried out with another person (that can be identified), witnessed by another person (identifiable person), or recorded by CCTV cameras. Details that do not meet any of these criteria are considered to be unverifiable."

After writing the statement, using 5-point Likert scales (1 = Completely unmotivated; 5 = Strongly motivated), we asked participants how motivated they had been to write down a convincing statement and to what extent they thought to have succeeded in this. Participants were also asked to report how strongly they believed that the details they provided would be checked by researchers on 5-point Likert scale (1 = Definitely no; 5 = Definitely yes).

We asked malingerers whether they had been using bluffing as a strategy in writing their statements. Bluffing was defined as providing false verifiable details. The possible answers were “Yes”, “Maybe”, and “No”. Perhaps the easiest way of bluffing is to confabulate about a person who can confirm the story (Culhane, Hosch, & Kehn, 2008). In this context, this might be a person who the individual claims has witnessed them experiencing the symptoms or who they have told about their symptoms. To investigate whether malingerers referred to false witnesses, we coded every statement in which a close person (parents, girlfriend/boyfriend, flat mate) was mentioned. After finishing the task, all participants were thanked for participating and rewarded with one research credit.

Coding

As in Study 1, all statements were coded by one coder, while the second coder scored a randomly selected 20% of all statements. Both coders were blind to the veracity of statements. The ICC's between coders was excellent for verifiable details (ICC = .94), non-verifiable details (ICC = .97), and for the total sum of details (ICC = .98). The ICC's for other categories of details also indicated good agreement (all ICC's > .80; see Supplemental Table 2.1).

Results

Motivation, Estimation of Success, Difficulty of the Task, and Belief that Statements Will be Checked

Truth tellers reported ($M = 3.53$, $SD = .79$) a comparable level of motivation as malingerers ($M = 3.43$, $SD = .80$), $t(103) = .57$, $p = .57$. Also, truth tellers ($M = 3.53$, $SD = .76$) did not differ from malingerers ($M = 3.31$, $SD = .96$) in how they rated their success, $t(103) = 1.17$, $p = .24$. As in Study 1, difficulty of the task was rated on a 7-point Likert scale. Truth tellers ($M = 3.76$, $SD = 1.28$) and malingerers ($M = 3.42$, $SD = 1.29$) did not differ with respect to their difficulty ratings, $t(103) = 1.32$, $p = .19$.

Both truth tellers ($M = 3.50$, $SD = .89$) and malingerers ($M = 3.15$, $SD = .96$) considered the possibility that the veracity of their statements would be checked and the groups did not significantly differ in that respect, $t(103) = 1.85$, $p = .07$.

Length of the Statements

On average, participants produced 142.47 words per statement. The length of the statements was not significantly different for truth tellers ($M = 152.63$, $SD = 111.16$) and malingerers ($M = 136.70$, $SD = 85.29$), $t(103) = .82$, $p = .41$.

We compared the results with those of Study 1. The additional instruction in Study 2 affected truth tellers so that they wrote significantly longer statements than in the Study 1, $t(77) = 4.50$, $p = .001$, *Cohen's d* = 1.02. A similar pattern emerged for malingerers, $t(149) = 2.64$, $p = .01$, *Cohen's d* = .04.

Number of Verifiable and Non-Verifiable Details

Truth tellers ($M = 8.26$, $SD = 15.31$) and malingerers ($M = 6.66$, $SD = 9.02$) did not differ in number of generated verifiable details, $t(103) = .68$, $p = .50$ ⁶. The group difference in number of non-verifiable details was not significant either, $t(103) = .63$, $p = .53$, with truth tellers ($M = 48.92$, $SD = 32.76$) and malingerers ($M = 45.03$, $SD = 29.44$) producing a comparable number of such details.

Proportions of Verifiable Details

The number of participants providing verifiable details was much higher than in Study 1 (57.1% vs. 16.8%). In fact, 50% of truth tellers and 63.4% of malingerers reporting at least one verifiable detail, $\chi^2(1, 105) = 1.24$, $p = .26$. Verifiable details formed 13.1% (2.4% in Study 1) of the overall number of details in all statements. As in Study 1, we calculated the proportion of provided information that could be checked. The average proportion of verifiable details (verifiable details / total of details) was .12 ($SD = .15$) for truth tellers, and .13 ($SD = .12$) for malingerers; this difference was not significant, $t(103) = .30$, $p = .77$.

Bluffing as a Strategy

From the total number of malingerers, 17 participants (25.4%) reported that they “maybe” had used bluffing, while 19 malingerers (28.4%) admitted providing false verifiable details.

From a total of 41 malingerers who provided (false) verifiable details, 63.4% mentioned a close person who could confirm their story. On the other hand, 57.9% of truth tellers also provided information about family members or close people who could confirm their story. The groups did not differ significantly in the frequency with which they mentioned close people, $X^2(1) = 1.03$, $p = .31$.

We checked whether malingerers' belief that their statements would be checked correlated with their bluffing strategies using Spearman's rho correlation coefficient. However, the relation was not significant, $r_s(105) = .09$, $p = .35$. Similarly, the correlation between malingerers' belief and the number of verifiable details remained non-significant, $r_s(105) = .01$, $p = .92$.

⁶ The main analysis for the number of verifiable details was run again with the belief of participants that their statements would be checked as a covariate. The covariate did not change the result, $F(1, 122) = .34$, $p = .56$.

Table 2.2.
Summary of main findings in Study 1 and Study 2.

| Results | Length of the statements | Difficulty of the task | Number of verifiable details | Number of non-verifiable details | Proportion of verifiable details |
|----------------|--------------------------|------------------------|------------------------------|----------------------------------|----------------------------------|
| Study 1 | >* | ~ | ~ | >** | <* |
| Study 2 | ~ | ~ | ~ | ~ | ~ |

Notes: * $p < 0.05$; ** $p < 0.01$; “>” Malingers having higher scores; “<” Malingers having lower scores; “~” No significant differences between groups.

Discussion

We examined whether the Verifiability Approach (Nahari & Vrij, 2014) could differentiate between people who are suffering from common physical symptoms and those who are malingering such symptoms. Our main findings in Study 1 and Study 2 are presented in Table 2.2. Truth tellers included a higher proportion of verifiable details despite generating shorter statements than malingers. This finding was also evident when we looked at the number of verifiable details, while controlling for the length of the statement. However, this effect only emerged in Study 1, where participants were not provided with an instruction to include verifiable details. In Study 1, the overall production of verifiable details remained low (2% of total details). When the pertinent instruction was provided (Study 2), the overall production of verifiable details increased (13%), and no difference between truth tellers and malingers in non/verifiable details emerged. This pattern appears to indicate that the instruction weakened the effect of the Verifiability Approach, which opposes previous studies where the use of such an instruction enhanced the differences between truthful and deceptive accounts. This discrepancy suggests that detecting malingers using the Verifiability Approach may be more effective if the patients are required to provide their reports spontaneously – rather than warning them that their reports will be examined for verifiable details. However, it might come with a risk that patients will not spontaneously include verifiable details in their reports.

We believe that the discrepancy in findings between our two studies is related to different levels of difficulty to incorporate false verifiable details (bluffing) into an account in this specific setting. The most popular way of bluffing observed in the current research was claiming that another person could confirm the account, most frequently a person closely related to the malingerer (e.g., parents, boyfriend, flatmate) (see also Culhane et al., 2008). Actually, mentioning close people as witnesses is a clever strategy because the majority of people are willing to corroborate a statement of a close friend or relative in order to help that person (Hosch et al., 2011). It may well be the case that this strategy works better in a malingering context than in a criminal/insurance setting. For example, a criminal or a fraudulent claimant needs to inform a false witness beforehand to pretend that s/he was with or spoke with the criminal/claimant, which means that the friend will be aware of the falsehood of the statement. Consistent with this, Vrij et al. (2016) found that only 17% of liars

reported discussing the incident with the person they mentioned in their statements, compared with 77% of truth tellers. In a malingering situation, even if a friend denied noticing a malingerer's symptom (e.g., headache), a symptom report would not necessarily be exposed as a form of malingering, because people often do not mention to others that they suffer from a particular symptom. Furthermore, in the context of symptoms it is easy to actually fool friends because common physical symptoms are often not clearly visible to others. Also, malingering often includes preparation and it occurs more frequently in the form of exaggeration of existing symptoms (Rogers, 2008). This means that a person is likely to plan how to include witnesses in the deception without them being aware, and to overreport the intensity of already present symptoms, which is even more difficult to disprove due to its subjective nature. Consistent with this is the result showing that malingerers did not differ from truth tellers in ratings of the difficulty of writing the statements about their symptoms. This may indicate that providing false potentially checkable information did not pose a big challenge for malingerers.

An interesting point concerns the length of the statements in both studies. While truth tellers spontaneously wrote significantly shorter statements than malingerers in Study 1, this difference disappeared in Study 2 when we provided the Information Protocol. It is plausible that malingerers, even when not instructed, tend to provide long, non-verifiable statements in order to conceal the lack of truthful information (e.g., Vrij et al., 2016). In contrast, truth tellers' parsimony in details might be a result of their belief that their honesty will "shine through", commonly referred to as the "transparency effect" (Savitsky & Gilovich, 2003). The DSM-4 assumes that malingerers are uncooperative and reluctant to talk about their symptoms (Rogers, 2008). However, the finding that malingerers spontaneously write longer statements contradicts these assumptions and warrants further study. On the other hand, when provided with additional instructions, truth tellers wrote longer statements which were comparable in length to those of malingerers. Thus, with these instructions, the difference between groups was no longer evident. Similar to these findings are the results of previous insurance claims studies in which researchers suggested that providing a detailed model statement about an unrelated topic would elicit more verbal clues of deception, such as longer and more detailed reports among truth tellers than among liars. However, the results showed that, even with the model statement, liars provided statements that were comparable in length with those of truth tellers (Leal, Vrij, Warmelink, Verham, & Fisher, 2013).

The main disadvantage of the Verifiability Approach in a medico-legal context seems to be the low spontaneous production rate for verifiable details in genuine patients' report. The percentage of truth tellers who reported verifiable details in Study 1 was around 17%, while it was 57% in Study 2 following instructions to provide such details. This suggests that the majority of genuine patients

experiencing physical symptoms of ill-health do not spontaneously provide checkable details, or simply might not have any checkable details to report about. In the context of symptoms, the majority of information provided by patients is subjective, and mostly concentrated on their internal state, rather than on external or visible condition. Thus, the situations in which professionals have the option to verify persons' symptoms complaints via cameras or witnesses may be extremely rare (Resnick et al., 2008). Additionally, people differ in the way they perceive their symptoms and behave when experiencing them (see Kolk, Hanewald, Schagen, & van Wijk, 2003; see also van Wijk, Huisman, & Kolk, 1999). While one person immediately calls a doctor or goes to the pharmacy, another person may just keep up with daily activities, without complaining to anybody. Both persons are truth tellers, but the second one would not have any verifiable details to report concerning their physical symptoms of ill-health. The additional concern is that a malingerer could also go to the doctor or pharmacy just to prepare a verifiable report. Unfortunately, in those cases, the Verifiability Approach would not be able to assist the detection of the false (verifiable) report.

One important limitation of the current study is that we relied on self-reports for the selection of truth tellers and malingerers, without any independent check as to whether they actually suffered from the reported symptoms or not. However, the incidence of selected symptoms amongst truth tellers was consistent with previous research about symptoms most frequently experienced by non-malingerers (see Petrie et al, 2014; Dandachi-FitzGerald & Merckelbach, 2013).

Future Research

Despite limitations of the Verifiability Approach, we do not exclude the possibility that, with certain adjustments, it may be possible to extend the Verifiability Approach into an efficient tool for the detection of malingered symptoms. The adjustments should focus on the consequences of including false verifiable details, because in reality people are confronted with losses if it is established that they malingered. However, such a warning could also influence truth tellers, who then might realize that they could be seen as malingerers.. Furthermore, our results seem to indicate that when the length of the statement is controlled for, the differences in verifiable details between groups become more pronounced. Therefore, introducing certain limitations to the length of symptom reports might help in differentiating between genuine and fabricated statements.

Conclusion

Our results suggest that the Verifiability Approach is not as effective in a malingering setting as in previously studied settings. However, using this method may be informative, especially in a domain in which self-report is inevitable, such as a medico-legal context. Therefore, future studies are needed in order to determine whether certain versions of the Verifiability Approach might be useful in a medico-legal context in which malingering is an issue.

THE VERIFIABILITY APPROACH TO FABRICATED SYMPTOMS

STUDY 3

Abstract

Several studies on the Verifiability Approach found that truth-tellers report more verifiable details than liars. Therefore, we wanted to test whether such a difference would emerge in the context of malingered symptoms. We obtained statements from undergraduates ($N = 53$) who had been allocated to three different conditions: truth-tellers, coached malingerers, and naïve malingerers. Truth-tellers carried out an intensive physical exercise and after a short interval wrote a report about their experience and elicited symptoms. The two malingering groups had to fabricate a story about the physical activity and its symptoms. Truth-tellers did not generate more verifiable details than malingerers. However, malingerers reported more non-verifiable details than truth tellers. Coached and naïve malingerers did not differ in this respect. Relative to truth-tellers, naïve malingerers reported more symptoms-related non-verifiable details, while coached malingerers reported more exercise-related non-verifiable details. Focusing on non-verifiable details may inform the detection of malingered symptoms.

Key words: Malingering, Verifiability Approach; Symptoms; Deception Detection.

Verifiability on the Run: An Experimental Study on The Verifiability Approach to Malingered Symptoms

Malingering is defined as “the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives” (DSM-5; American Psychiatric Association 2013, p. 726). Malingering may be driven by several incentives, such as financial compensation (e.g., benefits eligibility), legal outcomes (e.g., reducing or avoiding sentences), privileges (e.g., receiving stimulant medication), or other advantages (e.g., avoiding undesirable work). Because malingering may obscure diagnostic decision making, it is important to rule it out whenever these incentives may play a role in patients presenting with symptoms (Vilar-López et al., 2007). To this end, clinicians or researchers may employ screening instruments, and collateral data from different sources (e.g., medical records; Mittenberg, Patton, Canyock, & Condit, 2002).

Currently, the majority of research papers on malingering are concerned with tests and tasks that intend to identify malingerers (for reviews see Smith, 2008; Sollman & Berry, 2011). However, how malingerers talk about their symptoms and whether their speech or written reports may contain cues to malingering (e.g., verbal cues of malingering) has received less attention in the research literature. According to the DSM-5 (APA, 2013), malingerers exhibit a lack of co-operation when they are evaluated by medical professionals. This would seem to imply that malingerers are reluctant to talk about their symptoms (see also Worley, Feldman & Hamilton, 2015). However, there is no empirical evidence to support this assumption and in fact, there are even indications to the contrary. In a recent study by Akehurst et al. (2017), participants were instructed to undergo a cold pressor procedure and then either honestly reported or exaggerated their symptoms. To screen for malingering, half of the interviewers used a checklist based on criteria from different verbal veracity assessment methods such as Criteria Based Content Analysis (CBCA; see Blandon-Gitlin, Pezdek, Lindsay, & Hagen, 2009; Steller & Kohnken, 1989) and Reality Monitoring (RM; Johnson & Raye, 1981; see Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2013). Interviewers who used the checklist obtained a 75% correct classification of truth-tellers and a 66% correct classification of the exaggerators. Without the checklist, interviewers performed at a level no better than chance (50%). Interestingly, exaggerators generated *more* unusual details in their accounts than truth tellers, which runs counter to the clinical impression that malingerers are reluctant to talk about their symptoms.

Another approach that may help in identifying malingered symptom reports is the Verifiability Approach (Nahari & Vrij, 2014). It is based on the idea that liars want to provide statements that are rich in details, because they believe that such statements will be convincing. At the same time, they are reluctant to include too many details because they fear that too many details will provide leads for investigators. One way to resolve this dilemma is to provide an abundance of details that an

investigator cannot check (i.e., unverifiable details). Research has shown that relative to truth-tellers, liars do indeed provide fewer details that can be verified and more details that cannot be verified (Nahari et al., 2014a,b; Nahari, Leal, Vrij, Warmelink & Vernham, 2014). This so-called verifiability effect has been observed in mock crimes (Nahari, Vrij & Fisher, 2014a,b) and in false insurance claims settings (Harvey, Vrij, Nahari, & Ludwig, 2016; Nahari et al., 2014; Vrij, Nahari, Isitt, & Leal, 2016).

The Verifiability Approach states that the verifiability of provided details distinguishes liars from truth-tellers (Nahari et al., 2014). This focus on potential checkability deviates from traditional verbal veracity assessment methods (i.e., CBCA and RM) that primarily look at aspects such as the quantity of perceptual details and the reproduction of dialogues in the reports of liars and truth tellers. The difference between the Verifiability Approach and CBCA in terms of ability to detect deception becomes more pronounced when individuals have prior knowledge about how the veracity assessment method works (Nahari et al., 2014). That is, the effectiveness of the CBCA approach as a tool for detecting fabricated statements is impaired when interviewees are informed beforehand about the working of CBCA (Vrij, Akehurst, Soukara & Bull, 2002). In contrast, the verifiability effect appears to become stronger after interviewees have been given an Information Protocol informing interviewees that their statements might be checked for verifiable details (Harvey, Vrij, Nahari, & Ludwig, 2016; Vrij et al., 2016).

In two exploratory studies (Boskovic, Bogaard, Merckelbach, Vrij, & Hope, 2017), we examined the verifiability effect in the context of malingered symptoms. Previous verifiability studies were primarily oriented towards coding external details (e.g., perceptual, spatial, temporal), without including subjective details such as emotions or cognitive operations because of their unverifiable status (Nahari, Vrij, & Fisher, 2012). Like emotions, symptoms are subjective experiences. However, there is an important difference between emotions and symptom reports. As a rule, genuine symptoms lead to specific behaviours associated with those symptoms (e.g., going to/calling the doctor, taking medications, restricted activity, not going to work/studies, complaining to others, googling symptoms) that can be documented, witnessed, or carried out with another person. For example, 58% of people experiencing common symptoms (e.g., headaches) tend to use medication (prescribed or not), 48% make complaints to a friend or a family member, 24% reduce activity, and 5% actually consult a medical professional (Verbrugge & Ascione, 1987). Thus, the behavioural sequelae of their symptoms are, in principle, verifiable.

In our exploratory studies (Boskovic et al., 2017), we asked people to describe genuine physical symptoms or to fabricate an account about these symptoms. When participants were not informed about the details they should provide, a heightened number of non-verifiable details (rather

than a lack of verifiable details) was typical for fabricated symptom reports. Previous verifiability studies also noted that liars provide more non-verifiable details than truth-tellers (Vrij et al., 2016). In our studies, the abundance of non-verifiable details in fabricated symptom reports was so pronounced that it led to a marked difference between malingerers and honest participants in the length of their symptom descriptions, with malingerers' statements being significantly longer (Boskovic et al., 2017). This pattern (i.e., longer statements by deceptive interviewees) contradicts many deception studies (see DePaulo, 2003; Vrij, 2000), but is in accordance with people's belief that deceptive reports include more details rather than fewer (Granhag, Andersson, Stromwall, & Hartwig, 2004). Our finding is also consistent with Akehurst et al. (2017), who observed that exaggerators provide more unusual symptom descriptions than do truth-tellers.

In our exploratory work, we also evaluated the Information Protocol (i.e., instructing participants that details might be checked). This manipulation did not reduce the volume of verifiable details in malingerers as opposed to truth-tellers. Instead, malingerers reported false verifiable details and in particular reported false witnesses who they claimed could confirm their stories. Thus, we did not observe the typical verifiability effect. This discrepancy with previous verifiability studies might be related to the differences in investigated contexts (malingerers' symptoms versus lying about events or actions). For example, providing a false witness in a criminal context requires a conspirator who is willing to confirm a false alibi. Things are quite different for malingerers. They do not need to reveal that they are being deceptive to those who might be asked to confirm their story (e.g., friends, family). Thus, a malingerer can lie about his/her symptoms in front of others, which makes it easier to create a convincing account with false verifiable details.

The absence of a typical verifiability effect (i.e., honest people reporting more verifiable details than malingerers) in our exploratory studies might have been caused by truth-tellers' symptoms not being salient enough. That is, truth-tellers may have described mild symptoms that had no behavioural sequelae. Furthermore, their statements were based on retrospective self-reports, and we could not determine whether their symptoms were also present at the moment they described them (Boskovic et al., 2017).

With this in mind, we wanted to test whether a verifiability effect occurs when symptoms are actually elicited in truth-tellers, while malingerers only fabricate a story about them. Eliciting symptoms allows for controlling the time and duration of the symptoms in truth tellers. This way, truth-tellers share an identical experience, and report about acute symptoms shortly after they have occurred, which makes symptom reports less dependent on memory (e.g., Miranda, Gold, Gore, & Punnett, 2006). Thus, in the current study, participants either took part in or imagined taking part in an exhausting activity (running up and down the stairs) and were then told to report about the activity.

Several studies have found that liars often use details of previous experiences when lying (i.e., embedded lies; Leins, Fisher, & Ross, 2013), and that malingerers mostly report about symptoms they are familiar with (Dandachi-FitzGerald & Merckelbach, 2013). Thus, successful malingerers may embed a lie about an activity (e.g. Nahari & Vrij, 2014; Gnisci, Caso, & Vrij, 2010) and its symptoms in an event that they really experienced at one time. Thus, we coached one malingering group to use embedded lies (e.g., recalling previous experiences of running and related symptoms and report them as current symptoms) to investigate the efficacy of that strategy. Naïve malingerers were not given any specific instructions how to mangle the physical symptoms of running. We predicted that truth-tellers would produce more verifiable details and a higher proportion of verifiable details, whereas malingerers would generate more non-verifiable details (Hypothesis 1). Furthermore, we expected to find differences between the two malingering groups, such that an embedded lies-strategy would enable coached malingerers to fabricate more verifiable details about their experience than naïve malingerers (Hypothesis 2).

Method

Participants

A total of 53 (42 women) university students participated in the study. Their average age was $M = 21.13$ years ($SD = 2.56$), with a range from 18 to 29 years. Participants were assigned to three different conditions: Truth-tellers ($n = 18$), naïve malingerers ($n = 17$), and coached malingerers ($n = 18$). Students were compensated with either a course credit point or a voucher valued €7.50.

All participants were pre-screened for health problems and poor physical condition. None of 18 participants who were assigned to the physical exercise condition reported any serious health issues, nor were they taking medication at the time of the experiment. The majority ($n=16$) exercised more than three times a week, while two people reported working out only once a month. In the malingering groups ($n = 35$), three participants reported having health problems and four were taking medications. In total, 33 participants exercised regularly, while two participants reported not being physically active.

Procedure

Truth-tellers had to perform two tasks. The first task intended to induce symptoms, while the second gave participants the opportunity to exhibit certain symptom-related behaviours (e.g., talk to a friend about it, go to the nearby pharmacy). The first task was carried out in small groups in the stairways of a university building, which was next to the hospital (a picture of the stairs is given in Appendix 1). Participants were instructed to run down and up the stairs, from the third floor to the ground floor, twice, as quickly as they could. This exercise was followed by a 30 minutes break (second task) in which participants could go wherever they wanted in (e.g., library, café, restaurant,

pharmacy, home). After their return to the laboratory, they needed to write a statement about their experience during the physical exercise, including descriptions of their symptoms, and all the details about the exercise itself, and about what they did during the break. They were instructed to describe the experience (e.g., surroundings, their actions, and symptoms during the exercise and the break) including as many details as possible.

Both malingering groups were told about truth-tellers' assignment and were instructed that they had to convince researchers they were truth-tellers. Malingers were given a general description of truth-tellers' exercise (where and when it happened) and of the instruction to take a break of 30 minutes after the exercise. They were students and so they were familiar with the environment in which the study took place (e.g., where the stairs were, the writings and pictures on the walls, the nearby library, hospital, and shops in which students could have gone during the break). Malingers were instructed to fabricate a statement that included as many details as possible about their actions and bodily sensations and overall experience during the running and the break. However, while one group was just told to fabricate the statements (naïve malingerers), the coached malingerers were given additional instructions how to fabricate their statements. First, they were instructed to recall the last time they had performed an intense physical activity (e.g., running, mountain climbing). Second, they were asked to write everything they could remember about the symptoms they experienced during that activity, but to pretend that they experienced those physical symptoms during and after participation in this study (running down and up the stairs). Thus, coached malingerers were explicitly instructed to use a previous experience to lie about the target experience in the current procedure. They were also instructed to confabulate (embed lies) about all the details of the exercise and about what they did during the break. As in the other conditions, participants were encouraged to describe the experience providing as many details as possible (see Appendix 1).

Written instructions for all tasks were handed to participants and also read aloud by the researcher. After reading the instructions, the researcher again repeated the instructions and participants were invited to ask questions. Participants were told that they could earn an extra credit point or voucher (€7.50), if they were able to convince the researchers that they had, indeed, carried out the whole experiment (running and break).

Measures

Coding. One coder evaluated all statements, while a second, independent coder coded a randomly selected 25% of statements. Both coders were blind to the three different conditions. In coding the statements, particular details were excluded, such as information about the researchers or anything that had been part of the instructions. We excluded paraphrases of the instructions to avoid artificially raised levels of verifiable details among all three groups.

Following Nahari and Vrij (2014), all details were coded either as *verifiable* or *non-verifiable*. For a detail to be coded as verifiable, it had to meet one of the following criteria. The activities 1) were documented and therefore potentially checkable (e.g., the receipts for drinks or food; descriptions of writings on the stairs or pictures on the walls); 2) involved an action carried out together with (an) other identified person(s) (rather than alone or with a stranger who could not easily be traced) (e.g., identifying a person in the group who participated in the exercise as well); 3) pertained to something that was witnessed by (an) other identified person(s) (e.g., complaining to a friend during the break about the symptoms); 4) recorded, as mentioned by the examinee, on CCTV cameras (e.g., being in the library/pharmacy/nearby shops); 5) used potentially traceable technology (e.g., use of cash machine, bank cards, phone, tablet, computer); or 6) could be checked by blood analysis and medical tests (e.g., taking specific pills). The remaining details were classified as unverifiable.

Inter-rater agreement between the two coders, measured with inter-class correlation coefficients (ICC), was excellent for both the verifiable (.99) and non-verifiable (.97) details.

Symptom-related, exercise-related, and neutral details. We looked at overall frequencies of verifiable and non-verifiable details. However, because we were primarily interested in statements about feigned and genuine symptoms and their behavioural expressions, we also carried out a more fine-grained analysis and coded three categories: 1) Symptom-related, describing the symptoms (e.g., “sweating”; “pain in legs”; “shaking in front of Barry”); 2) Exercise-related, if details were referring to running (“I came second”; “I changed the tempo”; “I bumped into my ex-tutor while running”); and 3) Neutral, describing activities that were not related to the symptoms or to the exercise, but still provided insight in participants’ behaviour during the break (“I called my boyfriend”; “Bought a sandwich”; “Sat on the sofa”).

All three types of details were coded as verifiable or non-verifiable. ICC’s indicated strong agreement between the coders (all ICC’s $> .79$ and $< .99$).

Results

Number of Verifiable and Non-Verifiable Details

Table 2.3. summarizes the main results. There was no main effect of a group with respect to the number of verifiable details, $F(2, 50) = .78, p = .46$. To test whether coached malingerers would include more verifiable details in their accounts than naïve malingerers (Hypothesis 2), we contrasted the two malingering groups with regard to their verifiable details. Contrary to our prediction, the difference was not significant, $t(33) = 1.32, p = .25$, *Cohen’s d* = 0.39. Neither did truth tellers differ

from naïve malingerers, $t(33) = .17, p = .87$, or coached malingerers, $t(34) = 1.18, p = .25$, in terms of verifiable details.

There was a significant main effect of a group with respect to the number of non-verifiable details, $F(2, 50) = 7.82, p = .001, \eta^2 = 0.24$. Follow-up t -tests indicated that naïve malingers and coached malingerers produced significantly more non-verifiable details than truth tellers, $t(33) = 3.30, p = .002, \text{Cohen's } d = 1.11$ and $t(34) = 3.93, p = .001, \text{Cohen's } d = 1.31$, respectively. The two malingering groups did not differ with respect to the number of non-verifiable details, $t(33) = .15, p = .88$. Thus, the pattern of non-verifiable and verifiable details across groups only partially supports Hypothesis 1.

Proportion of Verifiable Details

Across the sample, verifiable details were reported by 28 participants (52.8%). From the total number of details reported, only 4.4% were verifiable. As in previous studies on verifiability, we calculated for each participant the proportion of verifiable details, i.e., the ratio between the total number of checkable details and overall number of details (verifiable details / total of details). Next, we ran a one-way Analysis of Variance (ANOVA). The three groups did not differ regarding the proportion of verifiable details reported, $F(2, 50) = 2.07, p = .13, \eta^2 = .08^7$.

Table 2.3.

Means and standard deviations of different detail categories and number of words in truth tellers, naïve and coached malingerers.

| | Group | | |
|----------------------------------|--------------------------------|------------------------------------|--------------------------------------|
| | Truth tellers <i>M (SD)</i> | Naïve malingerers <i>M (SD)</i> | Coached malingerers <i>M (SD)</i> |
| <i>N</i> | 18 | 17 | 18 |
| Verifiable details | 6.78 (9.85) | 7.41 (12.27) | 3.67 (5.38) |
| Proportion of verifiable details | .07 (.11) | .04 (.06) | .02 (.02) |
| Non-verifiable details* | 91.72 (38.19) | 147.00 (59.10) | 149.72 (49.43) |
| Length of the statements* | 264.44 (100.77) | 411.00 (154.39) | 427.11 (157.43) |

Notes: * $p < .01$; Length of statements = Number of words.

Symptom-Related, Exercise-Related, Neutral Verifiable and Non-Verifiable details

Table 2.4. shows the means and standard deviations of the three types of details across groups. The groups differed significantly in terms of symptom-related non-verifiable ($F(2, 50) = 4.88; p = .01; \eta^2 = 0.16$) and exercise-related non-verifiable details ($F(2, 50) = 7.26; p = .001; \eta^2 = 0.23$). Post hoc comparisons using Bonferroni procedure indicated that naïve malingerers provided significantly

⁷ Analyses were also performed using *Kruskal-Wallis* independent tests, and the results remained non-significant.

more symptom-related non-verifiable details than truth tellers ($p = .01$), and coached malingerers reported significantly more exercise-related non-verifiable details than truth tellers ($p = .001$). The two malingering groups did not differ with regard to the number of symptom-related and exercise-related non-verifiable details, $p = .16$; $p = .26$, respectively.

No significant group differences emerged for number of symptom-related verifiable ($F(2, 50) = .59$; $p = .56$), exercise-related verifiable ($F(2, 50) = .17$; $p = .84$), and neutral verifiable and non-verifiable details ($F(2, 50) = .45$; $p = .64$ and $F(2, 50) = 1.94$; $p = .15$, respectively).

Length of Statements

We calculated the total number of words for each of the three groups (length of statements). The three groups did differ with regard to this parameter, $F(2, 50) = 7.36$, $p = .002$, $\eta^2 = 0.23$. Both naïve malingerers and coached malingerers produced longer statements than truth tellers, $t(33) = 3.34$, $p = .002$, $d = 1.13$ and $t(34) = 3.69$, $p = .001$, *Cohen's d* = 1.23, respectively. Follow-up *t*-tests indicated that the two malingering groups did not differ from each other with respect to statement length, $t(33) < 1.0$.

Table 2.4.

Means and standard deviations across groups on symptom-related, exercise-related, and neutral verifiable and non-verifiable details.

| Codes | | Truth tellers <i>M (SD)</i> | Naïve malingerers <i>M (SD)</i> | Coached malingerers <i>M (SD)</i> |
|----------------|--------------------|--------------------------------|------------------------------------|--------------------------------------|
| Verifiable | Symptom-related | 1.00 (4.00) | 1.76 (5.32) | .34 (1.44) |
| | Exercise-related | 2.22 (4.71) | 1.82 (4.33) | 1.45 (2.64) |
| | Neutral | 3.56 (7.75) | 3.82 (6.89) | 1.89 (4.69) |
| Non-verifiable | Symptom-related* | 37.95 (22.51) | 66.47 (36.36) | 48.12 (21.01) |
| | Exercise-related** | 43.28 (22.84) | 63.53 (32.32) | 81.00 (33.08) |
| | Neutral | 10.67 (6.24) | 17.00 (14.55) | 20.62 (21.27) |

Notes: * $p < .01$; ** $p < .001$

Discussion

The principal aim of this research was to investigate whether the Verifiability Approach (Nahari & Vrij, 2014) could be used to discriminate effectively between truth-tellers, naïve malingerers, and coached malingerers. Their reports concerned a physical exercise and the symptoms it elicited. Additionally, we examined whether an explicit strategy of embedding lies in previous true experiences would result in coached malingerers providing more verifiable details and, in doing so, would render their reports more convincing relative to those of naïve malingerers.

The results of our study can be summarized as follows. First, both malingering groups provided significantly more non-verifiable details than truth-tellers. More precisely, compared with truth-tellers, coached malingerers produced more non-verifiable information describing the exercise itself, whereas naïve malingerers generated more non-verifiable details about their symptoms. This might be a result of differences in the instructions, or shows how malingerers had different strategies compared with truth-tellers in presenting themselves as honest. We have no ready explanation for this pattern and it requires replication to determine whether differences in types of non-verifiable details between malingerers and truth-tellers is a robust phenomenon.

Second, truth-tellers and malingerers did not differ in terms of number or proportion of verifiable details reported. Furthermore, the use of embedded lies as a strategy did not contribute to more (false) verifiable statements by participants in the coached malingering condition, while truth-tellers generated verifiable details at a low base rate. The low overall production of verifiable details (4.4%) suggests that even for truth tellers, reporting verifiable details when describing the physical exercise and the symptoms it elicited was an arduous task. Together with our previous findings (Boskovic et al., 2017), our results indicate that in the context of symptoms reporting, an extensive number of non-verifiable details might be more of a red flag for malingering than the lack of verifiable details. The increased number of non-verifiable details in malingerers' reports fits with some clinical observations about the way malingerers talk about their symptoms. For example, Resnick and Knoll (2005) noted malingerers' tendency to provide vague descriptions of their symptoms. Our findings are also consistent with Akehurst et al. (2017), who found malingerers to produce more unusual details than truth-tellers.

Closely related to their tendency to report more non-verifiable details than truth tellers, malingerers' reports were significantly longer than those of truth-tellers. This replicates our previous results (Boskovic et al., 2017). It appears that malingerers in the current study tried to avoid reporting information that might have enabled the researchers to detect that they did not actually participate in the symptom-eliciting exercise or in the break that followed the exercise. Thus, they compensated for the absence of specific information by providing more non-verifiable details (e.g., Vrij et al., 2016).

On the other hand, truth-tellers did not generate an abundance of verifiable details, possibly because they believed that their honesty would shine through. The “illusion of transparency” that truth-tellers might have, has been described in several domains (e.g., interrogations; Hartwig, Granhag, & Stromwall, 2007; Savitsky & Gilovich, 2003).

The current study, as well as previous work (Akehurst et al., 2017; Boskovic et al., 2017), indicates that if anything, malingerers produce more lengthy reports about their symptoms than do truth tellers. This pattern is difficult to reconcile with the DSM assumption that malingerers are uncooperative and reluctant to talk about their symptoms (Rogers, 2008; Worley, et al., 2009). It also contradicts the widespread belief that malingerers will experience difficulties when elaborating their fabricated symptoms and that their symptom reports will, therefore, be brief and less convincing (Ali, Jabeen, & Alam, 2015).

One limitation of our study was the lack of consequences for providing false verifiable details. To illustrate, one of the naïve malingerers wrote: “I am obese and I have a herniated disc.”. In theory, the health history and current health state of an individual are checkable (Nahari & Vrij, 2014), and so this was scored as a verifiable detail. However, we did pre-screen every participant for any health problems, and this participant did not report these problems in advance of the study. Thus, he provided a false verifiable detail. Given the absence of consequences for deceiving the interviewer, it is possible that participants were not particularly concerned about lying about a witness alibi or other verifiable details. The Verifiability Approach is likely to be more effective in settings where serious consequences for providing false verifiable details are present. Therefore, future research might want to examine verifiability effects in clinical settings where there are real consequences associated with the detection of malingering. A second limitation is that the exercise that participants had to perform was familiar to everyone, and therefore an easy starting point for confabulation and malingering. It might be worthwhile to explore verifiability effects with less common symptoms (e.g., hearing voices) in order to determine whether genuine patients report more verifiable details than malingerers.

One could argue that due to the subjective nature of symptoms, the Verifiability Approach cannot be used as a tool to detect malingering. However, even everyday symptoms often have behavioural consequences that are open to verification (e.g., telling a family member or friend about the symptom; Verbrugge & Ascione, 1987). Moreover, there are types of psychopathology that include both subjective symptoms and verifiable elements. Based on our studies so far, it might be that the Verifiability Approach is not a powerful tool to detect feigned symptoms per se. However, it might well be that the Verifiability Approach is effective in screening for confabulated stories about trauma exposure. This issue requires further study.

Conclusion

Based on the current results and our previous study, we conclude that non-verifiable details are a better indicator of malingering than lack of verifiable details. Relatedly, unlike the clinical impression that malingerers provide brief accounts of their symptoms, we and others (Akehurst et al., 2017) found that malingerers tend to produce extensive symptom reports.

CHAPTER 3

THE MODIFIED STROOP TASK AND THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED SYMPTOMS

This chapter draws on the following papers:

- Boskovic, I., Biermans, A., Merten, T., Jelacic, 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*. In press.
- Boskovic, I., Hope, L., Ost, J., Orthey, R., & Merckelbach, H. (2017). Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task. Accepted manuscript pending revision.

**THE MODIFIED STROOP TASK AND THE SELF-REPORT SYMPTOM
INVENTORY IN FABRICATED SYMPTOMS
THE CASE OF TEST ANXIETY
STUDY 4**

Abstract

Some researchers argue that the Modified Stroop Task (MST) can be employed to rule out feigning. According to these authors, Modified Stroop interference effects are beyond conscious control and therefore indicative of genuine psychopathology. We examined this assumption using a within-subject design. In the first session, students ($N = 22$) responded honestly, while in the second session they were asked to read a vignette about test anxiety and then fake this condition. During both sessions, we administered an MST consisting of neutral, anxiety-related, and test anxiety-related words. Participants also completed the Self-Report Symptom Inventory (SRSI; Merten et al., 2016) that focuses on over-reporting of pseudosymptoms. Our feigning instructions were successful in that students succeeded in generating the typical MST effect by providing longer response latencies on anxiety related ($r = 0.43$) and test anxiety – related ($r = 0.31$) words, compared with neutral words. Furthermore, students endorsed significantly more pseudosymptoms on the SRSI ($r = 0.62$) in the feigning session than in the honest control condition. We conclude that the MST effect is not immune to feigning tendencies, while the SRSI provides promising results that require future research.

Key words: Feigning; Over-reporting; Feigning; Test anxiety; SRSI; Modified Stroop task.

The Modified Stroop Task Is Susceptible To Feigning: Stroop Performance and Symptom Over-endorsement in Feigned Test Anxiety

The Modified Stroop Task (MST) is widely used in research on various psychological problems. For example, the MST has been applied in investigating the cognitive underpinnings of addiction, such as alcoholism (Kramer & Goldman, 2003) or gambling (Boyer & Dickerson, 2002). Moreover, it is also used in evaluations of treatments in patients with eating disorders (Ball et al., 2004), and among sex offenders (Price & Hanson, 2007). However, the MST has been the most frequently applied in research on anxiety symptoms (see Mathews & MacLeod, 1985; Richards, French, Johnson, Naparstek, & Williams, 1992; Lovett, 2005, for reviews). In this task, participants have to name the color of neutral and anxiety-related words as quickly as possible, while disregarding the content of the words. Typically, participants with high anxiety levels show longer reaction times for threatening than for neutral words. The MST effect is highly specific (Mathews & Macleod, 1985) and has been documented in patients with obsessive-compulsive disorder, posttraumatic stress disorder, panic disorder, social phobia, and specific phobias (Becker, Rinck, Margraf, & Roth, 2001). The effect is often conceptualized as reflecting the attentional bias that anxious people have towards threatening stimuli (Mogg & Marden, 1990).

The specificity of the MST effect led researchers to test its sensitivity to intentionally feigned symptoms. For example, Buckley, Galovski, Blanchard, and Hickling (2003) administered the MST to actors instructed to feign PTSD, healthy controls, and patients with genuine PTSD ($N = 18$). Although the overall reaction time of actors ($n = 6$) was slower than that of healthy controls, the actors did not display the typical MST effect that was observed in the subsample of PTSD patients. It might also be important to note that in one of the previous studies, Buckley and his colleagues (2002) did not find the specific MST effect, but rather an overall slowing down in the PTSD group (Buckley, Blanchard, & Hickling, 2002). However, the authors concluded that “reaction-time-based information-processing tasks such as the Stroop may be harder to fake than face valid self-report instruments” (Buckley et al., 2003; p. 64). Recent studies have drawn on this argument so as to exclude feigning as a scenario. For example, Constans et al. (2014) administered the MST to veterans with PTSD who engaged or did not engage in symptom over-reporting measured using the Miller Forensic Assessment of Symptoms Test (M-FAST; Miller, 2001). The over-reporters exhibited a stronger rather than a weaker MST effect when compared with non-over-reporters. Based on these results, the authors concluded that symptom over-reporting in their sample reflected heightened distress rather than intentional feigning. Thus, the MST effect has been used to rule out feigning.

However, Cannon (2003) pointed out that “specificity” of MST effects may reflect current concerns that people have (e.g., anxiety symptoms). For example, Mathews and Macleod (1985)

demonstrated that patients who worried mostly about physical harm were especially slow in color naming words describing physical threat, while patients who mostly worried about social threat were especially slow in color naming social threat-related words. If true, one would predict that people who feign symptoms will exhibit delayed reactions specifically for words referring to feigning (e.g., lie, fake) to the extent that they have concerns about being detected. Cannon (2003) found, indeed, that students instructed to feign mild brain trauma symptoms performed significantly worse on words pertaining to feigning relative to honest controls. With this in mind, one could attribute Buckley et al.'s failure to obtain an MST effect in actors feigning PTSD to actors' indifference towards their task words. Perhaps, they were not familiar enough with PTSD and/or they may have found the possibility that their feigning might be detected not important.

In the current study, we wanted to test whether the MST effect occurs in students instructed to feign test anxiety. We chose test anxiety as a target for feigning because tests are an important feature of students' lives. Thus, students have to cope with frequent examinations and due to the pressure to perform well, they often experience heightened levels of stress during tests (Lawson, 2006). Accordingly, test anxiety, i.e., a chronic preoccupation (worry) with and physiological responsiveness to test situations, is a widespread problem among students worldwide (Nelson, Lindstrom, & Foelds, 2014). Recent studies (e.g., Yeo, Goh & Liem, 2016) reported that the prevalence estimates for test anxiety in students range from 10 to 40 percent. On the other hand, conditions such as test anxiety, dyslexia and ADHD, which are correlated (Nelson et al., 2014), are feigned on a non-trivial scale because doing so may result in incentives (i.e., special academic privileges, such as extra time for completing exams; e.g., Musso & Gouvier, 2012). Given their experience with stressful exams, one would therefore expect that students would find it relatively easy to feign an extreme form of test anxiety.

Students who suffer from high test anxiety exhibit higher reactivity to test-related stimuli (Keogh, Bond, French, Richards, & Davis, 2004). This suggests that individuals who truly suffer from test anxiety will display the MST effect. Indeed, these individuals have been found to provide longer reaction times on Stroop trials with relevant threat words (e.g., test, inept) than on trials with neutral words (e.g., MacLeod & Rutherford, 1992). In the current study, we examined whether this pattern can be simulated by non-test anxious students who feign test anxiety. To establish that participants did indeed comply with feigning instructions, we administered the Self-Report Symptom Inventory (SRSI; Merten, Merckelbach, Giger, & Stevens, 2016). The SRSI is a recently developed measure of symptom over-reporting and includes two main scales, one consisting of genuine symptoms and the other of pseudosymptoms scale. Both scales cover a wide range of psychological (e.g., anxiety, depression) and physical (e.g., pain) complaints. People with authentic complaints

endorse more genuine symptoms than pseudosymptoms, while the reverse is true for people who feign their complaints (Merten et al., 2016). We examined whether feigning instructions will cause heightened levels of pseudosymptom endorsement on the SRSI. Finally, we also included the Brief Symptom Inventory-18 (BSI-18; Derogatis, 2001) in order to screen for general distress among students.

According to the aforementioned findings, in the current study, we anticipated that the students, when not instructed about the task, would not show any abnormalities in their response latencies regardless of word type. In contrast, when instructed to feign test anxiety, we expected that students would exhibit a corresponding MST effect. Furthermore, we foresaw that the symptom endorsement on SRSI would be considerably amplified in the second session compared with the neutral, uninstructed, testing.

Methods

Participants

Using the effect size (*Cohen's d* = 1.15⁸) reported by Buckley et al., (2003) and with α set at 0.05 and β at 0.95, the lower bound sample size was found to be 13. Originally, the study included 28 participants from Maastricht University and Hogeschool Zuyd, The Netherlands. However, to ensure that participants did not in any way suffer from genuine test anxiety or any other closely related form of anxiety (e.g., social anxiety; Muris, 2002), potential participants were prescreened. People with high test anxiety typically report high trait anxiety levels, which reflects relatively stable general anxiety proneness (Spielberger, 1972; Keogh & French, 2001). Thus, we administered the Spielberger State-Trait-Anxiety-Inventory (STAI-T; Spielberger, Gorsuch, & Lushene, 1970), using a cutoff score of 46 (see Fisher & Durham, 1999). Potential participants who scored above that cut-point were excluded. The mean STAI score was 33.85 (*SD* = 6.95), and two participants were excluded. We decided to include all the participants who successfully passed the pre-screening for two reasons: First, the robustness of statistical tests usually requires a minimum of 20 participants per cell, and second, we could not predict how many participants would proceed with the study, and how many would withdraw their participation. As an additional check on high distress symptoms among participants, we administered the BSI-18 (see below). Four participants scored above the cutoff point of 11, and they were also excluded from further analysis.

The final sample included 22 undergraduates (14 women). Their mean age was 21 years (*SD* = 1.73). Students were given two vouchers of € 7.5 euros each, or two credit points. The study involved two sessions, each session lasting about one hour.

⁸ Comparison of the Actors and Non-anxious group. Corresponding *r* value being .49.

Measurements

Brief Symptom Inventory-18 (BSI-18; Derogatis, 2001). The Dutch version of the BSI-18 is designed to measure general distress and is often used as a screen for psychological problems (Meijer, de Vries & van Bruggen, 2011). Typical items are: "I feel like I am going to faint", "I feel worthless", and "I am so restless that I can't sit still". Respondents evaluate items on a 0 (not at all) to 4 (very much) Likert scale. Total scores range from zero to 72, with higher scores indicating higher general distress (Meijer et al., 2011). Cronbach's alpha of the BSI-18 in the current study was .72. We employed a cutoff of 11 in order to eliminate participants with clinically significant levels of distress (De Beurs, 2011).

The Self-Report Symptom Inventory (SRSI; Merten et al., 2016). The SRSI consists of two main scales with 50 symptom items each. One scale lists genuine symptoms (e.g., "I am often exhausted"), while the other scale lists pseudosymptoms (e.g., "On some days my left arm is good for nothing, on other days the right one is useless"). Both main scales include five subscales that gauge plausible (potentially genuine) or unlikely (probably non-authentic) manifestations of cognitive complaints, depression, pain, somatic, and anxiety/posttraumatic stress disorder symptoms. The rationale behind the SRSI is that honest participants/patients will endorse more genuine symptoms than pseudosymptoms, whereas this differential endorsement pattern will be absent in people who feign complaints. Previous studies on the psychometric merits of the SRSI (e.g., Merten et al., 2016) found supportive evidence for this rationale. For both scales total scores range from zero to 50. For the pseudosymptoms scale, a cut point of 9 has been proposed (Merten et al., 2016). At this cut point, sensitivity is .80 and specificity is 1.00, which is probably an overestimation caused by using a non-clinical sample (van Impelen et al., 2014). In the current study, Cronbach's alpha's for the full scale were .76 (first session; T1) and .93 (second session; T2).

The Modified Stroop task. This task was created using an E-prime application, version 2.0.10.353 (see pstnet.com), and its word stimuli were presented on a computer screen (41.1 by 40.2 cm) to the participants. Words were presented 1000 ms after a fixation cross 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 (blue, green, and red) on the screen. The reaction time was measured in milliseconds (ms; measurement error = 1 ms). We included three groups of words: neutral, anxiety-related, and test anxiety-related (see Supplemental table 3.1). Anxiety related words were derived from Becker et al. (2001), test anxiety-related words taken from Lawson (2006), whereas neutral words were derived from both articles (for more details, see Appendix 2). Each word was presented three times, in a different 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. The total number of trials was 108 (12 words x 3 word groups x 3 colors). Prior to the experiment trials, participants were presented with 15 (5 words x 3 colors) practice trials with neutral words (e.g., belt, candle, map). The reaction time for neutral, anxiety-related, and test anxiety-related words per condition was calculated as the average response latency to all stimuli presented from the corresponding category of words.

Procedure

We used a within-subject design in order to investigate whether a non-symptomatic sample is able to produce the MST effect considered to reflect genuine test anxiety complaints. To avoid carry-over effects, we always started with the honest session and approximately one week later the feign session took place. During the first session (T1), participants were instructed to respond honestly to the BSI-18 and the SRSI. They were also presented with the MST containing neutral, anxiety-related, and test-related words, and asked to name the color of the word as fast as possible, without focusing on the meaning of the words. Participants were first given the opportunity to practice with the task. After 15 practice trials, participants completed 108 active trials. During the second session (T2), participants were first given a vignette that described the following scenario (see Appendix 2 for English version⁹): a student with serious test anxiety has to leave school to care for his/her ill mother. Therefore, this student is not able to take an exam that is necessary for passing the academic year. The only chance to still pass the academic year is to have his/her best friend talk to the exam committee as if he/she were the person with test anxiety. So, the friend has to convince the committee of the seriousness of the test anxiety by completing an anxiety-test and feign test anxiety. If the exam committee is not convinced of the test anxiety, the person with high test anxiety has to retake the whole academic year. Participants read the vignette and were then informed that if they were convincing in feigning test anxiety, they would take part in a lottery in which they could win an extra bonus of € 20. After reading the vignette, participants filled in the SRSI again and had the MST once more. The participants did not receive any specific instruction how to respond to the task that followed. At the end of the second session, participants were asked to fill out a questionnaire. In this questionnaire, participants were asked to rate their understanding of the task, the plausibility of the vignette, and also their motivation, success, and their opinion about the task difficulty on a five-point Likert scale (anchors: 1 = Low; 5 = High). Our primary interest was in whether non-symptomatic participants exhibit the typical MST effect when they are instructed to feign. Because participants with raised anxiety levels might obscure MST results, we excluded students with BSI-18 scores that exceeded the cutoff.

Statistical analysis

⁹ A Dutch (original) version of the instructions can be obtained from the first author.

Because our data were skewed, we used non-parametric tests, notably the *Wilcoxon Signed Rank* test and *r* for effect size using the formula proposed by Rosenthal (1994; z/\sqrt{N} (number of observations), and *Mann-Whitney U test*. According to Cohen's (1988) criteria, a value of 0.1 = small effect, 0.3 = medium effect, and 0.5 = large effect. For clarity's sake, we report means and standard deviations.

Results

Exit questionnaire

Participants' understanding of the task was high ($M = 4.32$, $SD = 0.78$), and they found the task moderately plausible ($M = 3.77$, $SD = 0.75$). Participants did not experience great difficulty in carrying out the task ($M = 2.18$, $SD = 0.91$). Furthermore, students were moderately motivated ($M = 4.00$, $SD = 0.87$), and they judged the success of their performance as rather modest ($M = 2.77$, $SD = 0.61$).

Endorsement of SRSI Pseudosymptoms

Table 3.1. shows mean BSI-18 scores (only T1), mean SRSI (T1 and T2) scores, and Stroop latency data for the three categories of words (T1 and T2).

From T1 to T2, there was a significant increase in endorsement of both genuine symptoms and pseudosymptoms, $Z = 4.10$, $p < .001$, $r = 0.62$, and $Z = 4.11$, $p < .001$, $r = 0.62$, respectively.

Table 3.1.

Means and standard deviations of the BSI-18 scores (only T1), SRSI scores for genuine and pseudosymptoms (T1 and T2), and MST latency data on neutral, anxiety and test anxiety words in both conditions (N = 22).

| Measures | | T1 <i>M (SD)</i> | T2 <i>M (SD)</i> | Wilcoxon signed rank test <i>Z</i> | <i>r</i> |
|-------------------------|-------------------------|---------------------|---------------------|---------------------------------------------|----------|
| BSI-18 | | 3.77 (2.37) | / | / | / |
| SRSI | Genuine symptoms | 6.82 (4.72) | 31.27 (7.37) | 4.10** | 0.62 |
| | Pseudosymptoms | 1.23 (1.85) | 15.64 (8.94) | 4.11** | 0.62 |
| Stroop task (ms) | Neutral | 512.95 (65.68) | 745.63 (285.78) | 3.65** | 0.55 |
| | Anxiety | 517.10 (80.54) | 795.95 (267.54) | 3.94** | 0.60 |
| | Test-anxiety | 509.81 (75.68) | 790.39 (286.01) | 3.81** | 0.57 |

Notes: ** $p < .001$; T1 = honest way of responding; T2 = feigning test anxiety. *BSI-18* – Brief Symptom Inventory; *SRSI* – The Self-Report Symptom Inventory.

Importantly, at T2, the large majority of participants ($n = 13$; 77%) scored above the cutoff of 9 pseudosymptoms. Looking at separate subscales of pseudosymptoms, it was apparent that the rise in pseudosymptoms was particularly evident for the Anxiety/ Depression/ PTSD subscale ($Z = 4.12$,

$p < .001$, $r = 0.62$; see Supplemental table 3.2). As to SRSI subscales that involve genuine symptoms, raised symptom scores during T2 were also particularly evident for the PTSD/ Anxiety ($Z = 4.13$, $p < .001$, $r = 0.62$), and Depression ($Z = 4.12$, $p < .001$, $r = 0.62$). All in all, these significant increases in symptom endorsement indicate that our instructions to feign test anxiety at T2 were effective.

The MST Effect

Response latencies during T2 were significantly longer than those during T1 (see Table 3.2). This was true for neutral words: $Z = 3.65$, $p < .001$, $r = 0.55$, anxiety-related words: $Z = 3.94$, $p < .001$, $r = 0.60$, and test-anxiety related words: $Z = 3.81$, $p < .001$, $r = 0.57$.

Comparing latencies between word categories revealed no significant differences at T1 (all Z 's < 1.35 , all p 's $> .18$). However, at T2, differences emerged. The reaction time was significantly longer when anxiety-related words were presented compared with neutral words, $Z = 2.84$, $p = .005$, $r = .43$. The same pattern emerged when comparing latency for test-anxiety words with that for neutral words, $Z = 2.06$, $p = .039$, $r = .32$. The difference between anxiety and test-anxiety words remained insignificant ($p = .20$) (see Table 3.2).

Table 3.2.

Wilcoxon Signed Rank tests (Z) and significance levels for differences in reaction times between different pairs at T1 and T2.

| Pair of words | T1 | | T2 | | |
|------------------------|----------|----------|----------|----------|----------|
| | <i>Z</i> | <i>p</i> | <i>Z</i> | <i>p</i> | <i>r</i> |
| Neutral - Anxiety | -.50 | .61 | 2.84 | .005 | 0.43 |
| Anxiety – Test-anxiety | .27 | .78 | 1.28 | .200 | 0.19 |
| Neutral – Test-anxiety | 1.35 | .18 | 2.06 | .039 | 0.31 |

Notes: T1 = honest way of responding; T2 = feigning test anxiety.

The magnitude of the MST effect at T2 ($M = 44.75$, $SD = 92.65$) was significantly larger than at the T1 ($M = -3.15$, $SD = 31.72$), $Z = 2.06$, $p = .039$, $r = .31$ ¹⁰.

Furthermore, we compared the MST effect at T2 between participants who scored above ($n = 17$) and participants who scored under the cut-point on the SRSI ($n = 5$). The MST effect in SRSI high scorers ($M = 42.92$, $SD = 100.74$) was not significantly raised from not detected feigners ($M = 51.01$, $SD = 66.45$), Mann Whitney U test = 38.00, $z = -.35$, $p = .724$), but this might be due to the small sample size.

¹⁰ Due to the nature of the tests (ranks), the relative distance remains unchanged, hence, the same outcomes as for the MST effect at T2.

Discussion

The Modified Stroop Task (MST) has been proposed as a method for detecting pathology related attentional bias (e.g., Becker et al., 2001). Some authors (Buckley et al., 2003; Constans et al., 2014; but see Thomas & Fremouw, 2009) have argued that the MST involves a reaction time pattern that is difficult to simulate and that therefore flags genuine psychopathology (e.g., PTSD). For example, Constans et al. (2014; p. 83) reasoned that it is “unlikely that someone who feigns PTSD symptoms could (1) correctly deduce that PTSD is associated with slightly slower responding on combat words, and (2) adjust color-naming response time to create the desired MST effect.”

The current study examined whether an MST effect can be obtained when non-symptomatic students are instructed to feign test anxiety. Our findings can be summarized as follows. First, looking into the SRSI scores, our feigning instructions were successful, meaning that they triggered the expected symptom over-endorsement pattern in feigning condition. That is, at T2 (feigning session), participants endorsed significantly more genuine and pseudosymptoms on the SRSI than at T1 (honest control session). The dominant strategy among participants instructed to feign test anxiety was to raise all symptom scores. However, symptom over-endorsement was particularly pronounced for anxiety-related symptoms. Furthermore, using the cutoff of 9, the SRSI pseudosymptoms subscale was able to detect 77% of participants in the feigning condition (T2).

Second, and most importantly, when participants were instructed to feign test anxiety (T2) without being informed about the MST effect, they produced longer reaction times for all three types of words. This overall slowing down likely reflects the increased cognitive load that is caused by feigning (see Vrij, Fisher, Mann, & Leal, 2008). However, a similar response pattern can be produced by patients, who are processing emotionally provocative material. Even though our study design limits us in discussing how would a genuine test anxiety group respond to the MST, we do know from the literature (e.g., Lawson, 2006) that the expected response pattern would be the MST effect. Some researchers, such as Buckley and his colleagues (2002), even consider an overall delayed response latency to be a sign of genuine complaints. We disagree with this position. That is to say, genuine psychopathology might lead to delayed reaction times, but not each instance of a delayed reaction indicates psychopathology. For example, in one recent study (Boskovic, Hope, Orthey & Merckelbach, 2018), we applied the MST to three groups: participants who had experienced high impact life events, participants with low impact life experiences, and actors, with low impact experiences, but coached to feign PTSD-related symptoms. We found the most delayed RTs in the actors (feigning) group, which opposes the position of Buckley et al. (2002). However, claiming that longer RTs can help detecting feigning, would be equally an overstatement of the case. The issue here is of course that people who feign symptoms on the MST are likely to produce longer response

latencies (e.g., Boskovic et al., 2018), yet people who exhibit longer response latencies are not necessarily feigning. For example, Becker et al (2001) compared MST performance of people with generalized anxiety disorder (GAD), people with social phobia (SP), and non-anxious participants. People with GAD had the overall longest reaction time while responding to both GAD-related and SP-related words.

Additionally, besides overall longer response latencies, participants specifically exhibited longer reaction times when responding to anxiety-related and test anxiety-related words compared with neutral words (i.e., the MST effect). This differential response pattern was not evident during the honest control session (T1). Thus, instructed students were able to simulate the MST effect in the feigning condition, indicating that this effect does not rule out feigning as previous studies claimed (e.g., Buckley et al., 2003; Constants et al., 2014). Our findings suggest that feigning and the cognitive load it induces (e.g., Cannon, 2003) might result in a response latency pattern (i.e., the MST effect) that closely resembles a response pattern that is widely considered to be a sign of genuine symptomatology. Cannon (2003) speculated that the MST effect reflects current concerns rather than authenticity of the complaints. This interpretation makes sense, given that people who are motivated to simulate a certain type of psychological problem will be preoccupied with key words referring to the features constituting that problem. The additional cognitive load caused by instructions to feign certain problems may in itself be sufficient for longer response latencies.

Even though both Buckley et al. (2003) and Constants et al. (2014) observed pronounced MST effects for trauma-related words in PTSD patients, there are several issues that preclude a straightforward interpretation of their findings. For example, Buckley et al.'s study relied on a small sample of actors, PTSD patients, and non-anxious controls (each $n = 6$; $N = 18$). The findings by Constants et al. (2014) actually fit nicely with our study, although the authors come up with a completely different interpretation of their results. The MST effect in their study was most pronounced in PTSD patients who were over-reporting symptoms on the Miller Forensic Assessment of Symptoms test (M-FAST; Miller, 2001). Thus, it is a distinct possibility that the delayed reaction times on the MST in this group did reflect preoccupation with symptoms that were intentionally exaggerated rather than an attentional bias related to genuine PTSD.

A few limitations of the current study warrant comment. First, participants were tested during two different sessions, separated by a one-week interval. This could possibly have resulted in a learning effect, because the tasks in the two sessions shared the same content. Participants had to react differently to the same tasks during the two sessions, and we assumed that a learning effect was limited in this way. Still the within-subject design we employed might have influenced our results. However, if anything, one would expect that, during the retest, practice would have speeded up Stroop

performance, thereby lowering the chances of finding a specific MST effect. If the learning effect did occur, then one can assume that the actual differences between the two conditions might be larger than what we observed. Second, we instructed students to feign test anxiety, a condition that will be relatively easy for them to relate to. Participants' ratings of the difficulty to feign test anxiety indicated that students did not experience significant troubles in fabricating this particular issue. Whether students who are instructed to feign panic disorder, PTSD, or another less prevalent anxiety disorder also manifest an MST effect remains to be seen. Third, our findings are based on a non-symptomatic sample, therefore, we are restricted when making any conclusion regarding the responses of a genuine test anxiety group. We wanted to investigate whether providing a healthy, non-symptomatic sample with feigning instructions would produce a response pattern that is assumed to represent genuine test-anxiety complaints (i.e., the MST effect). However, comparing feigners with a symptomatic group would be a logical and a necessary next step in investigating the meaning of the MST effect. It might well be the case that both groups exhibit MST effects, but that those in genuine patients are much higher than those in feigners. Finally, the employed within-subject design limited the in-depth analysis of the diagnostic utility of the MST (e.g., the ROC), which should be further explored.

Future studies might also want to test patients instructed to feign certain conditions, which is not without ethical problems. Including people with genuine symptoms would imply that one would instruct half of them to deny problems and/or exaggerate problems, which is problematic. However, that would enable researchers to investigate more precisely the utility of both the Modified Stroop task and the SRSI in differentiating individuals who experience genuine symptoms from those who feign symptoms.

To sum up, we found that the MST effect could be relatively easy to elicit with feigning instructions (see also, Cannon, 2003). Thus, the MST effect is not only sensitive to genuine psychopathology (Kimble et al., 2009), but also to the additional cognitive load (e.g., preoccupation of healthy people with certain themes and topics) caused by feigning, which introduces noise in the strength of this effect across clinical groups (e.g., people with PTSD). This may explain why Kimble et al. (2009) found that the MST effect is sometimes difficult to replicate. It might well be the case that our current understanding of the MST effect is hindered by a publishing bias (e.g., file drawer problem) (Kimble et al., 2009). Future work involving comparison of genuine patients and participants who feign is necessary in order to closely examine the quality and the utility of the MST effect in the symptom validity assessment. However, given the fragility of any MST-related findings, one needs to be careful in making diagnostic decisions only using isolated thresholds, such as reaction time.

THE MODIFIED STROOP TASK AND THE SELF-REPORT SYMPTOM
INVENTORY IN FABRICATED SYMPTOMS
THE CASE OF HIGH IMPACT EXPERIENCES
STUDY 5

Abstract

The Modified Stroop Task (MST) effect refers to a prolonged reaction time (RT) in color-naming words related to an individual's disorder, compared to the RT when neutral words are presented. Some authors argue that it is impossible to feign, thus, its absence in people claiming symptoms might be a red flag for feigning. We tested whether the MST effect is robust against feigning and compared it with a symptom validity questionnaire (i.e., the Self-Reported Symptom Inventory (SRSI)). The SRSI includes two main scales tapping into genuine symptoms and pseudosymptoms. We included participants ($n = 22$) who reported current high impact of aversive experiences (High Impact), participants ($n = 24$) who reported current low impact (Low Impact) of aversive experiences, and actors ($n = 18$) who also reported low impact, but were instructed to feign a high impact experience (Simulators). We administered the MST, including impact-related, neutral, and feigning-related words, and the SRSI. We found no MST effect for impact-related words in the high impact group, or for feigning-related words in the simulators. Actually, the impact-related MST effect was more pronounced in simulators than in the high impact group. Also, simulators exhibited significantly longer RTs on all types of words and they also endorsed significantly more pseudosymptoms on the SRSI. Thus, the SRSI was a more sensitive measure of feigning than the absence of an MST effect. Our findings cast doubts on the popular idea that the MST can be used to differentiate between genuine and feigned complaints.

Key words: Modified Stroop task, Feigning, Over-reporting, SRSI, Simulators.

Detecting Feigned High Impact Experiences: A Symptom Over-report Questionnaire Outperforms the Emotional Stroop Task

A recent review revealed that the Stroop task (Stroop, 1935) remains a popular measure of executive function in Canada and US (Rabin, Paolillo, & Barr, 2016). The modified version of this task (i.e., the Modified Stroop task; MST) is a popular paradigm for exploring attentional bias in different conditions, such as posttraumatic stress disorder (PTSD; e.g., Buckley, Galovski, Blanchard, & Hickling, 2003). The MST includes the presentation of disorder-related and neutral words the colors of which have to be named as quickly as possible while ignoring the meaning of the words (e.g., Buckley, Blanchard, & Neill, 2000). 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. This delayed response is considered to be a sign of attentional bias (Elsesser, Sartory & Tackenberg, 2004).

As reaction-time measures do not rely on patient self-reports, it has been argued that the observed effects are less vulnerable to manipulation and reflect an objective measure of attentional biases associated with certain disorders (Buckley et al., 2003; Constans et al., 2014). In this vein, some authors argue that the presence of a delayed response for trauma-related words in presenting PTSD cases – the so-called *MST effect* – reflects genuine trauma symptoms, whereas the absence of the effect in those who claim trauma symptoms might be a sign of intentional feigning. As a result, several authors have speculated that the Modified Stroop method could function as a screening tool for differentiating between feigners and people truly suffering from PTSD (Constans et al., 2014; Beck, Freeman, Shipherd, Hamblen, & Lackner, 2001; Kaspi, McNally, & Amir, 1995).

To test this hypothesis, Buckley et al. (2003) used the Modified Stroop method 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 Stroop task (i.e., the MST effect). 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 and colleagues (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, Blanchard, & Hickling, 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? Constants 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) found similar reaction times 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.

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 than 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 test the conclusions of previous researchers (Buckley et al., 2003) and to compare the MST with the SRSI to examine the extent to which both methods discriminate between feigned and genuine PTSD-like symptoms. For ethical reasons, we were restrained from including students with full-blown PTSD in our sample. Therefore, we included a subclinical student sample comprising (i) individuals who reported currently experiencing a high psychological impact of previous aversive life events; (ii) individuals who reported a low impact of previous aversive life events (controls); 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 impact groups, 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., 2017). In order to more closely examine 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 has been shown to be related to over-endorsement of atypical items on symptoms measures (Merckelbach, 2004), thus, we wanted to investigate whether this trait would play any role in participants' performance.

Method

Sample

In total, we recruited 138 participants, 92 for the two impact groups and 46 actors. An a priori *G-power* analysis based on the effect (*Cohen's d* = 0.92¹¹) found in Buckley et al. (2003) revealed an ideal sample size of 51 participants. Due to applying elimination criteria ($n_{\text{impact groups}} = 29$ 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 (see Procedure part below). Participants were assigned to: 1) a High Impact group ($n = 22$); 2) a Low Impact group ($n = 24$), or 3) Actors (with low impact scores)

¹¹ Comparison of the Actors and PTSD group. Corresponding *r* value being .41.

($n = 18$). Participants' age ranged from 18 to 35 years, with the average being 19.83 ($SD = 3.61$) years. The majority was female (81.2%). Participants from the High Impact and Low Impact groups received two credit points (0.5 for pre-screening and 1.5 for the experiment) for participation, while Actors received £10 as compensation.

The High and Low Impact groups comprised participants recruited from the undergraduate participant pool in the Department of Psychology of University of Portsmouth. Actors were recruited from Faculty of Media and Performing Arts in the same university, as well as from local city theatres. 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).

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 et al., 1981), we created a list of 18 aversive events (e.g., end of intimate relationship, death of family member/close friend; see Supplemental table 3.4.). 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 the most aversive event from the list that had happened to them in the previous six months. To adhere to ethical principles, we did not include extremely traumatic events (e.g., rape, murder of a friend or relative), and excluded participants who added any of such events to the scale. In total, three participants were excluded based on this criterion.

Impact of Event Scale – Revised (IES-R; Weiss, 2004; *Cronbach's alpha* current study = .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 items are “Any reminder brought back feelings about it” and “I had waves of strong 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. We employed this scale as a screening measure for allocating participants to the High Impact group.

PTSD Checklist for DSM 5 (PCL-5; Weathers et al., 2013; *Cronbach's alpha* current study = .92). The PCL-5 includes 20 symptom items and participants are asked whether they experienced any of these symptoms in the previous month. We used this measure to exclude participants who may

be suffering from full-blown PTSD symptoms. 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 items are “Feeling very upset when something reminded you of the stressful experience” and “Avoiding memories, thoughts, or feelings related to the stressful experience”. We excluded participants who scored 3 (“often”) or higher on any of the 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 in different colors were presented on a computer screen (41.1 by 40.2 cm), 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 (blue, green, and red) 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, pain; Buckley et al., 2002; Moradi et al., 1999); Feigning-related (e.g., liar, fake, truth, scam; Cannon, 2003), and Neutral words (e.g., chair, wall, pencil, solid; Becker, Rinck, Margraf, & Roth, 2001). The three word categories did not differ in their average length ($F(2, 35) = .63, p = .54$). Each word was presented three times, in a different 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 x 3 word groups x 3 colors). Prior to the experimental trials, participants were presented with 15 (5 words x 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.

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 indicate 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 cut point of more than 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. “Yes” answers are summed, with higher scores indicating higher

levels of fantasy proneness. Typical items are: “As a child, I thought that the dolls, teddy bears, and stuffed animals that I played with were living creatures” and “I am never bored because I start fantasizing when things get boring.” The scale has shown good psychometric characteristics in previous studies (e.g., Merckelbach et al., 2001).

Procedure

Participants were informed about the study using posters and advertisements. The study included two parts: A pre-screening phase and the main experiment. To be eligible for the main experiment, participants needed to pass the pre-screening phase. This 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 one event that had happened to them in previous six months from the Aversive Events List (see Supplemental table 3.4.), with an option to add an aversive event that was not listed. After they had selected or reported an event, we asked participants to complete two questionnaires with reference to that event: the IES-R (Weiss, 2004) and the PCL-5 (Weathers et al., 2013).

Group assignment. To assign participants to the High Impact, Low Impact or Actors group, we used the following criteria. Participants were assigned to the High Impact group if they had a mean score above 1.5 on the total IES-R scale, but also on its three separate subscales. This score reflects the cutoff point with the best diagnostic accuracy of PTSD cases (Creamer et al., 2003). Participants were assigned to the Low Impact group if their mean total and subscale IES-R scores were below 1.5. The same rule was employed for the Actors group. Furthermore, the Low Impact group and Actors needed to have a PCL-5 mean score under 1.5, while scores for participants in the High Impact group did not exceed 3 on this scale.

Exclusion of participants. 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 any of the PTSD symptoms and/or their mean PCL-5 score was above three (“often”; $n = 3$); and 3) there was a discrepancy between the total mean score on the IES-R and the mean scores of its three subscales (Intrusion; Avoidance, and Hyperarousal; $n = 43$). 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 3.3.).

Experimental phase. Once participants passed the pre-screening, they were invited to join the study. Testing was performed in groups of up to 4-5 people. Participants in the High and Low Impact groups were first given the CEQ questionnaire to fill in and then completed the MST and the

SRSI. Actors first filled out the CEQ, and next they were asked to carefully watch a short video in which people who had had a traumatic experience talked about the consequences of their experience (video available on <https://www.youtube.com/watch?v=PFW4hYsYF>). Following this, they were given a vignette, presenting a case of Alex (gender neutral), who went through a rough break-up with a threatening, manipulative, and abusive partner. Alex files an official complaint and wants the court to formulate a restraining order for the ex-partner. In order for Alex to receive a restraining order, s/he needs to convince the authorities that s/he is currently under a high psychological impact (i.e., traumatization) from the experiences with his/her ex-partner. Actors were asked to take their time and try 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, they received an exit questionnaire concerning (e.g., “How convincing/educative/stressful was the vignette/video?” etc.) to which they could respond on a 5-point scale (1 = Not at all; 5 = Extremely). Furthermore, they were asked whether they used a particular strategy to feign high levels of 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.

Statistical analysis

Our data were analysed using the non-parametric replacement for ANOVA, *Kruskal-Wallis test*, the non-parametric replacement for independent *t*-test, *Mann-Whitney U* test, and the non-parametric replacement for a paired *t*-test, *Wilcoxon Signed Rank* test. We calculated *r* for effect size using the formula proposed by Rosenthal (1994; $z/\text{squared root of } N$). The effect size estimates for *Kruskal-Wallis test* were calculated following the formula provided by Green and Salkind (2005): $\eta_p^2 = \chi^2 / N - 1$. For clarity's sake, we present means and standard deviations.

Results

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 3.4.).

We ran a *Kruskal-Wallis test* involving five pre-screening scores (IES-R total scale and subscales, and the PLC-5 scale) as dependent variables and groups (High Impact, Low impact, and Actors) as independent variables. The results are summarized in the Table 3.3. All groups significantly differed on all of the pre-screening measures (χ^2 s > 29.62, $p < .001$). Follow up *Mann-Whitney U* tests indicated that High and Low Impact groups significantly differed from each other (all *Us* < 15.50, z s > 4.90, $ps < .01$), as well as the High Impact and Actors groups (all *Us* < 46.00, z s > 4.15, $ps < .01$). However, there were no significant differences between Low Impact and Actors

groups (all $U_s > 140.00$, $z_s < 1.95$, $p_s > .05$). These findings are in accordance with our pre-screening selection.

Exit Questions

Actors rated the video as moderately educational ($M = 3.84$, $SD = .71$), highly helpful ($M = 4.00$, $SD = .69$), and as somewhat stressful ($M = 2.22$, $SD = 1.35$). The vignette was rated as highly understandable ($M = 4.17$, $SD = .92$), as well as convincing ($M = 4.22$, $SD = .87$), and somewhat stressful ($M = 2.50$, $SD = 1.04$). Overall, the motivation of Actors to present themselves as Alex and effectively feign psychological impact was high ($M = 4.22$, $SD = 1.06$). Furthermore, 61% of them reported having a strategy in order to perform better on the task. The majority of actors tried to evoke their own memories or memories of their close ones going through a similar experience to the one presented in the vignette, and used these memories while reporting. Also, they tried to visualize the situation and to analyze the emotional response that should follow.

Fantasy Proneness

We ran a *Kruskal-Wallis* one-way ANOVA to test differences between three groups in their fantasy proneness scores. High Impact ($M = 11.63$; $SD = 4.58$), Low Impact ($M = 8.25$, $SD = 3.39$), and Actors group ($M = 13.28$, $SD = 4.00$) had significantly different scores on the CEQ, $\chi^2(2, N = 64) = 14.48$, $p = .001$; $\eta_p^2 = .23$. Follow-up *Mann-Whitney U tests* indicated significant differences between the two Impact groups ($U = 148.50$, $z = -2.55$, $p = .011$; $r = .37$), and between the Actors and Low Impact group ($U = 67.50$, $z = -3.79$, $p = .001$; $r = .58$). Actors and High Impact group did not significantly differ ($U = 169.00$, $z = -.79$, $p = .44$).

Modified Stroop Task

We ran a *Kruskal-Wallis* one-way ANOVA with types of words as dependent variables (impact-related, feigning-related, and neutral) and groups as an independent (High Impact, Low Impact, and Actors) variable. There was an overall significant difference between groups on all type of words: impact-related, $\chi^2(2, N = 64) = 15.83$, $p = .001$; $\eta_p^2 = .25$, feigning-related, $\chi^2(2, N = 64) = 14.66$, $p = .001$; $\eta_p^2 = .23$, and neutral words, $\chi^2(2, N = 64) = 13.54$, $p = .001$; $\eta_p^2 = .21$. *Mann-Whitney U tests* indicated that Actors had significantly longer response latencies on all three types of words than both Impact groups ($r_s > .50$; see Table 3.3.).

Next, we ran two *Kruskal-Wallis* ANOVAs with Impact-related and Feigning-related MST effect as dependent variables and group as an independent variable in order to compare the groups on their interference scores for Impact (Impact-related words RT – Neutral words RT) and Feigning words (Feigning-related words RT – Neutral words RT). Groups significantly differed with respect to the interference scores for Impact-related words ($\chi^2(2, N = 64) = 6.63$, $p = .036$; $\eta_p^2 = .10$).

Actors had a significantly higher Impact-related interference score than the High Impact group ($U = 107.00$, $z = -2.47$, $p = .013$, $r = .39$) but there was no significant between Actors and Low Impact group ($p > .05$), or between two Impact groups ($p > .05$). Group differences in interference scores for Feigning-related words did not reach significance, $\chi^2 (2, N = 64) = 3.26$, $p = .196$ (see Table 3.3.).

Table 3.3.

Pre-screening measures (IES-R total and subscales, and PCL-5 score), the Modified Stroop Task, and the SRSI results across groups.

| | Groups | | | |
|-------------------------------------|-----------------------------|----------------------------|------------------------|-----------------|
| | High Impact ($n = 22$) | Low Impact ($n = 24$) | Actors ($n = 18$) | χ^2 |
| | $M (SD)$ | $M (SD)$ | $M (SD)$ | ($2, N = 64$) |
| Pre-screening measures | | | | |
| IES-R | 2.05 (.30) | .53 (.29) | .75 (.39) | 44.23** |
| Intrusion | 2.30 (.58) | .61 (.37) | .83 (.36) | 43.65** |
| Avoidance | 2.13 (.64) | .58 (.38) | .81 (.45) | 40.02** |
| Hyperarousal | 1.62 (.54) | .35 (.34) | .48 (.38) | 37.60** |
| PLC-5 | 1.61 (.70) | .37 (.08) | .43 (.10) | 29.62** |
| The Modified Stroop Task | | | | |
| Impact-related words | 574.24(93.97) | 586.15(78.11) | 833.68(307.93) | 15.83** |
| Feigning-related words | 570.62(82.06) | 585.16(84.45) | 803.72(263.84) | 14.66** |
| Neutral words | 587.74(101.30) | 588.99(97.84) | 788.10(222.83) | 13.54** |
| Impact-related interference score | -13.50 (35.30) | -2.85 (37.86) | 45.58 (132.35) | 6.63* |
| Feigning-related interference score | -17.12 (46.73) | -3.83 (34.57) | 15.61 (82.72) | 3.26 |
| SRSI | | | | |
| Genuine symptoms | 21.00 (9.97) | 16.08 (7.06) | 39.84 (7.34) | 32.76** |
| Pseudosymptoms | 6.96 (6.45) | 3.96 (3.11) | 25.00 (11.54) | 34.07** |

Note: ** $p < .01$; * $p < .05$. *Pre-screening measures:* All scores given on a 4-point Likert scale (0 to 4).

The Modified Stroop task: Reaction time was measured in milliseconds. *SRSI:* Scores present the number of endorsed symptoms (0 to 50).

Furthermore, we performed series of *Wilcoxon Signed Rank* tests in order to investigate the differences in response latencies between Impact-related, Feigning-related, and Neutral words *within* each group. The differences in response latencies did not attain significance in any of the groups and so, we did not obtain a significant standard MST effect in the High Impact group or a significant feigning MST effect in the Actors group (see Table 3.4.).

Table 3.4.

Wilcoxon Signed Rank test (Z) and p levels for differences in reaction times (RT) between word categories within each group.

| Word type pairs | Group | | |
|-----------------------------------------------|--------------------------------------|-------------------------------------|---------------------------------|
| | High Impact <i>z</i> (<i>p</i>) | Low Impact <i>z</i> (<i>p</i>) | Actors <i>z</i> (<i>p</i>) |
| Impact-related words – Neutral words | -1.70 (.09) | -.48 (.63) | -1.89 (.06) |
| Feigning-related words – Neutral words | -1.51 (.13) | -.34 (.73) | -.76 (.44) |
| Impact-related words – Feigning-related words | -.18 (.86) | -.03 (.98) | -1.20 (.23) |

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.

To test the sensitivity of the MST in detecting over-reporting, we compared the High Impact, Low Impact, and the Actors group, and 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 Impact group and Actors was below the chance level (41% and 34%, respectively), and somewhat above it for the Low Impact group (58%; see Table 3.5.).

Self-Report Symptom Inventory

Table 3.3. shows means and standard deviations of the three groups on the main SRSI scales and relevant subscales. A *Kruskal-Wallis* one-way ANOVA showed that groups significantly differed on both genuine and pseudosymptom scales of the SRSI ($\chi^2(2, N = 64) = 32.76, p < .001, \eta_p^2 = .52$; $\chi^2(2, N = 64) = 34.07, p < .001, \eta_p^2 = .54$, respectively). Follow up *Mann-Whitney U* tests indicated that Actors scored significantly higher than High and Low Impact groups ($U_s < 12.50, ps < .001, rs > .69$). However, the Impact groups did not differ from each other in their endorsement of either genuine or pseudosymptoms ($U_s > 174.50, ps > .05^{12}$).

We also wanted to see whether participants differed specifically with regard to their symptom endorsement in the subscales of genuine and pseudo PTSD/Anxiety/Depression symptoms. Overall, Actors scored significantly higher on both subscales than the Low Impact ($U_s < 12.50, ps < .001; rs$

¹² The difference between two impact groups was marginally significant on genuine symptoms scale ($p = .049$), however, after correcting the alpha levels (p level being .017), the difference was no longer significant.

> .81) and High Impact group ($U_s < 25.00$, $p_s < .001$; $r_s > .75$). High and Low Impact groups did not significantly differ ($U_s > 198.00$, $p_s > .13$) with regard to these scores (see Supplemental table 3.5.).

Table 3.5.

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 Impact | 9 (41%) | 13 (59%) | 16 (73%) | 6 (27%) | 41% | 73% |
| Low Impact | 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.

We used the standard cut point of more than nine pseudosymptoms (Merten et al., 2016). Table 3.5. shows the number and percentages of correct classifications and false positives (i.e., High Impact and Low Impact participants misclassified as feigners) and false negatives (i.e., Actors misclassified as High Impact participants). 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 Impact group, and 3 (12.5%) from the Low Impact group were misclassified by the SRSI as feigners.

Explorative analysis

To explore whether fantasy proneness was related standard and feigning MST effects and the genuine and pseudosymptom scales of the SRSI, we calculated Pearson product-moment correlations including only the two Impact groups (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 ($r = .03$; $p = .84$). 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 *Mann-Whitney U* test to investigate possible differences in fantasy proneness between participants from the Impact groups who were misclassified by SRSI as feigners (24% of the High and 12.5% of the Low Impact group, $n = 9$) 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 ($M = 13.78$, $SD = 2.82$; U test = 55.00, $z = -3.10$, $p = .002$, $r = .45$), than participants who were correctly classified ($M = 8.92$, $SD = 4.09$)¹³.

¹³ The data and analysis are available on Open Science Framework (<https://osf.io/v3uhq/>).

Discussion

We compared the efficacy of the Modified Stroop Task and a symptom over-report questionnaire (SRSI) in detecting feigned high impact life 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, unlike in Buckley's study (Buckley et al., 2003), our High Impact group did not manifest longer reaction times for Impact-related words than for other categories of words.

Second, in line with Constants et al. (2014), who found patients with over-reporting tendencies to exhibit overall longer response times than PTSD patients without over-reporting style, we found that Actors exhibited "overall slowing down" on all types of words (impact-related, feigning-related, and neutral). However, Actors did manifest the most pronounced interference score for Impact-related words (Impact-related words RT – Neutral words RT), significantly higher than the interference score of the High Impact group. Thus, Actors did not arbitrarily prolong their response latencies, but rather consciously identified words that might be triggering for the character they were playing (Alex). This finding contradicts the assumption of low face validity of reaction time measures (e.g., MST), often taken as the main reason behind applying such measures in a forensic context (Buckley et al., 2003)

Third, actors failed to exhibit 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 a priori reason to assume that in real life, dedicated feigners do have strong connotations when confronted with such target words. They too might 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 Impact and Actor groups were below chance level, and slightly above for the Low Impact 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 pseudosymptoms of the SRSI, compared with participants in the High and Low Impact groups, 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 Impact group and 12.5% of the Low Impact group were classified by the SRSI as over-reporters. These might be false positives, but taking into account the sensitivity 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 two impact groups on SRSI 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). CEQ scores significantly correlated with endorsement of both genuine and pseudo SRSI symptoms, which confirms the previously established connection between high fantasy proneness and atypical responding (Peace & Masliuk, 2011). Interestingly, participants in both Impact 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 groups who exhibited elevated levels of fantasy proneness, engaged in over-reporting of their symptoms.

A few limitations of the current study warrant comment. First, 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, which could have led to inaccurate group assignment of participants to groups. Third, taking into consideration the events participants reported, it was difficult to select words for the MST that were uniquely associated with every possible aversive event (e.g., fight with a friend, accident, or a close friend or family member dying). Therefore, it is possible that the words that we included did not trigger all participants' current impact. Fourth, despite having a sizeable initial sample, the number of participants left after applying the exclusion criteria might have led to underpowered results. On that note, 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. There is literature suggesting that actors tend to report more aversive events than non-actors (e.g., Thomson, Keehn, & Gumpel, 2009). However, whether these events are experienced as "traumatic" mainly depends on actors' evaluation (Elal & Slade, 2005). 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., Thomson, Keehn, & Gumpel, 2009).

In sum, 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 events from those who fabricated such impact. We believe that our results, along with those reported earlier (Boskovic et al., 2018), indicate that the MST does not allow for an accurate differentiating between honest and feigned symptom presentations. This is likely to be because MST effects are fragile, and if this is the case then it renders the MST unsuitable as a diagnostic tool. On a positive note, although not perfect, the SRSI might be a promising alternative for differentiating between honest and feigned symptom presentation.

CHAPTER 4

THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED PHYSICAL AND PSYCHOLOGICAL SYMPTOMS

This chapter draws on the following paper:

Boskovic, I., Merckelbach, H., Merten, T., Hope, L., & Jellicic, M. (in press). The Self-Report Symptom Inventory as an instrument for detecting symptom over-reporting: An explorative study with instructed simulants. *European Journal of Psychological Assessment*.

Abstract

The recently developed Self-Report Symptom Inventory (SRSI) intends to provide an alternative approach to the detection of symptom over-reporting. Unlike other measures, the SRSI includes both non-existent symptoms (i.e., pseudosymptoms), and genuine symptoms. Previous research using the German SRSI showed that people who exaggerate their complaints over-endorse both types of symptoms. In the current simulation experiment, we tested whether the Dutch and English SRSI are effective in identifying over-reporting by comparing SRSI scores of an honest group ($n = 51$) with those of two experimental simulators groups (pain, $n = 54$; anxiety, $n = 53$). The pain and anxiety simulators endorsed significantly more genuine symptoms and pseudosymptoms than honest participants ($\eta_p^2 = .50$ and $\eta_p^2 = .30$, respectively). Furthermore, pain and anxiety over-reporters specifically over-endorsed symptoms corresponding to their simulation instructions (*Cohen's* d s > 0.77). Using the recommended cutoff score, the SRSI detected 48% of pain over-reporters and 73% of anxiety over-reporters, with areas under the curve (AUCs) ranging from .88 to .91. These results indicate that the SRSI is a promising tool for identifying over-reporting, but further research with clinical samples is needed.

Key words: SRSI, over-reporting, simulation, symptoms.

The Self-Report Symptom Inventory as an instrument for detecting symptom over-reporting: An exploratory study with instructed simulators

Intentional fabrication or exaggeration of symptoms occurs on a non-trivial scale and in a variety of settings (e.g., medico-legal evaluations, psychiatric assessment of asylum seekers, rehabilitation clinics). Base rate estimates close to or exceeding the 50 percent mark have been found in litigating patients after whiplash injury (Schmand et al., 1998), criminal defendants (Ardolf, Denney, & Houston, 2007), social security disability claimants in the U.S. (Chafetz, Abrahams, & Kohlmaier, 2007; Miller, Boyd, & Cohn, 2006), students with claimed attention deficit disorders, but without learning disability (Sullivan, May & Galbally, 2007), U.S. veterans claiming mild traumatic brain injury (Armistead-Jehle, 2010), and asylum seekers in a psychiatric hospital (Van der Heide, Boskovic & Merckelbach, 2017). Research in psychosomatic rehabilitation clinics in Germany revealed a non-trivial occurrence of over-reporting among their patients (e.g., Göbber, Petermann, Piegza, & Kobelt, 2012). In a recent survey, Santamaría, Ramírez, and Ordi (2013) asked 161 Spanish medical doctors to estimate the percentages of symptom exaggeration in patients who claim temporary disability benefits. Mean estimates of 50 percent or higher were reported for patients who presented with whiplash injury, fibromyalgia, chronic cervicgia, depression, and anxiety disorders.

The first tests for detecting symptom exaggeration date back to the 1940s, but it was the introduction of the forced-choice performance validity tests (PVTs) in the 1980s that provided a major impetus to this field (for reviews see Sollman & Berry, 2011; Lippa, 2017). The rationale behind these tests is that when given a simple task in which a person can obtain at least 50% of correct answers just by guessing (i.e., forced-choice task), below chance level performance will be indicative of *intentional* avoidance of correct answers (Bianchini, Mathias, & Greve, 2001; Merten & Merckelbach, 2013).

Self-report scales that intend to measure the validity of symptom presentation have an even longer history and started with an early study of Hartshorne and May (1928; see also Ben-Porath, 2012). Next to validity scales embedded within standard measures of psychopathology such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) and the Personality Assessment Inventory (PAI; Morey, 2007), several stand-alone instruments for detecting symptom over-reporting have been developed. These tests are predicated on the idea that people who exaggerate their complaints will tend to over-endorse symptoms, even when these symptoms are extreme, bizarre or quite rare across different cultures and settings (e.g., Boskovic, van der Heide, Hope, Merckelbach, & Jelicic, 2017).

According to several surveys (Martin, Schroeder, & Odland, 2015; Dandachi-FitzGerald, Ponds, & Merten, 2013), one widely used measure to assist in this type of symptom validation is the

Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; Merckelbach & Smith, 2003; for a meta-analysis see Van Impelen, Merckelbach, Jelicic, & Merten, 2014). The SIMS consists of 75 items describing bizarre symptoms (e.g., “Sometimes my headaches are so strong that my feet hurt”) or approximate answers (“If you have US\$1.50 and I take fifty cents away, you will have 75 cents left”) aimed at detecting feigned neurological, cognitive, and psychiatric problems. The SIMS has satisfactory psychometric properties in various samples. In a recent meta-analysis on the cross-cultural assessment of feigning, the SIMS was found to possess high classification accuracy with low variability across cultures (Nijdam-Jones & Rosenfeld, 2017; see also Boskovic et al., 2017).

However, one limitation of the SIMS is its heavy reliance on bizarre symptomatology, which may make it readily identifiable as a symptom validity instrument. Another drawback is that the SIMS focuses on types of feigning that might be prevalent in the criminal context (e.g., amnesia, psychosis, low intelligence), but are less common in civil litigation (e.g., personal injury or workers’ compensation claims). In civil forensic settings, less blatant forms of symptom exaggeration are expected to occur (e.g., exaggerated anxiety and pain problems, mild cognitive impairment; Merten, Merckelbach, Giger, & Stevens, 2016).

With these limitations in mind, a new stand-alone measure of over-reporting, the Self-Report Symptom Inventory (SRSI, Merten et al., 2016; Merten, Giger, Merckelbach, & Stevens, in press), was developed. The SRSI is a 107-item questionnaire with a dichotomous response format (*true* or *false*). Items are simple and easy to understand; they require a low reading level (end of primary school) and it takes 10 to 15 minutes to complete the instrument.¹⁴ Importantly, the SRSI includes both genuine and pseudosymptoms that gauge diagnostic problems frequently occurring in personal injury cases. Thus, one main scale taps into plausible symptoms (e.g., “I am often exhausted”) and the other main scale involves unlikely symptoms (e.g., “On some days my left arm is good for nothing, on other days the right one is useless”). The two main symptoms scales include five subscales that either cover plausible (potentially genuine) cognitive, depressive, nonspecific somatic (e.g., fatigue, sleep disorders, physical exhaustion), anxiety, and pain symptoms, or unlikely (probably non-authentic) manifestations of cognitive complaints, pain, motor, sensory, and depression/anxiety symptoms. The range of possible subscale scores varies between 0 and 10 for each of the ten subscales, and between 0 and 50 for both main scales.

¹⁴ The SRSI manual (Merten et al., in press) stipulates as preliminary criterion that patients/respondents should have a reading level equivalent to eight years of formal schooling in the language that the SRSI version is given in. Potential limits of the SRSI in people with intellectual disability, developmental dyslexia, mild dementia or aphasia have yet to be tested. One explorative study with patients of an Austrian memory clinic (described in Merten et al., in press) suggested that the SRSI does not yield false-positive classifications in this particular group (in which early or mild dementia is prevalent).

The empirical item selection procedure and the results obtained with the original German SRSI were described by Merten et al. (2016). So far, the SRSI was tested in various European countries (total $N > 1500$). Thus, English, Dutch, Norwegian, French, Portuguese, and Russian adaptations of the SRSI are currently available, and work on an Italian version is in progress. Giger and Merten (in press) studied the equivalence of the French and German SRSI in a sample of Swiss bilingual respondents and found no systematic differences between the two versions, neither in an honest condition nor in an experimental simulation design.

On the basis of 520 SRSI protocols from diverse samples (civil forensic assessment, normal controls, participants of experimental analogue studies, forensic in-patients, criminal offenders; Merten et al., 2016), the reliability of the genuine symptom and pseudosymptom scores (*Cronbach's alpha*) was estimated to be 0.95 and 0.92, respectively. At the current stage of scale development, it is the pseudosymptom scale of the SRSI that is used for identifying symptom over-reporting, whereas the genuine symptom scale can only be interpreted qualitatively. A Receiver Operating Characteristics (*ROC*) analysis with SIMS scores as external criterion suggested two cutoffs (Merten et al., 2016): a cut scores of >6 pseudosymptoms that might be used for screening purposes (with a maximum of 10% false positives) and a more stringent cut score of >9 for standard diagnostic purposes (with a maximum of 5% false positives). Sensitivity and specificity estimates for the screening cut score were found to be .83 and .91, respectively and those for the standard cut score were found to be .62 and .96, respectively.

Follow-up research provided further evidence for the notion that honest participants/patients endorse genuine symptoms, but only few pseudosymptoms, whereas this differential endorsement is largely absent among people who engage in over-reporting (Boskovic, Hope et al., 2018; Boskovic, Biermans et al., 2018; Merten et al., in press). Helvoort, Merckelbach, and Merten (under review) recently administered the SRSI to forensic inpatients who had no motive to engage in symptom distortion. They found a generally low endorsement of pseudosymptoms among patients, whereas genuine symptoms were endorsed at moderate levels. This study, together with other research (e.g., Giger & Merten, in press) suggests that the pseudosymptom scale of the SRSI is not sensitive to real psychopathology.

So far, studies on the SRSI did not look into whether different types of over-reporting instructions generate different patterns of endorsed pseudosymptoms. The critical question here is whether subscales of pseudosymptoms are differentially sensitive to simulation attempts targeted at different psychopathological domains (e.g., pain versus anxiety). With these considerations in mind, we tested the sensitivity of the SRSI to simulated pain or anxiety-related problems. For both pain and anxiety problems, clinicians largely depend on self-reports of patients (Kucyi, Scheinman, & Defrin,

2015). Thus, adding an instrument such as the SRSI to self-report inventories designed to measure pain or anxiety might lead to a more accurate detection of patients who over-report symptoms or feign those conditions. In the current study, we tested both honest participants and participants who were, on the basis of a brief scenario, instructed to over-report anxiety or pain symptoms. This way we were able to compare overall SRSI scores of honest and instructed simulators and also, and most importantly, to look into the sensitivity of the pain and anxiety scales to pick up different forms of symptom over-reporting.

Method

Participants

A priori *G*Power* analysis suggested a minimum sample size of 147 participants, based on alpha set at 0.05, beta at 0.95, and an $\eta_p^2 = .33$ (Boskovic, Hope et al., 2018). Our final sample included 158 participants. The mean age of participants was 21.36 years ($SD = 2.61$; range 18 to 34 years), and 83% were female. Participants were randomly assigned to one of three groups: Honest comparison group ($n_1 = 51$); simulated pain group ($n_2 = 54$), and simulated anxiety group ($n_3 = 53$). Participants were psychology bachelor students (83%) or studying at a postgraduate level (17%). Because of the international background of Maastricht University students, all participants had the option to answer questions in English or Dutch: 68% opted to respond in English while the remainder responded in Dutch. The distribution of Dutch students was similar across groups: 31.4% in the honest control group, 33.4% in the simulated pain, and 32% in simulated anxiety group, $\chi^2(2) = .05, p = .97$. Participants were compensated with either a financial reward of 7.50 euros or one research participation credit.

Measures

The Self-Report Symptom Inventory (SRSI; Merten et al., 2016; in press). As said before, the SRSI includes two main scales: genuine symptoms and pseudosymptoms. 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”. In the current study, *Cronbach’s alpha’s* for both scales were .95. There are five subscales addressing potentially genuine symptoms in the domains of: 1. Cognitive problems; 2. Depression; 3. Pain; 4. Nonspecific somatic complaints; 5. PTSD/Anxiety; and five subscales relate to pseudosymptoms, such as: 1. Cognitive/memory problems; 2. Neurological (motoric complaints); 3. Neurological (sensory complaints); 4. Pain, and 5. Anxiety/Depression. In this study, we specifically focused on scales that tap into genuine and pseudopain symptom and scales that gauge genuine symptoms and pseudosymptoms of anxiety, depression, and posttraumatic stress (i.e., “anxiety-related” scales).

In addition to genuine and pseudosymptom items, the SRSI contains a safeguard against careless or random responding. That is, a 5-item supplementary scale describes aspects of perfect health or mental abilities (e.g., “My mental abilities are excellent”; “I am happy and without worries”). In case of careless or random responding, people would endorse these items, which would then contradict their endorsement of symptoms. In the current study the honest participants scored relatively high on the perfect health items (and relatively low on symptom items), whereas the reverse pattern emerged in instructed simulators. In short, there were no obvious cases of random or careless responding.¹⁵

To make the SRSI less salient, we asked participants also to complete a number of other items, including questions concerning their motivation (see below) and the **Brief Symptom Inventory-18** (BSI-18; Derogatis, 2001). The BSI-18 is designed to measure general distress and is often used as a screen for psychological problems (Meijer, de Vries, & van Bruggen, 2011). Typical items are: “I feel pain in the heart or chest” and “I feel numbness or tingling in parts of my body.”. Respondents evaluate the 18 items on a 0 (not at all) to 4 (very much) Likert scale. Total scores range from zero to 72, with higher scores indicating higher general distress (Meijer et al., 2011). Scores above 11 are often considered as reflecting clinically raised levels of psychological distress (De Beurs, 2011; Merckelbach, Langeland, de Vries & Draijer, 2014). *Cronbach’s alpha* of the BSI-18 was .83. In the current study, we did not look into BSI-18 scores of the instructed simulators, but focused on those of the honest comparison group. To investigate the sensitivity of the SRSI to genuine symptomatology, we did explore how honest comparison participants with raised BSI-18 scores (>11) and those with low scores (< 11) performed on the SRSI main scales.

Procedure

Students signed up for the study via Research Participation System (SONA, <http://maastricht-fpn.sona-systems.com>). They received a link directing them to the study that was conducted using Qualtrics. Participants could choose whether to complete the English or Dutch version. After answering demographic questions, they were asked to fill out the BSI-18 and the SRSI. All participants were asked to respond honestly to all BSI-18 items. The instructions for completion of the SRSI were randomized. One group of participants were instructed to complete the questionnaires as honestly as possible (honest comparison group), whereas the other two groups received one of two scenarios describing a character who either misrepresented pain (simulated pain) or anxiety (simulated anxiety). The text about simulated pain described a situation in which a person was hit by a car, but did not suffer any injury from the accident. However, the person, in order to “get back” at

¹⁵ More precisely, honest participants ($M = 3.19$, $SD = 1.62$) scored significantly higher on the five perfect health items than pain ($M = .31$, $SD = .72$) and anxiety ($M = .62$, $SD = 1.15$) simulators, $F(2, 155) = 88.22$, $p < .001$, $\eta_p^2 = .53$, who generally scored well below 1.

the driver for being rude, decided to make an official report of the accident, to see an insurance doctor and to simulate serious pain. The text about simulated anxiety concerned a police officer who needed time off, but whose boss was not interested in hearing his reasoning. Thus, the police officer decided to go to the police doctor and to simulate suffering from a trauma (see Appendix 3). Participants in the instructed simulation groups were asked to take on the role of the protagonist and, with that in mind, complete the measures in a convincing manner. Following the completion of the survey, participants were given a debriefing written form that explained the true purpose of the research¹⁶. There were no missing data due to the rule that we implemented in our Qualtrics procedure: participants could not go forward with the task when a previous question had been left unanswered. Of course, participants still had the option to exit the link at any given moment. Three participants stopped the study while answering the initial questions (demographic items), and their data were not used.

Statistical Approach

We performed a Multivariate Analysis of Variance (*MANOVA*) in order to evaluate differences between the three groups (honest comparison group, pain simulators, and anxiety simulators) in their scores on the SRSI main scales (genuine symptoms and pseudosymptoms scale). We also conducted a *MANOVA* investigating the differences between the two simulator groups on pertinent SRSI subscales (genuine pain symptoms, genuine anxiety symptoms, pseudopain symptoms, and pseudoanxiety symptoms). *Bonferroni post-hoc tests*, together with the *t*-tests were employed for the follow-up examination of the potential differences between groups. We used *Cohen's ds* and η_p^2 for effect size calculations. We also performed *ROC* analysis to calculate the area under the curve (*AUC*).

Results

Exit questions

Participants rated their motivation and clarity of instructions using five-point scales (anchors: 1 = “Not at all”, 5 = “A great deal”). Instructed simulators were also asked about the plausibility of the scenarios, how convincing they thought they were in reporting target symptoms, and how difficult they found it to imagine themselves in the described role. For the full sample ($N = 158$), mean values for motivation and clearness of instructions were $M = 3.82$ ($SD = .72$) and $M = 4.44$ ($SD = .72$), respectively. Groups did not significantly differ with regard to these variables, $F_s(2, 155) < 2.48$, $p_s > .09$. Overall, instructed simulators ($n = 103$) rated the scenarios as plausible, and their performance as convincing: $M = 3.85$ ($SD = .07$) and $M = 3.41$ ($SD = .08$), respectively. Also, they found the task

¹⁶–The study was preregistered on the OSF platform (<https://osf.io/jpnqr/>). Originally, we intended to include three instructed simulation groups. Yet, after consultation with practitioners in the field, we decided to focus only on the pain-related and anxiety-related scenarios to keep a clear distinction between psychological and physical symptoms.

moderately difficult, $M = 2.57$ ($SD = .09$). The two simulators groups did not significantly differ in their ratings, $ts(97) < .28$, $ps > .77$, *Cohen's ds* < 0.02 . Eight instructed simulators (8%) had relatively low plausibility and/or motivation scores (≤ 2). Removing them from the analyses to be described below did not change the basic findings, thus, their data were retained in the main analyses.¹⁷

We also asked instructed simulators whether they currently suffered from symptoms similar to those described in the scenarios. Four participants in the simulated pain group reported back, stomach, shoulder, and neuropathic pain, and seven participants in the simulated anxiety group suffered from anxiety and fatigue (total $n = 11$). We ran the analyses to be presented below with and without these participants, but the results were similar. *Welch's t*-test indicated no significant differences between these participants and the other instructed simulators on the genuine and pseudosymptoms scales of the SRSI, *Welch's t* (14.25) = $-.93$, $p = .37$, and *Welch's t* (15.22) = $.08$, $p = .94$, respectively.

BSI-18 and SRSI scores in the honest comparison group

We next tested whether the BSI-18 scores of honest comparison participants were related to their SRSI scores. The BSI-18 scores and the genuine and pseudosymptoms scales were correlated, *Pearson's r* = $.59$, $p < .001$, and *Pearson's r* = $.62$, $p < .001$, respectively. Consequently, we checked whether participants with clinically raised BSI-18 scores (> 11 ; $n = 34$) produced a similar proportion of pseudosymptoms as those with relatively low BSI-18 levels (≤ 11 ; $n = 17$). The difference between the number of endorsed genuine symptoms and the number of endorsed pseudosymptoms¹⁸ between these groups did not reach significance: *Welch's t* (44.66) = 1.98 , $p = .054$, *Cohen's d* = 0.54 . Although the moderate effect size suggests that participants with raised levels of psychological problems endorsed relatively more pseudosymptoms than participants with low levels of psychological problems, one should keep in mind that the mean number of endorsed pseudosymptoms remained low in both groups (< 3) and were well below the cutoffs.

Main SRSI scales: Honest participants versus simulators

To test whether groups differed in their SRSI scores, we conducted a *MANOVA* with a Group (honest comparison group vs. simulators of pain vs. simulators of anxiety) as an independent factor and SRSI main scales (genuine and pseudosymptoms) as dependent variables. There was an overall significant effect of Group, $\lambda = .48$, $F(4, 308) = 33.40$, $p < .001$, $\eta_p^2 = .30$. As can be seen in Table 4.1, simulators scored higher on both genuine and pseudosymptoms than honest comparison participants.

¹⁷ For example, we evaluated the difference between honest participants and motivated simulators (participants with low motivation/plausibility scores removed) in number of pseudosymptoms and found a significant effect, $F(2, 147) = 34.23$, $p < .001$, $\eta_p^2 = .32$, that is similar to that reported in Table 1 (full subsample of simulators).

¹⁸ For both groups, we calculated the following difference: (number of endorsed genuine symptoms – number of endorsed pseudosymptoms).

SRSI in Fabricated Physical and Psychological Symptoms

Table 4.1.

Endorsement of genuine and pseudosymptoms on the SRSI of honest comparison group, pain simulators, and anxiety simulators.

| SRSI scales | Honest comparison ¹ <i>M</i> (<i>SD</i>) | Pain simulators ² <i>M</i> (<i>SD</i>) | Anxiety simulators ³ <i>M</i> (<i>SD</i>) | <i>F</i> (2, 155) | Effect size (η^2) | <i>Cohen's d</i> |
|-----------------------|----------------------------------------------------------|--------------------------------------------------------|-----------------------------------------------------------|----------------------|-----------------------------|-------------------------------------------------------------------|
| Genuine | 12.34 (7.56) | 30.26 (9.43) | 33.77 (10.66) | 78.73* | .50 | 2.09 ^{1,2} 2.32 ^{1,3} 0.35 ^{2,3} |
| Pseudosymptoms | 2.68 (2.90) | 12.87 (10.50) | 17.38 (12.00) | 33.08* | .30 | 1.33 ^{1,2} 1.68 ^{1,3} 0.40 ^{2,3} |

Note: * $p < .001$; “^{1,2}” indicates contrast between the honest comparison group and pain simulators, “^{1,3}” indicates contrast between the honest comparison group and anxiety simulators, “^{2,3}” indicates contrast between the pain simulators and anxiety simulators.

Specifically, *Bonferroni post-hoc* tests indicated that the two simulators group (simulated pain, $M = 30.26$, $SD = 9.43$, and simulated anxiety, $M = 33.77$, $SD = 10.66$) endorsed significantly more items from the genuine symptom scale than the honest group, $M = 12.34$, $SD = 7.56$, $ps < .001$. *Cohen's d* was 2.10 for the contrast between honest participants and pain simulators, and 2.32 for the contrast between honest participants and anxiety simulators. The two simulator groups did not differ in the mean number of genuine symptoms they endorsed, $p = .16$. Furthermore, both simulated pain ($M = 12.87$, $SD = 10.50$) and simulated anxiety ($M = 17.38$, $SD = 12.00$) participants endorsed more pseudosymptoms than the honest group ($M = 2.68$, $SD = 2.90$), $ps < .001$. *Cohen's d* was 1.33 for the contrast between honest participants and pain simulators and 1.68 for that between honest participants and anxiety simulators. The simulated anxiety group attained somewhat higher pseudosymptom scores than the simulated pain group, $p = .043$, *Cohen's d* = 0.40.

We also performed a *MANOVA* with Groups (honest comparison vs pain simulators vs anxiety simulators) and Language (English vs Dutch) as an independent, and the SRSI main scales (genuine and pseudosymptoms) as dependent variables. Language had an overall significant effect on at least one of the SRSI main scales, $\lambda = .95$, $F(2, 151) = 3.84$, $p = .024$, $\eta_p^2 = .05$. Follow-up univariate tests showed that there was an higher endorsement of pseudosymptoms among English speaking participants ($n = 107$; $M_{pseudo} = 12.48$, $SD = 12.35$) than among Dutch speaking students ($n = 51$; $M_{pseudo} = 8.19$, $SD = 7.51$), $F(1, 154) = 7.69$, $p = .006$, $\eta_p^2 = .05$. However, the number of endorsed genuine symptoms in English ($M_{genuine} = 26.45$, $SD = 13.17$) and Dutch people ($M_{genuine} = 23.95$, $SD = 13.08$) did not differ, $F(1, 154) = 2.90$, $p = .091$. Most importantly, the group differences described above (honest vs. pain simulators vs. anxiety simulators) did not change when Language was taken into account, and this was true for group differences in genuine symptoms $F(2, 152) = 79.95$, $p < .001$, $\eta_p^2 = .51$, and for group differences in pseudosymptoms, $F(2, 152) = 34.77$, $p < .001$, $\eta_p^2 = .31$.

Thus, there was no significant interaction between Group and Language for the genuine symptom scale, $F(2, 152) = .02, p = .981$, or the pseudosymptom scale, $F(2, 152) = .77, p = .466$.

SRSI subscales: pain simulators versus anxiety simulators

To investigate more specifically the pattern of symptom endorsement in the two simulator groups, we performed a *MANOVA* with scores on the four pertinent SRSI subscales (genuine pain symptoms, pain pseudosymptoms, genuine anxiety symptoms, and anxiety-related pseudosymptoms) as dependent variables. There was an overall significant difference between the two groups, $\lambda = .37$, $F(4, 102) = 43.73, p < .001, \eta_p^2 = .63$. Follow-up *t*-tests showed that the two groups differed significantly on all four subscales. That is, the instructed pain simulators endorsed significantly more pain-related genuine symptoms and pain-related pseudosymptoms than instructed anxiety simulators. On the other hand, instructed anxiety simulators endorsed significantly more symptoms on both genuine anxiety symptoms subscales and anxiety-related pseudosymptoms subscales than did instructed pain simulators (see Table 4.2). The Supplemental Table 4.1 gives scores of the three groups on all (i.e., ten) separate subscales (Appendix 3).

Table 4.2.
Scores on pertinent SRSI subscales of instructed simulators.

| SRSI scales | | Simulators | | <i>t</i> -test | <i>Cohen's d</i> |
|-----------------------|---------|-----------------------|--------------------------|----------------|------------------|
| | | Pain <i>M (SD)</i> | Anxiety <i>M (SD)</i> | | |
| <i>N</i> | | 54 | 53 | | |
| Genuine | Pain | 8.59 (2.30) | 3.71 (3.29) | 8.89* | 1.72 |
| | Anxiety | 5.39 (2.70) | 8.02 (2.54) | 5.17* | 1.00 |
| Pseudosymptoms | Pain | 4.23 (2.96) | 1.94 (3.01) | 3.95* | 0.77 |
| | Anxiety | 2.57 (2.65) | 6.19 (2.74) | 6.92* | 1.34 |

Note: * $p < .001$

Detection accuracy

Table 4.3 summarizes detection accuracy parameters (Area under the Curve (*AUC*), sensitivity, specificity). With the screening cutoff of > 6 endorsed pseudosymptoms, 78% of the instructed simulators were detected (pain simulators: 74%, anxiety simulators: 83%). With this cutoff, 10% of honest participants were misclassified as simulators (i.e., “false positives”).

With the standard cutoff score of 9, 60.7% of instructed simulators were detected. Looking at the simulator groups separately, more anxiety simulators (73%) than pain simulators (48%) were detected, $\chi^2(1) = 7.26, p = .007$. With this cutoff, there were two “false positive” cases (4%). Apart from the recommended cutoffs, Table 4.3 also includes accuracy information for the cutoffs of >4

and >15, which the SRSI manual (Merten et al., in press) describes as liberal and rigorous, respectively.

Table 4.3.

Area under the curve (AUC), and sensitivity and specificity for liberal, screening, diagnostic, and strict cutoffs (>4, .6, >9, and > 15, respectively).

| Groups | | SRSI Pseudosymptoms scale | | | | | | | | |
|-------------------------|------------|---------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| Honest comparison group | <i>AUC</i> | <i>95%CI</i> | >4 | | >6 | | >9 | | >15 | |
| vs. | | | <i>Sensitivity</i> | <i>Specificity</i> | <i>Sensitivity</i> | <i>Specificity</i> | <i>Sensitivity</i> | <i>Specificity</i> | <i>Sensitivity</i> | <i>Specificity</i> |
| Simulators combined | .90 | .85 .95 | .87 | .78 | .78 | .90 | .60 | .96 | .39 | 1.00 |
| Pain simulators | .88 | .82 .95 | .85 | .79 | .74 | .90 | .48 | .96 | .30 | 1.00 |
| Anxiety simulators | .91 | .85 .97 | .87 | .79 | .83 | .90 | .73 | .96 | .49 | 1.00 |

Note: All AUCs were significant, $p < .001$.

Discussion

We tested the sensitivity of the SRSI to instructed pain and anxiety simulators. Our findings can be summarized as follows. First, pain and anxiety simulators endorsed more genuine and pseudosymptoms than honest comparison participants. The two simulator groups did not differ in their overall endorsement of genuine symptom, and marginally differed in their reports of pseudosymptoms.

Second, and most importantly, simulators endorsed significantly more pseudosymptoms consistent with the scenarios they were instructed to mimic. The pain simulators reported significantly more genuine pain symptoms as well as pain-related pseudosymptoms than anxiety simulators. In contrast, the anxiety simulators endorsed significantly more anxiety-related symptoms than the pain simulators. These findings confirm the sensitivity of specific subscales of the SRSI to certain forms of over-reporting, in this case, pain and anxiety (Merten et al., 2016). Yet, pseudosymptom endorsement was not at all limited to the specific symptom domain stipulated in the scenario's, which is in accordance with exaggerators' tendency to go over the top in their symptom presentation (Conroy & Kwartner, 2006). Whether the other subscales of the SRSI exhibit a similar level of sensitivity to attempts to simulate a specific type of complaint remains to be established.

Third, the screening and the standard cutoff scores of 6 or 9 yielded considerable percentages of detected simulators (83.2% and 60.7%, respectively). The sensitivity of the standard cutoff is in line with previous instructed simulator studies that looked into the diagnostic accuracy of the SRSI. For example, when asked to simulate test anxiety, the standard cut score yielded a sensitivity of 77%

(Boskovic, Hope et al., 2018) and in a study about feigning a high impact negative experience, 89% of simulators were detected using this cutoff (Boskovic, Biermans et al., 2018). Admittedly, the sensitivity obtained in pain simulators may look disappointingly low. However, this reflects the problems in the diagnostics of pain experiences, which are so difficult to calibrate that it is often impossible to draw precise demarcation lines between exaggerated and authentic pain reports. It is no coincidence that in the domain of pain assessment there are only few SVTs available. From this perspective, the SRSI might be a valuable albeit imperfect asset. Clearly, it provides clinicians with only one source of data and therefore it should be embedded in a multimethod approach and a careful analysis of consistency and plausibility of pain reports (Bianchini, Greve & Glynn, 2005). Still, our results suggest that the SRSI is better at detecting anxiety simulators than pain simulators and this is an issue that warrants further exploration. It may well be the case that people have a better intuitive understanding of pain experiences and how they should be reported (Kucyi et al., 2015). If true, a lower cutoff point, such as 6, could be more efficient in the screening for simulated pain-related claims. However, due to the higher false positive rate associated with this cutoff, it should be used with much caution. Germane to this discussion is also our observation that anxiety simulators exhibited a stronger tendency to over-generalize their (pseudo-)symptom reports to other domains than pain simulators. In fact, compared with pain simulators, anxiety simulators endorsed significantly more cognitive and sensory pseudosymptoms, which contributed to their higher total pseudosymptom scores (see Supplemental table 4.1). The tendency to over-generalize symptom reports across several functional domains is why SVTs such as the SIMS and the SRSI seem to function (van Impelen et al., 2014; Merten, Friedel, & Stevens, 2007). A lack of over-generalization in some patients (e.g., sophisticated or coached respondents) puts upper limits to the sensitivity of SVTs to detect over-reporting.

Fourth, overall, we found a satisfying *AUC* of .90 for the SRSI, which corresponds well with *AUCs* that were reported by Merten et al. (2016). Yet, as far as diagnostic accuracy is concerned, we also observed two problems. To begin with, English speaking participants endorsed more pseudosymptoms than Dutch speaking participants. This might be the result of a poorer understanding of items, given that the English speaking group largely consisted of international students. A similar pattern was found for the English version of the SIMS (van Impelen et al., 2014). Another and more important point concerns the correlation of a standard psychopathology measure (i.e., the BSI-18) with SRSI pseudosymptoms, a finding that has also been observed for the SIMS (e.g., Edens et al., 2007). One relevant consideration in the case of the SRSI, however, is that the association of high BSI-18 scores with raised endorsement of pseudosymptoms was found to pertain to the lower end of the distribution of pseudosymptom scores. That is to say, there were only few honest participants with

raised BSI-18 scores who exceeded the cutoffs of the SRSI (but see below). This nuances the apparent sensitivity of the SRSI to real psychopathology. In accordance with this, recent studies have concluded that patients with serious psychopathology seldomly attain SRSI pseudosymptom scores above the cutoff (Giger & Merten, in press; Helvoort et al., under review).

A few restrictions of our study warrant comment. First, our study relied on a healthy student sample, which is not representative of the general public, let alone of patients. Interestingly, our honest group scored higher on genuine symptoms and pseudosymptoms than comparison groups that have been described in the SRSI literature and that were primarily composed of non-student participants (e.g., Merten et al., in press; $N = 155$; genuine symptoms: $M = 7.64$, $SD = 6.98$; pseudosymptoms: $M = 0.81$, $SD = 1.80$). In these non-student comparison groups fewer than 1% scored above the screening and the standard cut score. Thus, compliance with the honest response instructions might have been a problem in our honest comparison group, which restricts the generalizability of our findings. Future studies should compare simulators with patients in order to more closely examine the specific response patterns of these groups (Rogers, 2008) on the SRSI.

Second, and related to the previous points, we did not use pre-experimental checks in order to test participants' understanding and compliance with the role described in the simulator scenarios. By checking this, we might have improved participants' adherence to their scenarios, and would have been able to determine whether a number of participants did not comply with their roles (e.g., Nies & Sweet, 1994).

Third, our study was conducted using an online tool, which might have led to a more liberal endorsement of symptoms and pseudosymptoms. The anonymous context of online surveys may foster a cavalier way of responding (Fleischer, Mead, & Huang, 2015). Thus, we cannot rule out that some of the honest participants had a casual attitude responding and ended up with pseudosymptom levels above the cutoffs. It is because this consideration that we don't know for sure whether our "false positives" were truly false positives. However, in our previous lab-based studies that involved face-to-face contact with participants (Boskovic, Hope et al., 2018; Boskovic, Biermans et al., 2018), the rate of endorsed symptoms was similar to our current findings, for both genuine and simulators group. On the other hand, SRSI studies that were based on non-student participants who were given individualized test sessions (Giger & Merten, 2013, in press; Merten et al., in press) found rates of above cutoff scores that were lower and generally, fewer genuine symptoms and pseudosymptoms were endorsed in these studies. Thus, the relatively high rate of above cutoff pseudosymptom endorsement in the current study might flag correct positives in low motivated student respondents rather than false positives. Clearly, this issue warrants systematic research.

Finally, the employment of the experimental simulation design in the current study poses significant limitations with regard to external validity (e.g., Rogers & Cruise, 1998). Healthy experimental participants and the conditions under which they are tested differ in important aspects from patients undergoing real-world forensic or clinical assessments. Thus, even with the most careful consideration of optimizing a simulation design (such as described by Nies & Sweet, 1994, or more recently Rogers, 2018; Niesten et al., 2017), the results of experimental simulation studies will be of limited generalizability. In particular, estimates of sensitivity and specificity have to be judged with greatest caution; they rather reflect to upper limits of classification accuracy of an instrument than its real diagnostic potential. Relatedly, effect sizes from simulation designs tend to be substantially larger than those expected for real-world evaluations (Ingram & Ternes, 2016). Similar limits apply to the meaning of *AUC* estimates obtained from experimental simulation studies. However, in contrast to studies relying on patient samples, their advantage is a higher internal validity because they allow for a thorough experimental control of test conditions (Vickery et al., 2004).

In sum, we believe that our data add to the growing body of evidence indicating that the SRSI is a promising instrument for detecting over-reporting of various sorts of symptoms. The next step would be to thoroughly evaluate its diagnostic merits in patient samples. For instance, an interesting topic for future research might be to examine to what extent real-world pain patients with invalid symptom claims generalize their symptom reports to other pathological domains. More generally, this point touches upon the difficulties involved in the detection of exaggerated pain symptoms (see Fishbain et al., 1999; Bianchini et al., 2005; Greve, Bianchini & Brewer, 2013).

CHAPTER 5

THE VERIFIABILITY APPROACH AND THE SELF-REPORT SYMPTOM INVENTORY IN FABRICATED SYMPTOMS

This chapter draws on the following paper:

Boskovic, I., Diebets, P., Hope, L., & Jelicic, M. (in press). Verify the scene, report the symptoms: Testing the Verifiability Approach and the SRSI in detection of fabricated PTSD. *Legal and Criminological Psychology*.

Abstract

In order to effectively feign Posttraumatic Stress Disorder (PTSD), a person needs to confabulate an exposure narrative and to fabricate symptoms of high distress. The Verifiability Approach is a lie-detection method based on the notion that truth tellers' narratives include more verifiable (e.g., documented) information than liars' narratives. Thus, in this study, we examined the verifiability of the truthful and fabricated narratives about a reported aversive exposure. We also tested symptom reports among participants using the PTSD-related subscales of the Self-Report Symptom Inventory (SRSI). One subscale taps into genuine symptoms, while the second subscale includes pseudosymptoms likely to be endorsed by those who are feigning PTSD. One group (truth tellers) was presented with a vehicle crash scene using the Virtual Reality paradigm ($n = 22$), while the other group (feigners) was instructed to fabricate such experience ($n = 46$). Our results indicate that the feigners produced non-verifiable and lengthier narratives than truth tellers. In contrast, truth tellers reported a higher proportion of verifiable information. Symptom reports were significantly different between groups, with feigners endorsing more of both PTSD-related genuine symptoms and pseudosymptoms. The presence of non-verifiable details and the proportion of verifiable details, together with the symptom endorsement on the SRSI subscales, should be inspected in order to enhance the detection of fabricated PTSD-like accounts.

Key words: PTSD; Feigning; Verifiability; Narrative; Symptom Reports.

Verify the Scene, Report the Symptoms: Testing the Verifiability Approach and SRSI in the Detection of Fabricated PTSD

The term Posttraumatic Stress Disorder (PTSD) was officially introduced to the DSM in 1980 (DSM-3; American Psychiatric Association, 1980) and pertains to a broad range of psychological disturbances as a consequence of a trauma experience (Resnik, West, & Payne, 2008). Originally, this type of diagnosis was associated with war veterans (Adamou & Hale, 2003), but it soon became clear that combat exposure is not the only possible trigger for PTSD. In fact, any negative life event can be experienced as traumatic (Resnik et al., 2008), but not every traumatic experience results in PTSD (Bonanno, 2005; Hall, Hall, & Chapman, 2006). The prevalence of traumatic exposure in general public is up to 70%, yet, the prevalence of PTSD is approximately 10% (Young, 2016). The highest prevalence of PTSD is among the victims of sexual abuse (up to 80%; Hall et al., 2006) and war veterans (up to 58%; Gurriel & Fremouw, 2003), compared with a general population (up to 15%; Hall & Hall, 2007).

As currently described in DSM 5 (American Psychiatric Association, 2013), the diagnosis of PTSD includes eight different criteria (A-H). In order to receive the PTSD diagnosis, a person must have 1) a traumatic experience (A), 2) symptoms that were caused by the traumatic experience: re-experiencing (B), avoidance (C), negative mood and cognition (D), and arousal (E), which 3) last at least one month (F), present a severe obstacle for daily functioning (G), and are not a product of medication or alcohol/drug abuse (H). All of these criteria are based on self-report, meaning that a person can easily over-report or fully fabricate his/her trauma-related complaints if incentivized to do so (i.e., malingering). Malingering presents a deliberate fabrication of symptoms in order to gain potential external benefits (American Psychiatric Association, 2013). Indeed, there are many potential financial (e.g., compensation) and legal (e.g., reduced criminal responsibility) benefits available for people suffering from PTSD (Knoll & Resnick, 2006; see also Rassin, Boskovic, & Merckelbach, 2018). Thus, due to the self-evident nature of its key symptoms, incentives surrounding this diagnosis, and professionals who admit their uncertainty in recognizing malingering (Cohen & Appelbaum, 2016), PTSD is one of the most easily and the most frequently feigned psychological disorders (Gurriel & Fremouw, 2003; Resnik et al., 2008; Resnick et al., West, & Wooley, 2018). The estimated prevalence of fabricated PTSD is above 30% (Freeman, Powell, & Kimbrell, 2008; Lees-Haley, 1997).

In order to successfully feign suffering from PTSD, a person needs to provide a convincing story about the exposure (A criterion), and to adequately report symptoms (B-E). Therefore, the detection of such cases must comprise assessment of the validity of the exposure statements, as well as the validity of reported symptoms.

Verifying exposure narratives

Previous research on verbal credibility assessment suggests that, overall, truthful narratives contain more details than deceptive narratives (Johnson, 2006), as well as more specific information, such as time and location (Porter, Peace, & Emmett, 2007). Furthermore, truthful narratives tend to be more emotionally charged and judged as more plausible than narratives about a fictitious event (Peace & Porter, 2011). Currently, there are many methods of credibility assessment that are based on the assumption that truthful narratives include details of certain quality. For example, the most often used tool, the Criteria Based Content Analysis (CBCA, Steller & Koehnken, 1989), includes 19 different criteria. The presence of each CBCA criterion indicates the truthfulness of the statement (Amado, Arce, Farina & Vilarino, 2015). Two meta-analyses performed on the utility of the CBCA (Amado et al., 2015; Hauch, Sporer, Masip, & Blandon-Gitlin, 2017) provided overall favourable findings. However, findings reported by Bogaard and colleagues (2013) suggested that CBCA criteria are vulnerable to contextual bias (Bogaard, Meijer, Vrij, Broers, & Merckelbach, 2013). In other words, the interpretation of the statements and scoring on CBCA can vary considerably if coders are presented with additional information, such as personality characteristics. This vulnerability to contextual bias limits the reliability of the CBCA method (Bogaard et al., 2013).

Alternatively, a recently developed technique, the Verifiability Approach (VA; Nahari, Vrij, & Fisher, 2014), has shown some success in facilitating the detection of fabricated statements (e.g., Harvey, Vrij, Nahari, & Ludwig, 2017; Jupe, Leal, Vrij, & Nahari, 2017; Nahari, 2018). According to this approach, people who are telling truth include more details that are verifiable (i.e. checkable in the real world) in their statements than people who fabricate their accounts. Liars cannot provide verifiable information without the risk of being caught, therefore, they opt to report more information unavailable for external validation (i.e., non-verifiable details; Nahari, 2018). For instance, reporting “I had coffee with my friend, Brianna” would be considered as verifiable due to the mention of an identifiable person who could (in principle) confirm this claim. However, a person intending to deceive the interviewer would likely offer a vaguer report, such as “I was sitting on a bench alone” (see also Nahari et al., 2014).

Furthermore, the statements of truth tellers and liars differ even more when the statement-givers are informed, using an ‘Information Protocol’, that the number of reported verifiable details will serve as an indicator of their veracity. While this instruction elicits more verifiable details from truth tellers, liars are unable to follow through without exposing their lies, thus their reports remain mostly non-verifiable (Harvey et al., 2017). This implies that disclosing the detection strategy actually facilitates the success of the VA (Nahari, Vrij, & Fisher, 2014b).

In previous studies (Boskovic, Bogaard, Merckelbach, & Vrij, 2017; Boskovic, Gallardo, Vrij, Hope, & Merckelbach, 2018), researchers aimed to test the utility of the VA outside of its original lie-detection context. They investigated whether the VA is effective in the context of symptom reporting and if the use of this approach would enhance the detection of fabricated complaints (i.e., headache). However, checking the number of verifiable details did not aid the detection of fabricated symptoms complaints because people with genuine complaints and malingerers reported similar amounts of verifiable information. Rather, the amount of non-verifiable information reported appeared to be a cue for deceptive symptom reports. Those who malingered provided notably more non-verifiable details in their reports, which also resulted in their reports being significantly longer than genuine accounts. Surprisingly, both of these characteristics, (increased non-verifiable details and longer reports) were no longer diagnostic once the Information Protocol was applied. It is likely that providing participants with the clear instructions on which type of details would indicate their veracity subsequently influenced both groups of participants to write similar statements. This finding challenges the pattern of findings prevalent in the lie-detection literature.

It has been suggested that the utility of the VA approach is limited in the context of malingering due to the unverifiable status of internal symptoms (Nahari, 2018). However, we would argue that symptoms, such as pain or a severe headache, usually result in behaviours that can, potentially, be verified (e.g., going to the doctor, taking medication etc.; Boskovic et al., 2018). A critical difference between the lie-detection and malingering contexts might be the feasibility of providing false verifiable details. For instance, in typical lie-detection contexts, where people report about an external event likely witnessed by others (e.g., a fight), a production of misleading verifiable information might be a riskier choice than when a person is describing his/her internal state (e.g., pain). Therefore, feigners, who report about subjective experiences, might have an easier task in providing checkable information when presented with the Information Protocol (Boskovic et al., 2017).

Nevertheless, in case of malingered PTSD, the contexts of lie-detection and malingering overlap. The detection of fabricated exposure narratives resembles the lie-detection contexts, and therefore, similar methods may prove useful, however, the symptom veracity assessment requires a malingering specific approach.

Symptom endorsement in feigned PTSD

Assessing the veracity of the symptoms using the verbal assessment and content analysis (e.g., Akehurst et al., 2015) has not been a dominant approach within the field of symptom validity assessment. Rather, most research focusses on the development of symptom scales that test for over-reporting and exaggerated complaints, known as Symptom Validity Tests (SVTs). Exaggerated

responding to symptoms inventories has been shown to be a trademark of fabricated PTSD reports (e.g., Tracy & Rix, 2017; Hall & Hall, 2007; Peace & Masliuk, 2011). Thus, the idea behind SVTs is that those who fabricate their symptoms will over-endorse items on symptom inventories, even if they contain bizarre or implausible complaints (Merckelbach & Smith, 2003). One such measure is the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997). The SIMS includes 75 atypical items, which are not likely to be true even for genuine patients. For example, items such as: “Sometimes when writing a phone number, I notice that the numbers come out backwards even though I don't mean to do it”. People who endorse such items above a proposed threshold score (e.g., 16, Smith & Burger, 1997) are believed to be over-reporting, thus, their reports should be viewed with some scepticism. Furthermore, recently a study showed that tendency to overendorse SIMS items is even more pronounced among people fabricating direct traumatic exposure than indirect aversive experience (Szogi & Sullivan, 2018). However, these atypical items are often obvious to the examinees, a shortcoming that diminishes the reliability of SIMS. Furthermore, the SIMS includes items pertaining to complaints such as amnesia, psychosis, and low intelligence, which are frequent within the criminal context, but not in civil medico-legal setting (Merten, Merckelbach, Giger, & Stevens, 2016).

A recently developed measure of over-reporting, the Self-Report Symptom Inventory (SRSI; Merten et al., 2016) may therefore be a better alternative for a couple of reasons. First, the SRSI includes a mix of genuine symptoms and pseudosymptom items, divided into two main scales. A typical item for the genuine symptom scale is “I have nightmares about things that happened to me”, while for the pseudosymptom scale a typical example is “I can't remember what happened to me, but I constantly dream about it.”. Because both types of items are distributed within the questionnaire, it is more difficult for an examinee to realize the real aim of the assessment. Second, the SRSI genuine symptoms scale of symptoms related to 1) cognitive complaints; 2) depression; 3) pain; 4) somatic problems, and; 5) anxiety/posttraumatic stress disorder. The pseudosymptoms scale taps into 1) cognitive/memory complaints; 2) neurological motor and 3) neurological sensory issues; 4) pain, and 5) anxiety/depression/PTSD. Thus, each of the two main (genuine symptoms and pseudosymptoms) scales includes five subscales describing the most prevalent complaints within civil medico-legal context (Merten et al., 2016). Research to date suggests that the SRSI is a promising tool for detection of over-reporting, with rates of detection above 77% (Merten et al., 2016). However, the utility of the subscales alone, such as genuine and pseudosymptoms of anxiety-related issues, has not been examined so far.

The current research

The aim of this study was to investigate the quality of narratives and symptom reports between people instructed to fabricate an aversive experience and its consequences and people who were exposed to an aversive experimental manipulation. In this study, we included two groups of participants. One group participated in a (separate) study conducted in the clinical psychology department, in which they were exposed to a Virtual Reality (VR) scene of a vehicle crash as a method of inducing trauma-like symptoms. The second group, participating only in the current study, were not exposed to the VR scene of the vehicle crash, but was instructed to act as if they witnessed the same VR scene. Thus, we then asked all participants to write a narrative about the scene as if they witnessed it in person. We anticipated that truth tellers would report more verifiable details while feigners would produce more non-verifiable narratives. Additionally, we asked participants to report their distress caused by the witnessed scene by using the anxiety/PTSD-related subscales of the SRSI. We expected that truth tellers would endorse significantly fewer symptoms of both the genuine symptoms and pseudosymptoms than feigners.

Method

Sample

We initially recruited 102 participants; 21 participants were subsequently excluded (see below) and a further 13 participants withdrew from the study. The final sample consisted of 67 students (77.6% females, *Mean age* = 22.5, *SD* = 3.83). Of this sample, 22 (truth telling condition) participants were recruited from a separate clinical study focused on inducing PTSD-like symptoms using the VR paradigm. On average, participants joined the current study 6 days (*SD* = 8.00; range 0-25) after the VR exposure. After careful inspection, one participant was additionally excluded because of the extremely long delay after the exposure (80 days), which resulted in a total of 21 truth tellers in the final sample. The rest of the participants (*n* = 46) were a newly collected sample who did not witness the VR scene. Participants in this second group were instructed to feign the experience of witnessing the VR scene as if they experienced it 7-15 days ago (feigning condition).

Measures and Materials

Jellinek-PTSD Screening Questionnaire (JPSQ, van Dam, Ehring, Vedel, & Emmelkamp, 2013). Participants in the VR study (truth tellers) were pre-screened for putative PTSD symptoms using the JPSQ. The JPSQ is a short self-report questionnaire and consists of four questions that can be answered with either yes or no. The score is the total sum of positive answers (range 0 – 4). Only participants with a score of 0 were allowed to participate in the current study. The JPSQ has shown to have high sensitivity (.87) and specificity (.75) (van Dam et al., 2013).

PTSD Checklist for DSM-5 (PCL-5; Weathers et al., 2013). Participants in the feigning condition were pre-screened for any current high distress using the PCL-5. To adhere to ethical restrictions, we excluded any participant ($n = 21$) who responded with an answer equal to or higher than 3 (*Quite a bit*). The *Cronbach's alpha* of PCL-5, based on the entire initial feigning sample ($N = 80$), was .94.

Self-Report Symptom Inventory (SRSI; Merten et al., 2016). The SRSI includes 107 items belonging to two superordinate scales: One that pertains to plausible symptoms and one that pertains to pseudosymptoms. Furthermore, each of the two main scales includes five subscales tapping into different issues. In this study, we focused only on the two subscales, pertaining to the symptoms expected to follow an exposure to an aversive event (i.e., anxiety and PTSD-like symptoms). Thus, we included 22 items, 2 control items and 20 items describing the anxiety/PTSD related complaints. Half of the items present genuine anxiety/PTSD-like symptoms (*Cronbach's alpha* = .88) and the other half tap into exaggerated complaints (e.g., pseudosymptoms; *Cronbach's alpha* = .75). For each symptom, participants indicate whether or not they suffer from it (False/True), thus the maximum score per scale (genuine and pseudosymptoms) is 10. Although these subscales include a mix of symptoms, among which also depression, for the sake of clarity, we will address them as “anxiety/PTSD subscales”.

Newspaper Article. Before witnessing the VR scene, participants in the clinical study (truth tellers) were given a newspaper article. Modeled on newspaper reports of actual crashes, the article described the accident, included information about the victims, and speculated about possible causes of the crash. The article included a picture of the crash scene. The article was 100 words in length and was presented in both English and Dutch (see Appendix). We also incorporated the article into our study design.

Virtual Reality Scenario. In order to elicit PTSD-like symptoms among the truth tellers group, participants were exposed to a Virtual Reality scenario depicting a vehicle crash between a car and a train. The VR simulation was created and run in Unity 5 on an Oculus Rift DK2 (Development Kit 2). The programming language used was C# and the graphics were created in Blender 3D. The VR scenario was shot from the first-person perspective, and a participant could look left and right within the immersive scene. The scene presents as follows: The crash involving a train and a car has already happened. The participant is on a bike, and in front of her/him there is a fence and three parked cars. Two people, a woman and a man, are standing next to their car, panicking and trying to call for help. The train is not moving and burning fire starts in the car. The victims, a man and a baby, are loudly screaming from the car on fire. No train passengers are visible. The crossing lights and

warning sirens are on, and it is raining. The car is consumed in flames and the victims trapped in the car fall silent. In the last minute of the scene, police and ambulance sirens can be heard approaching.

Procedure

Participants from a separate clinical study served as our truthful comparison group (truth telling condition). This study was concerned with inducing PTSD-like symptoms. Exclusion criteria were a non-zero score on the JPSQ (see above) and having witnessed or having been involved in a car crash. The participants first received a newspaper article describing a crash between a car and a train. After reading the article, they were exposed to the crash scene in the VR lab. After approximately 7 days, they returned for a follow-up session when they were referred to the current study.

Participants in the feigning condition were pre-screened using the PCL-5 for current levels of distress. In total, we excluded 21 people who reported high levels of distress. Another exclusion criterion was having witnessed (or having been involved in) a car crash in real life. However, no participants reported having that experience. Thirteen participants in the feigning condition initiated participation in the study but did not complete it. Thus, those participants were not included in our final dataset.

All of the participants in both conditions received a Qualtrics link which allowed them to access the study program. After responding to demographic questions (age, gender, and student status), truth tellers were also asked to indicate the delay (in days) since exposure to the VR scene. Then, all participants were given a battery of self-report questionnaires¹⁹ followed by two main tasks: 1) to write a narrative about their experiences of witnessing the crash, and 2) to fill out the SRSI subscales. Prior to writing the narrative, participants in both conditions were instructed to read the newspaper report about the crash. Truth tellers were instructed to report about the VR scene in as much detail as possible. Participants in the feigning condition received instructions to feign witnessing the crash scene, and were asked to imagine receiving financial compensation from the train company if they provided a convincing witness account. Research showed that, besides compensation, revenge is a strong motive for feigning (Peace & Masliuk, 2011), thus, participants were encouraged to think about any negative experience they previously had with a national train company. Participants could opt to write their narratives in either English or in Dutch (see Figure 1). In the truth telling condition, 14 participants wrote the narratives in Dutch (66%), whereas 13 participants opted for Dutch in malingering group (20%). In order to motivate all the participants to write detailed narratives, we informed them that participants who provided detailed accounts would be entered into a prize draw for €20.

¹⁹ For the sake of clarity, hereby we will only focus on the narrative and the SRSI findings.

Coding using Verifiability Approach (VA; Nahari & Vrij, 2014). According to the VA, details of statements can be coded as verifiable or non-verifiable. In the current study, for a detail to be coded as verifiable, it had to meet one of the following criteria: 1) to be documented (recorded, or left actual or virtual trace); 2) was related to occurrences that were carried out together with (an) other identified person(s) rather than alone or with a stranger who could not easily be traced; or 3) pertained to something that was witnessed by an other identified person(s) (Nahari & Vrij, 2014). Details that did not fulfill these requirements were labeled as non-verifiable. Coding was performed by two coders, both blind to the conditions. The primary coder coded all the statements, while the second coded a randomly selected sample of 20% ($n = 18$) of the statements. The details were coded in four different categories: 1) Verifiable details from the newspapers (VNP; “There were two victims, a man and a baby”), 2) New verifiable details (V; “A man and a woman were calling the ambulance”), 3) Non-verifiable from the newspapers (NNP; “Potential cause of the crash was a distracted car driver”), and 4) New non-verifiable details (NV; “I thought I would faint”).

The inter-rater reliabilities between coders intraclass correlation coefficients (ICC) very good for VNP details (ICC = .88, 95%CI [.69 - .96]) and for V details (ICC = .86, 95%CI [.33 - .96]), as well as for the NNV details (ICC = .78, 95%CI [.40 - .92]) and excellent for NV details (ICC = .91, 95%CI [.52 - .97]).

Statistical Approach

The differences in the narratives’ quality and symptoms reports between truth tellers and feigners were calculated using the *Welch’s t-tests*, due to unequal sample size between groups. For the effect size of our findings we report *Cohen’s d*. In order to investigate the relationship between the features of trauma-related narratives and the quality of symptom reports, we calculated the *Pearson’s r product-moment correlation coefficients*. To further examine the contribution of the verifiable and non-verifiable details, SRSI genuine symptoms and pseudosymptom subscales to the prediction of the group membership (truth tellers vs. feigners) we performed the *Discriminative function analyses* and calculated the *Areas Under the Curve (AUCs)*.

Results

Motivation

With respect to their motivation to report convincing narratives, the average rating was 3.73 ($SD = 1.10$) on a five-point Likert scale (anchors being: 1 = “Not motivated at all”; 5 = “Extremely motivated”). Regarding their motivation to report the symptoms, participants’ average rating was 3.82 ($SD = .88$)²⁰.

²⁰ We re-conducted all the main analyses excluding the three participants who reported very low motivation, however, our results did not significantly differ. Therefore, we retained them in the data.

Narrative reports

Participants of the feigning group provided significantly longer narrative reports than truth tellers, *Welch's* $t(55.77) = 2.03, p = .047, d = 0.49$. Looking into the frequency of the four categories of details, we observed that 98.5% reported at least one verifiable detail from the newspapers, 100% reported new verifiable details, only 15% reported non-verifiable details from the newspapers, and 95.5% added new non-verifiable information into their reports. The feigners and truth tellers differed in the total number of non-verifiable details (NNP + NV), *Welch's* $t(51.05) = 2.64, p = .020, Cohen's d = 0.65$, which might be a consequence of the difference in the number of new non-verifiable details (NV), *Welch's* $t(51.38) = 2.68, p = .010, Cohen's d = 0.66$. No other significant differences were found in the number of other categories (VNP, V, and NNP) of details (see Table 5.1).

We calculated the proportions of total verifiable information ((V + VNP) / Total details), verifiable details from the newspapers (VNP / Total details), and the new verifiable details (V / Total Details), controlling for length of statements. The proportion of total verifiable details (V + VNP / Total details) was significant, *Welch's* $t(37.36) = 3.37, p = .002, Cohen's d = 0.90$, with truth tellers having higher proportions of these details ($M = .77, SD = .15$) than feigners ($M = .64, SD = .14$).

Looking separately in the two categories of verifiable information, no significant differences were found. The proportion of the verifiable details from the newspapers (VNP) was not significantly different between truth tellers ($M = .24, SD = .19$) and feigners ($M = .19, SD = .14$), *Welch's* $t(30.91) = 1.19, p = .24, Cohen's d = 0.30$. Similarly, neither the difference in the proportion of the new verifiable details (V) was significant, *Welch's* $t(41.12) = 1.95, p = .057, Cohen's d = 0.55$, although there was a higher proportion of new verifiable details among the truth tellers ($M = .52, SD = .14$) than among feigners ($M = .44, SD = .15$).

Symptom reports using the SRSI anxiety/PTSD subscales

Truth tellers reported significantly fewer genuine symptoms ($M = 2.81, SD = 2.97$) and pseudosymptoms ($M = .71, SD = 1.42$), than feigners (genuine symptoms, $M = 5.85, SD = 3.23$, and pseudosymptoms, $M = 2.67, SD = 2.18$). For genuine symptoms *Welch's* $t(41.91) = 3.77, p < .001, Cohen's d = 0.98$, , and *Welch's* $t(56.95) = 4.39, p < .001, Cohen's d = 1.06$, for pseudosymptoms (see Table 5.1).

Table 5.1.

Comparison of means and standard deviations between the truth tellers and feigners on all used measures.

| | Truth tellers <i>M</i> (<i>SD</i>) | Feigners <i>M</i> (<i>SD</i>) | Welch's <i>t</i> -test (<i>df</i>) ²¹ | Cohen's <i>d</i> |
|----------------------------------------------|-----------------------------------------|------------------------------------|-------------------------------------------------------|---------------------|
| Narratives | | | | |
| Length | 161.09 (111.86) | 231.63 (167.10) | $t(55.77) = 2.03^*$ | 0.49 |
| Total of Verifiable details | 33.81 (19.90) | 34.52 (22.80) | $t(44.10) = .130$ | 0.03 |
| Total of Non-verifiable details | 11.43 (11.66) | 20.50 (15.65) | $t(51.05) = 2.64^*$ | 0.65 |
| Verifiable details from newspapers (VNP) | 8.62 (3.61) | 8.20 (3.56) | $t(38.29) = .45$ | 0.11 |
| New Verifiable details (V) | 25.19 (18.30) | 26.33 (21.25) | $t(44.75) = .22$ | 0.05 |
| Proportion of total Verifiable details | .77 (.15) | .64 (.14) | $t(37.36) = 3.37^{**}$ | 0.90 |
| Non-verifiable details from newspapers (NNP) | .24 (.54) | .17 (.53) | $t(38.20) = .45$ | 0.13 |
| New Non-verifiable details (NV) | 11.19 (11.54) | 20.33 (15.60) | $t(51.38) = 2.68^{**}$ | 0.66 |
| Symptom reports | | | | |
| SRSI PTSD Genuine | 2.81 (2.97) | 5.85 (3.23) | $t(41.91) = 3.77^{**}$ | 0.98 |
| SRSI PTSD Pseudo | .71 (1.42) | 2.67 (2.18) | $t(56.95) = 4.39^{**}$ | 1.06 |

Notes: * $p < .05$, ** $p < .01$; The proportion of verifiable information was calculated using the next formula: $(VNP + V) / ((VNP + V) + (NNP + NV))$.

Narratives and Symptom Reports

In order to investigate the relationship between the features of trauma-related narratives (the number and proportion of verifiable details and the number of non-verifiable details) and the quality of symptom reports (endorsement of genuine symptoms and pseudosymptoms), we calculated the *Pearson's r product-moment correlation coefficients*. For truth tellers, we also included *delay after the exposure* (number of days before joining our study) as a variable. However, none of the correlations reached significance (all $ps > .05$), indicating no association between the features of narratives and the quality of symptom reports, and no association between the delay and the quality of the truth tellers' accounts. Among feigners, the non-verifiable details were significantly related to the endorsement of both genuine symptoms and pseudosymptoms, *Pearson's r* being .32 and .31, respectively ($ps < .05$). None of the other correlations were significant (see Table 5.2).

²¹ All calculations were also performed using a non-parametric, *Mann-Whitney U test* and the results did not differ.

Table 5.2.

Pearson's correlation coefficients for the features of fabricated narratives and the symptom reports among feigners.

| | SRSI Genuine symptoms | SRSI Pseudosymptoms |
|--------------------------------------------|-----------------------|---------------------|
| Truth tellers ($n = 21$) | | |
| Verifiable details Total | .30 | -.04 |
| Verifiable details proportion | -.19 | -.41 |
| Non-Verifiable details Total | .42 | .19 |
| Feigners ($n = 46$) | | |
| Verifiable details Total | .19 | .15 |
| Verifiable details proportion | -.14 | -.11 |
| Non-Verifiable details Total | .32* | .31* |

*Note: * $p < .02$.*

Diagnostic Utility of the (Non-)Verifiable Details and the SRSI subscales

First, we examined general detection accuracy for each of the main measures (number and proportions of total verifiable and non-verifiable details, as well as the genuine and pseudosymptoms), using the Area Under Curve (AUCs) (see Table 5.3). All measures, except the number of verifiable details ($AUC = .50$, $p = .962$), detected feigners better than chance ($AUCs > .70$).

Table 5.3.

The Area Under the Curve (AUC), significance (p) level and Confidence Intervals (CI) of numbers and proportions of verifiable, number of non-verifiable details, and SRSI anxiety/PTSD genuine and pseudosymptoms.

| Measures | AUC | p | 95%CI |
|---------------------------------------|-----|-------|---------|
| Verifiable details Total | .50 | .962 | .36-.65 |
| Verifiable details Proportion | .72 | .004 | .59-85 |
| Non-Verifiable details Total | .70 | .008 | .57-.84 |
| Non-Verifiable details Proportion | .72 | .003 | .59-85 |
| SRSI genuine anxiety/PTSD symptoms | .74 | .001 | .62-87 |
| SRSI anxiety/PTSD pseudosymptoms | .78 | <.001 | .66-89 |

Second, we tested the incremental validity of combining these measures: i) the VA (proportion of verifiable detail & total number of non-verifiable details), ii) SRSI subscales (genuine anxiety/PTSD symptoms & pseudosymptoms), and iii) VA and SRSI subscales combined. We ran

three *Discriminative Analyses* to distinguish between the truth tellers and feigners. The group membership was in all three cases the classifying variable. The analysis including the VA (the proportion of verifiable details and total non-verifiable details) yielded significant discriminative function, $\chi^2(2) = 10.95$, $\lambda = .84$, $p = .004$ (*canonical correlation* = .40). The SRSI subscales also significantly predicted the group classification of participants, $\chi^2(2) = 14.98$, $\lambda = .79$, $p = .001$ (*canonical correlation* = .45). Finally, the discriminant function was also significant for the VA and the SRSI subscales combined, $\chi^2(4) = 20.30$, $\lambda = .72$, $p = .001$ (*canonical correlation* = .52). The sensitivity, specificity and the overall accuracy are provided in the Table 5.4.

Table 5.4.

Sensitivity (true positive), specificity (true negative), and overall accuracy in classifying participants to truth tellers and feigners using narrative features and the subscales of the SRSI.

| | Predictors | Sensitivity | Specificity | Overall accuracy |
|-------------------------|-------------------------------|-------------|-------------|------------------|
| Narrative | Verifiable details proportion | 47.6% | 93.5% | 79.1% |
| | Non-verifiable details | | | |
| SRSI subscales | SRSI genuine symptoms | 62% | 82.6% | 76.1% |
| | SRSI pseudosymptoms | | | |
| Narrative + SRSI | | 52% | 87% | 76.1% |

Discussion

In the current research, we examined whether a combination of different detection strategies worked to expose the reporting strategies of individuals feigning PTSD. Specifically, we tested the Verifiability Approach (VA; Nahari, Vrij, & Fisher, 2014) in the context of PTSD exposure narratives, and the utility of the anxiety/PTSD-related subscales of the Self-Report Symptom Inventory (SRSI; Merten et al., 2016) within symptom validity assessment.

Our results indicate that participants who were asked to feign an aversive exposure produced significantly longer narratives than truth tellers reporting on an actually experienced aversive event. This finding is consistent with earlier research showing that feigners elaborate on their injuries more than truth tellers (e.g., Purisch & Sbordone, 1997). Examination of the narrative content revealed that feigners inflated the length of their reports by overproducing new, non-verifiable details. For instance, they focused on describing their subjective state during the scene (e.g., “I was shocked”) rather than on external circumstances, as shown in previous symptom-focused studies (e.g., Boskovic et al., 2017, 2018). It would appear that, regardless of the origin of experience in question, feigners overcompensate for the lack of truthful verifiable information with non-verifiable details. Yet, every participant reported at least one detail that was, in principle, checkable. However, truth teller narratives included a greater proportion of verifiable information. These results align well with

findings regarding verifiability within the lie-detection context where the proportion of verifiable information has been shown to differentiate between truthful and fabricated accounts (e.g., Harvey et al., 2017). Overall, our findings suggest that inspecting the narratives of real-life trauma exposure for (non)verifiability, should be considered for future investigation of this topic.

Looking at the symptom report results, the anticipated response pattern emerged. Feigners endorsed significantly more symptoms from both genuine- anxiety/PTSD and pseudo- anxiety/PTSD subscales than truth tellers. This pattern of results fits well with previous research using the complete SRSI questionnaire (Merten et al., 2016), and it confirms the over-generalization strategy feigners exhibit when reporting about their symptoms (Merten, Thies, Schneider, & Stevens, 2007). Inflated symptoms reports were also previously reported in research investigating feigned victimization claims (e.g., Peace, Porter, & Cook, 2010). Furthermore, we found that feigners' symptom over-endorsement was associated with a higher frequency of non-verifiable details in their narratives. This pattern of results suggests that both the VA and the SRSI captured the *hyperbolicism* that is typically found in fabricated accounts.

Taken together, the current results suggest that the proportion of verifiable details, the total amount of non-verifiable details, and both SRSI subscales were valid predictors of feigning. However, this was not the case for the *number* of verifiable details, which might not be as an important feature of narratives in the symptom validity assessment (Boskovic et al., 2017). Yet, when these predictors were included in discriminative function, the VA (both proportion of verifiable details and number of non-verifiable details) resulted in the overall accuracy of 79%. The main advantage of the VA was the high true negative rate (specificity), meaning that truth tellers were correctly identified. In contrast, the SRSI subscales exhibited similar overall accuracy of 76%, with improved sensitivity at the cost of specificity. Combined, both measures produced an overall detection accuracy of 76%, with a balanced trade-off in terms of sensitivity and specificity, supporting a multi-method approach to the detection of feigned PTSD-like accounts.

A few methodological issues warrant comment. We included a healthy student sample, some of whom had been induced with trauma-like symptoms using the VR exposure. The VR paradigm is currently the most sophisticated method for induction of PTSD symptoms (Dibbets & Schulte-Ostermann, 2015), however, the VR exposure cannot fully replicate real-life exposure experiences, thus, our results have a limited generalizability to actual PTSD patients. Next, our sample was limited in size, which may have led to underpowered results, although the main findings regarding narratives and symptom reports correspond well to the results of previous studies in the field (e.g., Boskovic et al., 2017). In addition, because truth tellers were free to join the current study at any time after the VR exposure, they may have chosen to proceed with our study once the effect of the exposure had

declined. However, in this study, the correlation analyses indicated that the delay was not associated with the quality of truth tellers' accounts. Furthermore, all the participants in the truth telling condition were debriefed about the previous study before joining our study. There was no way of testing whether/how this influenced their reports, but it is likely that students anticipated our interest in the trauma-related symptomatology. It is important to mention that participants had an opportunity to choose in which language (Dutch or English) to write the narrative, but that we did not control for students' language proficiency. This factor could have had an impact on our findings, considering that language was found to be an important confounder in symptom validity studies (Nijdam-Jones & Rosenfeld, 2017). Another limitation was that we did not apply the full version of the SRSI, as recommended by Merten et al. (2016), thus, a closer examination of detection accuracy was limited. Therefore, the findings concerning the SRSI subscales should be taken with caution and further tested using the whole SRSI. Finally, participants in the feigning condition were asked to imagine having a chance to revenge to the train company because revenge was shown to be a strong motivator (Peace & Masliuk, 2011). However, in reality, people claiming PTSD are often confronted with a significantly stronger (financial or emotional) incentive (positive or negative), which can have a different influence on their response style (Peace & Masliuk, 2011; Peace & Richards, 2014; Resnick, West, & Payne, 2008).

In conclusion, non-verifiable and lengthier narratives remain a strong cue to fabricated exposure accounts, whereas an overendorsement of both genuine and pseudosymptoms should raise skepticism with respect to the truthfulness of the symptom reports. In order to achieve the most balanced detection of fabricated PTSD claims the examination of narratives' quality and symptom reports should be combined. However, in order to validate our results, future research including the PTSD patients or/and accounts of the real-life aversive events is critical.

CHAPTER 6

THE CROSS-CULTURAL INFLUENCE ON PRACTITIONERS' PLAUSIBILITY JUDGMENTS OF ATYPICAL SYMPTOMS

This chapter draws on the following paper:

Boskovic, I., van der Heide, D., Hope, L., Merckelbach, H., & Jelicic, M. (2017). Plausibility judgments of atypical symptoms across cultures: An explorative study among Western and non-Western experts. *Psychological Injury and Law*, 10, 274–281.

Abstract

Symptom validity tests (SVTs) are predicated on the assumption that overendorsement of atypical symptoms flags symptom exaggeration (i.e., questionable symptom validity). However, few studies have explored how practitioners from different cultural backgrounds evaluate such symptoms. We asked professionals working in Western ($n = 56$) and non-Western countries ($n = 37$) to rate the plausibility of uncommon symptoms taken from the Structured Inventory of Malingered Symptomatology (*SIMS*), dissociative symptoms from the Dissociative Experience Scale (*DES-T*), and standard symptoms (e.g., anxiety, depression) from the Brief Symptom Inventory-18 (*BSI-18*). Western and non-Western experts gave similar plausibility ratings to atypical, dissociative, and standard symptoms: both groups judged BSI-18 symptoms as significantly more plausible than either dissociative or atypical symptoms, while the latter two categories did not differ. Our results suggest that the strategy to detect symptom exaggeration by exploring overendorsement of atypical items might work in a non-western context as well, although dissociative symptoms should be additionally evaluated.

Key words: Symptom Validity Assessment; Atypical Symptoms; Cross-Cultural Research, Structured Inventory of Malingered Symptomatology.

Plausibility Judgments of Atypical Symptoms Across Cultures: An Explorative Study Among Western and Non-Western Experts

When a person is presenting with atypical mental problems (e.g., “My headaches are so severe that my feet hurt”), this may raise the suspicion of malingering. Malingering is defined as the intentional production of false or grossly exaggerated symptoms, motivated by external incentives. Such incentives may involve financial rewards, compensation, or reduced legal responsibility (American Psychiatric Association, 2000). However, what is considered to be an atypical symptom may depend on the cultural background of patients and evaluators (e.g., Weiss & Rosenfeld, 2012). Thus, cultural backgrounds may affect how patients express psychological or medical complaints and how healthcare professionals evaluate the plausibility of these complaints (e.g., Thakker & Ward, 1998; Kleinman & Cohen, 1997; Hausotter & Schouler-Ocak, 2007). Surveys suggest that professionals from different countries only moderately agree in their evaluation of mental disorders (Giosan, Glovsky, & Haslam, 2001) and neuropsychological symptoms (e.g., mild head injury; Ferrari, Constantoyannis, & Papadakis, 2001). Exaggerated symptoms might be more acceptable or even expected in one culture, but possibly an instant red flag for malingering in others (Charles, Gafni, Whelan, O’Brien, 2006). Furthermore, the language in which a medical or psychological examination is conducted may affect the response style of patients (Harzing, 2006; Johnson, Kulesa, Cho, & Shavitt, 2005), leading to possibly inaccurate conclusions about significantly different prevalence levels of exaggerated symptomatology across countries (Nijdam-Jones & Rosenfeld, 2017). However, there are only a few cross-cultural studies on symptom validity assessment (e.g., Merten & Rogers, 2017), and even less research has focused on practitioners’ judgments of atypical symptoms across cultures.

Symptom Validity Tests (SVTs) aim to detect an exaggerated response style in patients (e.g., Larrabee, 2012). Many SVTs are predicated on the assumption that endorsing a relatively high number of atypical symptoms is indicative of symptom exaggeration. One widely used instrument (e.g., Martin, Schroeder, & Odland, 2015) is the Structured Inventory of Malingered Symptomatology (SIMS; Widows & Smith, 2005), which lists 75 bizarre, uncommon, atypical, and rare symptoms such as “There is a constant ringing in my ear” and “The voices that I hear, have never stopped since they began”. Endorsing more than 16 of these atypical symptoms indicates a heightened probability of exaggerated symptom presentation (Merckelbach & Smith, 2003). Although the internal consistency of the SIMS is satisfactory, its test-retest stability is sufficient, and its ability to discriminate between symptom exaggeration and honest responding is fairly effective (with sensitivities varying between 75% and 100%; van Impelen et al., 2014), some authors have expressed concerns about using SVTs such as the SIMS in patients or defendants with a non-Western

background (Merten & Rogers, 2017; Nijdam-Jones & Rosenfeld, 2017). Specifically, Merten and Rogers (2017) note that the detection of exaggerated symptoms in minority groups might be complicated by culturally distinct illness expression and clinicians' stereotypes about malingering in migrant workers.

Whether bizarre or atypical symptoms of the SIMS are also bizarre and unlikely in a non-Western context is an empirical question. Some scholars have speculated that Eastern cultures focus more on the somatic manifestations of psychiatric conditions such as posttraumatic stress disorder (PTSD), while in Western countries patients emphasize the psychological impairments that accompany these conditions (Kleinman & Cohen, 1997; see also Duckers, Alisic, & Brewin, 2016; Terheggen, Stroebe, & Kleber, 2001). This would suggest that different thresholds across cultures might apply in detecting atypical symptomatology. On the other hand, Van der Heide and Merckelbach (2016) compared SVT outcomes of several groups of asylum seekers who stayed in a psychiatric facility. Their study involved the following groups: (1) asylum seekers who had incentives to exaggerate their mental problems; (2) asylum seekers who did not have such incentives; (3) asylum seekers with a poor proficiency in the language of the host country (Dutch), and (4) asylum seekers with a good proficiency in Dutch. The authors compared these groups with regard to their endorsement of atypical symptoms taken from the SIMS. Atypical symptom endorsement occurred on a nontrivial scale and was related to incentives rather than language proficiency. In line with this, Nijdam-Jones and Rosenfeld (2017) concluded in their recent meta-analysis on cross-cultural feigning assessment involving 34 different tools that of the four psychiatric symptom validity measures (i.e., M-FAST, MENT, PAI, and SIMS), that the SIMS had the highest overall classification accuracy, indicating the lowest level of variability across cultures and languages.

Unfortunately, no research has examined the cultural background of professionals who make decisions about the plausibility of various symptoms. With this in mind, we wanted to explore possible cultural variations in perception of atypical symptoms among professionals from Western and non-Western countries. Besides atypical symptoms, we also included common psychological problems such as depression and anxiety. Furthermore, we included dissociative symptoms because they might overlap with atypical symptoms (Merckelbach et al., 2015). We anticipated that experts with a Western background would find the common psychological problems more plausible than atypical symptoms taken from the SIMS, with dissociative symptoms occupying an intermediate position. We had no a priori hypothesis about the symptom plausibility rank order of experts with a non-Western background.

Method

Sample

Our study included a convenience sample of 93 professionals from 22 countries. The average working experience of the professionals was 9.55 ($SD = 8.43$) years, 11 years ($SD = 9.45$) for Western and 7 years ($SD = 5.50$) for non-Western professionals ($t(91) = 1.91, p < .05$). The majority of them (72%) were working in the field of clinical psychology and psychotherapy, while 23% had medicine as their work setting. Western and non-Western professionals were mostly working in a clinical (43% and 32.5%, respectively), forensic (35.7% and 13.5%, respectively), or therapy (14.3% and 43.2%; respectively) setting. Groups did only differ with regard to the latter setting; Mann-Whitney U test = 774.00, $z = -2.16, p = .03$.

Following Huntington (1993)²², we assigned professionals from North America, Western Europe, Australia, and New Zealand to the Western group. Professionals from East and South Europe, Asia, and Africa formed the non-Western group. In total, the Western group consisted of 56 professionals (60%), while the non-Western group consisted of 37 professionals (40%) (see Table 6.1).

Measures

We included the 37 items from the short form of the **Structured Inventory of Malingered Symptomatology** (SIMS; Smith & Burger, 1997; Malcore, Schutte, Van Dyke, & Axelrod, 2015²³), 8 items from the **taxon subscale of the Dissociative Experiences Scale** (DES; Bernstein & Putnam, 1986; Waller, Putnam, & Carlson, 1996), and 18 items from the **Brief Symptom Inventory-18** (BSI-18; Derogatis, 2001).

The SIMS items allude to psychological and neuropsychiatric symptoms that are, at least in a Western context, uncommon (Merckelbach & Smith, 2003). An illustrative item is: “Sometimes my muscles go limp for no apparent reason so that my arms and legs feel as though they weigh a ton”. The symptoms of the DES-taxon include the most pathological forms of dissociation and involve unusual phenomena such as “I have experienced being in a place and having no idea how I got there”. The items of the BSI-18 refer to the typical symptoms of depression and anxiety that are, at least in a Western context, relatively prevalent. Illustrative items are “Nervousness or shakiness inside ”and “Difficulty in breathing”. All symptoms ($37+8+ 18 = 63$) were reformulated into statements of patients (e.g., “I have difficulty breathing”), mixed, and then presented to the professionals.

²² Huntington proposed a differentiation between Western and non-Western civilizations primarily based on religion and economic and social development.

²³ Following the appendix in the manuscript by Malcore et al. (2015), 38 items were excluded from the original SIMS version. This resulted in the final version of 37 items, which is not in accordance with authors’ claim that the final version includes 36 items. Our questions to the authors were left unanswered.

Plausibility Judgments on Fabricated Symptoms

Table 6.1.

The frequencies of professionals from Western and non-Western countries.

| Country | Western | Non-Western |
|------------------------|-----------|-------------|
| Australia | 2 | 0 |
| Bosnia and Herzegovina | 0 | 1 |
| Canada | 4 | 0 |
| China | 0 | 3 |
| Croatia | 2 | 0 |
| Germany | 5 | 0 |
| Greece | 0 | 4 |
| Indonesia | 0 | 1 |
| Ireland | 1 | 0 |
| Italy | 7 | 0 |
| Japan | 0 | 1 |
| Lebanon | 0 | 2 |
| Lithuania | 6 | 0 |
| Malaysia | 0 | 2 |
| Netherlands | 8 | 0 |
| New Zealand | 1 | 0 |
| Rwanda | 0 | 1 |
| Serbia | 0 | 19 |
| South Africa | 0 | 2 |
| United Kingdom | 8 | 0 |
| USA | 12 | 0 |
| Vietnam | 0 | 1 |
| Total | 56 | 37 |

Procedure

The study was conducted using Qualtrics. Participants were contacted via email, at medical and psychological conferences, or personally invited via email to join the study. Professionals first completed a set of demographic questions (e.g., work experience and field of work) and then asked to imagine a patient who is presenting with a specific symptom. The task of professionals was to grade each symptom on a 5-point plausibility scale (anchors: 1 = “Definitely authentic”; 5 = “Definitely exaggerated”).

After professionals had rated the 63 symptoms, they were asked questions about prevalence issues (“How often do you think patients exaggerate symptoms?”) using 4-point scale (anchors: 1 = “Never”; 2 = “Almost never”, 3 = “Rarely”, 4 = “Often”), whether there are any clear signs for detection (anchors: “Yes” and “No”), and to provide a description of clues they considered to be important for the detection of exaggerated symptomatology.

Mean plausibility scores. We calculated the mean plausibility scores for SIMS, DES-taxon, and BSI-18 symptoms, separately (sum score/number of items). Thus, mean plausibility scores varied between 1 (definitely authentic) and 5 (definitely exaggerated). The data and the analysis can be found on Open Science Framework platform, following the link: <https://osf.io/f8pqk/>.

Results

Group Differences in Symptom Plausibility

We conducted a multivariate analysis of variance (*MANOVA*), where dependent variables were the plausibility scores on SIMS, DES-T, and BSI-18, and the independent variable was culture background of professionals (Western vs non-Western). The main effect of cultural background was non-significant, $\lambda = .98$, $F(3, 89) = 1.30$, $p = .28$, which indicates that this factor did not affect how practitioners judged the plausibility of symptoms²⁴. Mean scores of Western and non-Western professionals for the three categories of symptoms are presented in Table 6.2. The table shows that the results were not significant with respect to any of the three measures used.

Rank Ordering Symptom Plausibility

We next ranked the average plausibility judgments of all symptoms for the Western and non-Western group, and calculated a correlation between groups’ rank orders. The Spearman rank correlation, $r_s = .91$, $p < .01$, indicated high agreement between professionals’ judgment of items’ plausibility.

Using a series of *t*-tests (with *alpha* values adjusted to .02), we explored whether professionals from Western and non-Western background differed in their plausibility judgments for individual symptoms. The groups evaluated three symptoms significantly different, all from the SIMS: Item 11 (“Recently I’ve noticed that my memory is getting so bad that there have been entire days that I cannot recall”), $t(91) = 2.62$, $p = .01$, $d = .54$; Item 12 (“At times I’ve been unable to remember the names or faces of close relatives so that they seem like complete strangers”), $t(91) = 2.44$, $p = .017$, $d = .51$, and item 19 (“Sometimes my muscles go limp for no apparent reason so that my arms and legs feel as though they weigh a ton.”), $t(91) = 3.32$, $p < .01$, $d = .79$. Western professionals evaluated these symptoms as less plausible ($M = 3.17$, $SD = 1.17$, $M = 2.87$, $SD = 1.25$, and $M = 3.02$, $SD = 1.15$,

²⁴ We also ran this MANOVA with working experience as a covariate, and none of the results reached significance: $\lambda = .97$, $F(3, 89) = .98$, $p = .40$, SIMS $F(1, 90) = 2.71$, $p = .11$, DES-T $F(1, 90) = 2.01$, $p = .16$; BSI-18 $F(1, 90) = .30$, $p = .60$.

respectively) than the non-Western group ($M = 2.47$, $SD = 1.40$, $M = 2.22$, $SD = 1.29$, and $M = 2.20$, $SD = 1.17$, respectively). In Supplemental table 6.1. mean plausibility ratings and corresponding rank numbers can be found.

Table 6.2.

Mean plausibility scores of Western and non-Western professionals.

| Measures | Groups | N | M (SD) | t (91)* | p |
|---------------|--------------|----|------------|---------|-----|
| SIMS | Western | 56 | 2.61 (.57) | 1.85 | .07 |
| | Non-Western | 37 | 2.34 (.84) | | |
| | Total | 93 | 2.50 (.70) | | |
| DES-T | Western | 56 | 2.54 (.71) | 1.62 | .10 |
| | Non-Western | 37 | 2.26 (.89) | | |
| | Total | 93 | 2.43 (.80) | | |
| BSI-18 | Western | 56 | 1.80 (.84) | .52 | .60 |
| | Non-Western | 37 | 1.73 (.71) | | |
| | Total | 93 | 1.77 (.64) | | |

Notes: SIMS - *Structured Inventory of Malingered Symptomatology* (SIMS; Smith & Burger, 1997), DES-T - *taxon items of Dissociative Experiences Scale* (DES; Bernstein & Putnam, 1986); BSI-18 - the *Brief Symptom Inventory-18* (BSI-18; Derogatis, 2001); *We also performed *Welch's t*-tests, *p* levels remained non-significant ($ps > .091$).

Differences between Symptom Categories

We compared the average plausibility judgments of SIMS, DES-T, and BSI-18 symptoms with each other, using paired *t*-tests, in order to investigate whether practitioners would differentiate between atypical symptoms (SIMS), dissociative symptoms (DES-T), and common symptoms (BSI-18). SIMS ($M = 2.50$; $SD = 0.70$) and DES-T symptoms ($M = 2.43$; $SD = 0.80$) were evaluated as less plausible than BSI-18 symptoms ($M = 1.77$; $SD = 0.64$), $t(91) = 11.63$, $p < .01$ and $t(91) = 8.85$, $p < .01$. The difference in plausibility ratings for SIMS and DES-T symptoms was not significant: $t(91) = 1.72$, $p = .09$ (see Table 6.3).

Prevalence and Clues of Symptom Exaggeration

The majority of professionals (64.5%) agreed that exaggeration is rare, and a quarter (24.8%) indicated that it occurs often, while other categories were less frequently chosen (Almost never: 7.5% and Never: 3.2%). Overall, prevalence estimates were related to plausibility ratings for SIMS ($rs(93) = .30$, $p < .01$), DES-T ($rs(93) = .28$, $p < .01$), and BSI-18 symptoms ($rs(93) = .31$, $p < .01$).

More than half of professionals believed that there are no clear signs of exaggerated symptomatology (57%), while 36.6% responded positive to this question, and the rest did not provide an answer (6.4%). Only 27% of the total sample, 37.5% of the Western group and 11% of the Non-Western group, gave brief descriptions. In total, they generated 39 clues that were grouped into seven

different categories: Inconsistency within a report or an incongruence between reported symptoms and behavioural or anamnestic information (31%); Over-reporting of implausible symptoms (18%); Little or too specific (medical terminology) details of symptom reports (18%); Presence of external benefits (13%); Specific non-verbal clues (10%); Individual factors such as educational background (7.5%), and presence of personality disorders (histrionic or antisocial) (2.5%).

Table 6.3.

Contrasts in plausibility judgment of SIMS, DES-T, and BSI items for full sample.

| Contrasts | Means (SD) | <i>t</i> (91) | <i>p</i> | <i>Cohen's d</i> |
|----------------------|--------------------------|---------------|----------|------------------|
| SIMS - DES-T | 2.50 (.70) 2.43 (.80) | 1.72 | .09 | 0.16 |
| DES-T - BSI-18 | 2.43 (.80) 1.77 (.64) | 8.85 | .001 | .93 |
| BSI-18 - SIMS | 1.77 (.64) 2.50 (.70) | 11.63 | .001 | 1.20 |

Notes: SIMS - *Structured Inventory of Malingered Symptomatology* (SIMS; Smith & Burger, 1997), DES-T - *taxon items of Dissociative Experiences Scale* (DES; Bernstein & Putnam, 1986); BSI-18 - the *Brief Symptom Inventory-18* (BSI-18; Derogatis, 2001).

Discussion

Many SVTs are based on the rationale that overendorsement of atypical symptoms is reflective of symptom exaggeration. However, do atypical symptoms possess cross-cultural constancy? This question bears relevance to, for instance, the evaluation of asylum seekers with psychiatric problems, in which culturally shaped presentations of symptoms might be misjudged as feigning. This led some workers in the field to take a sceptical position as to the utility of SVTs across different cultural settings. For example, Merten and Rogers, (2017; p. 106) wrote: “Assuming that any feigning measure is universally applicable across languages and diverse cultures is categorically unacceptable.” However, virtually no studies investigated whether professionals from various countries evaluate the plausibility of atypical symptoms in a similar way. With this in mind, we asked professionals working in Western and non-Western countries to judge a mix of symptoms that are – in a Western context – common or atypical.

Our findings can be summarized as follows. First, there were no significant overall differences between professionals from Western and non-Western countries in how they evaluated the plausibility of atypical symptoms (SIMS), dissociative experiences, (DES-T), and common mental problems such as depression (BSI-18). Both Western and non-Western professionals found BSI-18 symptoms more plausible than SIMS symptoms. This finding provides support for the review of Nijdam-Jones and Rosenfeld (2017) in which they concluded that the SIMS can be used to

differentiate between exaggerating and non-exaggerating response styles in various language settings. Apparently, the atypical and bizarre nature of SIMS symptoms is constant across different cultural settings, which makes them useful for detecting an exaggerated symptom presentation.

Second, both groups regarded dissociative symptoms as less plausible than the common symptoms of the BSI-18. This might have to do with the fact that dissociative symptoms have a lower prevalence in the general population than symptoms such as depression and anxiety (Wittchen et al., 2011). However, professionals did not find dissociative symptoms more plausible than SIMS items. This observation is in line with recent research suggesting that in both healthy groups and clinical samples dissociative symptoms co-occur with symptom exaggeration (Merckelbach et al., 2015; Merckelbach, Boskovic, Pesy, Dalsklev, & Lynn, 2017). It might well be the case that individuals who engage in symptom exaggeration have a preference for dissociative symptoms because they regard these symptoms as indicating a profound impairment. For example, commenting on how malingering is portrayed in novels, Kuperman (2006; p. 70) concluded that “When madness is feigned, the eccentricity of simulation (...) sends a message to observers: ‘I’m not myself, so I’m not responsible’.” Thus, lay people may have the idea that dissociative symptoms (e.g., amnesia, depersonalization) compromise personal responsibility and in some settings (e.g., in court), this is precisely the impression that people would be motivated to convey.

Finally, the majority of professionals believed that exaggeration of symptoms occurs rarely. Inconsistencies between reported symptoms and behavioural or anamnestic information were seen as the most important clues for the detection of exaggeration. Both findings are in line with previous studies (Ruff et al., 2016; Keesler et al., 2017)

A few limitations of the current study warrant comment. First, our study was based on a relatively small, convenience sample, symptoms were only provided in English, and we did not ask professionals to judge their English proficiency. It is possible that our survey reached only people who are already engaged in some form of research-related activities (e.g., attending conferences), and who communicate primarily in the English language. Thus, our results might differ if the survey had been available to a wider population of practitioners and presented in their own language. Second, it might be the case that some of the professionals originally came from another country than the one they are currently working in. We did not obtain information as to their country of origin from all participants, but we assume that the cultural setting in which they presently work is more decisive for their evaluation of symptoms than the country in which they were born. Third, groups significantly differed in work experience and in the field of practice. However, even when we included work experience as a covariate, no differences were found. Fourth, and related to the previous point, the high agreement between Western and non-Western professionals in their plausibility ratings might

reflect a common Western oriented training in psychology and/or medicine. Fifth and most importantly, our study focused on the plausibility of single symptoms, when in clinical practice, professionals will look at the combination of symptoms. Given these limitations, future studies may want to survey larger groups of Western and non-Western professionals in their own language, including those who had a non-Western training, and ask them to evaluate the plausibility of symptom combinations.

In sum, Western and non-Western professionals were found to show a high level of agreement in their evaluation of symptoms. Importantly, SIMS symptoms are seen by Western and non-Western professionals as bizarre, lending some credit to the use of the SIMS in non-Western groups (e.g., asylum seekers; Van der Heide & Merckelbach, 2017). Both groups of professionals also rated dissociative symptoms as less plausible than common BSI-18 symptoms. This might reflect a representative heuristic (Tversky & Kahneman, 1974), with BSI-18 symptoms being evaluated as more plausible than dissociative symptoms simply because the former are more prevalent than the latter. Alternatively, it might reflect the inherent problematic nature of dissociative symptoms due to the fact that malingerers have a preference for eccentric psychopathology such as dissociation (Merckelbach et al., 2017). Given this ambiguity, it might be wise to develop measures that tap into dissociative symptomatology, but that also include validity scales that correct for over-reporting.

Our results in no way imply that cultural differences in symptom presentation should be disregarded. The rich literature on such differences showcases the importance of cultural considerations for clinicians' diagnostic routines (see e.g., Young, 2014; Nijdam-Jones et al., 2017; Weiss & Rosenfeld, 2012).

CHAPTER 7

GENERAL DISCUSSION

"I'd committed GBH (grievous bodily harm), he said. After they arrested me I sat in my cell and I thought, 'I'm looking at five to seven years'. So I asked the other prisoners what to do. They said, 'Easy! Tell them that you're mad! They'll put you in a country hospital. You'll have Sky TV and PlayStation. Nurses will bring you pizzas... Tony said faking madness was the easy part, especially when you're seventeen and you take drugs and watch a lot of scary movies..."

(Jon Ronson, 2011, "The Psychopath Test", p. 42-43)

Malingering occurs on a nontrivial scale in both the criminal (e.g., Mittenberg et al., 2002) and civil legal contexts (e.g., Larrabee, 2003), and it has a serious social and legal impact on society (Jelicic et al., 2018). The detection of malingering presents a challenge that needs to be tackled using multiple approaches (Scott & McDermott, 2011). In the course of research conducted for this thesis, we empirically examined the efficacy of three detection methods that assess different aspects of an examinee's behaviour in relation to his/her complaint.

First, we examined whether a novel verbal credibility assessment tool, the Verifiability Approach (VA; Nahari et al., 2014) could be applied to the detection of fabricated symptom reports. Routine symptom assessment usually begins with a clinical interview, in which patients' self-report plays an important role. However, the unstructured interview is an unreliable method for detecting fabricated symptoms (Jelicic et al., 2018). Moreover, the most frequently malingered symptoms (e.g., pain) are those that cannot be (dis)proven using other, more objective measures (see Mittenberg et al., 2002). Therefore, we wanted to test whether evaluating the content of symptom reports in terms of verifiable and non-verifiable details could assist the detection of symptom fabrication, and provide more information about malingerers' reporting strategies.

Second, in an attempt to find more objective parameters to assess examinees' health complaints, some researchers argued that using a reaction time measure, such as the Modified Stroop Task (MST) may produce more reliable results than self-report (e.g., Buckley et al., 2003; Constants et al., 2014). The MST effect occurs when a person exhibits prolonged response latency in color-naming the disorder-related words, compared with their reaction time when neutral words are presented. The presence of the MST effect is believed to signal genuine pathology (e.g., Buckley et al., 2003). However, research has provided inconsistent findings regarding the utility of MST in symptom validity assessment (e.g., Buckley et al., 2002). Furthermore, Kimble and his colleagues (2009) indicated the presence of a publishing bias for papers showing results supporting the existence of the MST effect more generally. We administered the MST in order to test whether the absence of the MST effect (i.e., patients' prolonged reaction time in color-naming the disorder-related words) signals malingering.

Third, following the dominant detection method in symptom validity assessment, the Self-Report Symptom Validity Tests (SRVTs), we evaluated a recently created measure, the Self-Report Symptom Inventory (SRSI; Merten et al., 2016). The SRSI is a self-report questionnaire used to screen for symptom over-reporting. It consists of two main scales tapping into genuine complaints (genuine symptoms scale) and into exaggerated, implausible symptoms (pseudosymptoms), respectively. The rationale behind the SRSI is that people with real problems will endorse more items from the genuine symptoms scale and a low number of pseudosymptoms. In contrast, people who exaggerate their complaints will overendorse items belonging to both the genuine symptoms and pseudosymptoms scale. Based on previous studies that reported initial evidence for the SRSI as a promising detection tool (e.g., Merten et al., 2016), we continued investigating the adequacy of the SRSI in the detection of fabricated test anxiety, high distress symptoms, and physical complaints.

Finally, we also examined the influence of culture on the symptom validity assessment. Previous research suggested that manifestation and interpretation of symptoms differs across countries (e.g., Kleinman & Cohen, 1997), and that language differences can lead to misunderstanding of items often included in symptom validity tests (Nijdam-Jones & Rosenfeld, 2017; Merten & Rogers, 2017). Investigating the cross-cultural differences in symptom validity assessment is of interest to, for example, clinicians who have to evaluate the psychiatric symptoms of asylum seekers (e.g., van der Heide, Boskovic, & Merckelbach, 2017). Hence, we investigated whether the practitioners from Western and non-Western countries differ in their judgments of the plausibility of common, dissociative, and atypical symptoms.

Below, the main findings of each chapter will be briefly summarized and integrated in terms of the over-arching purpose of this research: the quality enhancement of symptom validity assessment. We also review the limitations of our findings and elaborate on future directions and the practical implications of our research to date.

Main Findings of This Thesis

The research described in this thesis yielded a number of important and novel findings. First, our main findings with respect to the Verifiability Approach suggest that malingerers of physical symptoms, when not instructed on the details they should include in their reports, produced considerably more non-verifiable information, and thus longer statements than truth tellers. In contrast, truth tellers, even though they reported verifiable information infrequently, included a higher proportion of such details in their reports than malingerers (Study 1, Chapter 2). This latter finding replicated the results observed in the lie-detection literature (e.g., Nahari et al., 2014), which showed that the proportion of verifiable information contained in a statement is an important cue to veracity. Previous studies (e.g., Harvey et al., 2017) found that informing statement-givers that the verifiability

of details would serve as a cue for honesty (i.e., the Information Protocol) facilitates the differences in number of verifiable details reported by honest people and liars. To contrary, using this instruction in our study resulted in the disappearance of the previously established differences between the malingerers and the truth tellers (Study 2, Chapter 2). It is possible that administering the Information Protocol weakened the effect of the Verifiability Approach and provided enough information because it helped malingerers in generating false verifiable information. Malingerers' dominant strategy in creating false checkable details was to incorporate false witnesses (e.g., family members) in their stories.

The tendency of those who feign a condition to create or report false proofs has been observed before. Thomas (2001) examined the clinical features of patients believed to be malingering dissociative symptoms and compared them with genuine patients. Behaviours such as bringing “proofs” to support the diagnosis (e.g., telling others about their condition, and enrolment in the self-help groups) were most frequently seen among malingerers (Thomas, 2001). Moreover, in our experiment, in which we induced physical symptoms in one group and asked others to fabricate such complaints, the results once again revealed malingerers' preference for reporting non-verifiable details and lengthy statements (Study 3, Chapter 2). In general, this pattern of findings was not affected by the strategy of one group of malingerers to use embedded lies. Therefore, our results suggest that it is the number of non-verifiable details, rather than verifiable information, that should present a red flag for malingering. Specifically, a high number of vague details in the symptom reports should raise doubts about the validity of the complaints. These findings contradict the assumption of the DSM and other misconceptions surrounding the alleged uncooperativeness of malingerers, and their reluctance to talk (e.g., Ali et al., 2015). People who fabricate their symptoms seeking some kind of external benefit have a high motivation to convince the evaluator of the veracity of their claims and, as a result, tend to give detailed injury reports (Purisch & Sbordone, 1997). In contrast, the “transparency illusion” (Savitsky & Gilovich, 2003) that is often present among truth tellers, may lead people telling the truth to believe that their honesty is obvious to others, thus, they might provide poorer reports than malingerers.

Second, we failed to replicate findings suggesting that the absence of the MST effect can distinguish genuine from fabricated complaints (e.g., Buckley et al., 2003). Malingerers, without having any pre-knowledge about the response pattern they should provide in order to appear genuine, were nevertheless able to produce the test anxiety MST effect. In contrast, using the standard cutoff point of nine endorsed pseudosymptoms on the SRSI, 77% of malingerers were detected (Study 4, Chapter 3). The study in which we investigated the detection of fabricated high impact aversive experiences yielded similar findings (Study 5, Chapter 3). The MST did not occur in any of the

groups, neither among the people with an actual high impact personal history, nor among those malingering such reports. Still, malingerers were significantly slower in color-naming all the words, compared with participants with the high and low impact experiences. This finding runs against the assumption that a prolonged reaction time might indicate genuine complaints (Buckley et al., 2002). Rather, the increase in reaction time could just as easily be a consequence of malingerers' cognitive load while performing the task (Vrij et al., 2008). An increased cognitive load might also explain previous findings of Constants et al. (2014), who found the longest response latencies for over-reporting groups. Applying the SRSI to the three impact groups, however, led to the detection of 89% of malingerers, which reflects higher diagnostic accuracy than the absence of the MST effect in symptom validity assessment (34%). Yet, when using the standard SRSI cutoff point, 27% of the high impact group was misclassified as malingerers. The high rate of false positive results indicate two possible issues: First, that the SRSI's standard cutoff point might not be as accurate as reported by Merten et al. (2016) and, thus, requires further examination. Alternatively, it may be that participants in the high impact group had engaged in over-reporting during the pre-screening. Perhaps, then, they were from the beginning assigned to the wrong group. Despite this, both studies in this chapter indicated that while the MST is an unreliable tool for detecting fabricated symptoms, the SRSI has promising diagnostic value in the symptom validity assessment.

Third, we tested the psychodiagnostic accuracy of the SRSI in fabricated physical (i.e., pain) and psychological (i.e., anxiety) complaints (Chapter 4). In both cases, malingerers endorsed more genuine symptoms and pseudosymptoms than the honest comparison group. The SRSI performed better in facilitating the detection of malingerers who reported anxiety-related complaints (77%) than in the detection of malingerers who reported pain-related symptoms (56%). The lower detection rate among simulators of pain-related symptoms might have occurred because the face validity of pain-related items in the SRSI genuine symptoms scale may match participants' intuitive knowledge about the more or less uniformed manifestation of pain (Kucyi et al., 2015). Thus, people malingering pain may be readily able to recognize which symptoms do not seem plausible, and they will avoid endorsing those symptoms. In contrast, the interpretation of anxiety might be more vulnerable to individual differences and, as a result, the pseudosymptoms might be less obvious to malingerers meaning that they are more likely to be endorsed. Additionally, anxiety is often comorbid with other psychological issues, such as depression (Cummings, Coporino, & Kendall, 2014), somatic or concentration problems. Each of the two main scales of the SRSI includes five subscales from which at least three consist of symptoms that could also be found in anxiety patients (e.g., somatic problems, cognitive impairments, and depression). Thus, simulators of anxiety-related symptoms, compared with simulators of pain, had a higher chance of over-endorsing symptoms, and being detected by the

SRSI. Only two participants from the honest control group were misclassified as malingerers, which is a low false positive rate of 4%. Thus, overall, the SRSI was again shown to be a promising tool for detecting over-reporting.

Fourth, in the examination of PTSD-like complaints, in which a person needs to convincingly report about trauma exposure (criterion A), and about the symptoms (criteria B-E), we combined the VA and the anxiety/PTSD-related subscales of SRSI (Chapter 5). Advocates of the VA suggest that this approach might not be applicable for questioning the veracity of internal experiences, such as emotions or symptoms (Nahari et al., 2012, 2018). Thus, we tested the VA in exposure narratives, which are oriented towards describing an external aversive event. Our results were consistent with the first study in symptom-related settings. Those who lied about being exposed to an aversive event produced more non-verifiable information and lengthier statements. Conversely, truth tellers' narratives included a higher ratio of verifiable details than malingerers. Regarding the SRSI subscales, malingerers endorsed more anxiety/PTSD-related genuine symptoms and pseudosymptoms than truth tellers. Furthermore, malingerers' reporting strategies in writing the narratives by incorporating the extensive amount of non-verifiable information, were shown to be related with their overendorsement of all types of symptoms on the SRSI. In terms of detection accuracy, combining the VA and SRSI led to the most balanced detection of PTSD-like symptoms. When both methods were combined, the strengths of each method evened out, resulting in the overall detection accuracy of 76% in our sample. These findings highlight the benefit of employing a multi-method approach in the assessment of fabricated accounts.

Overall, what we observed as malingerers' dominant strategy in any given task across the whole thesis might be described as *hyperbolism*. Their narratives were excessive, the response latencies remarkably long, and the symptom self-reports amplified. The results of all three applied methods suggest that, regardless of the modality in which we tested malingerers' performance, exaggeration could serve as a trademark of intentional symptom fabrication (see also Jones, 2017; Porter, Yuille, & Lehman, 1999; Tan, Slick, Strauss, & Hultsch, 2002).

Finally, we investigated whether the plausibility of atypical symptoms varies across cultures. Frequently used measures in symptom validity assessment, such as the SIMS (Nijdam-Jones and Rosenfeld, 2017), rely on atypical symptoms as a detection tool. Endorsement of atypical symptoms might indicate malingering in one culture, but is this assumption also true for other cultures? In other words, how cross-cultural are atypical symptoms? What is atypical in one culture need not to be atypical in another. Thus, it is important to investigate whether practitioners from Western and non-Western countries differ in their plausibility judgments of such symptoms. We noticed that practitioners have a high agreement in their plausibility judgments of symptoms, ranging from

everyday complaints, dissociative symptoms, to bizarre items. Thus, we found no major differences in plausibility judgments of symptoms between practitioners working in Western and non-Western cultures. However, a concerning result was that, regardless of their cultural background, practitioners were not able to distinguish the plausibility of the bizarre items from dissociative symptoms. The practitioners in our sample possibly failed to make a distinction between the two because malingerers prefer to report dissociative symptoms, especially in court setting (Merckelbach et al., 2017). However, practitioners' lower plausibility scores of dissociative symptoms might be caused by the low prevalence of dissociative complaints in practice (i.e., representative heuristic, meaning that something that is less representative in general public is seen as less plausible; Tversky & Kahneman, 1974; see also Staniloiu & Markowitsch, 2014).

In light of our findings, we encourage future research exploring a multi-method approach including the VA and the SRSI in the context of symptom validity assessment. As shown above, both the VA and the SRSI helped detecting malingerers' dominant deceptive strategies, such as providing lengthier statements and vaguer details, and exaggerating their symptom complaints. In contrast, our investigation of an already established, although controversial task, the MST, did not provide any supportive results for this method as a detection tool for malingering. Possible reasons for our unfavourable MST findings might be the study designs we employed, or the samples we used. However, the ongoing controversy in regard to the MST's reliability (Kimble et al., 2009), suggests that the issues relating to the MST originate from the lack of a rigorous scientific investigation of the underlying mechanisms believed to be causing the MST effect. The need for deeper investigation of the MST might be additionally hindered by a publication bias that was shown to exist towards papers reporting positive MST results. Therefore, we strongly advise practitioners to avoid basing their decisions solely on data resulting from applying the MST.

Limitations

A number of limitations of the research conducted for this thesis warrants comment. First, we relied on student, healthy and/or subclinical samples. Thus, the generalizability of our findings to the wider public and clinical groups may be problematic. Due to ethical considerations, during the pre-screening of participants, we were restricted to include only students whose intensity of symptoms did not exceed certain thresholds. Therefore, it is possible that our results would differ among patients, or even in students with more serious complaints. This might be particularly the case for our results regarding the physical symptoms and PTSD-related complaints. For instance, people suffering from intense symptoms, such as chronic insomnia, would struggle with writing reports or performing tasks (Moul et al., 2002). Similarly, student reports about ending their relationships and moving away from their families may not be an entirely appropriate model of statements provided by

severely victimized or abused people. However, in real-life symptom validity assessment, even severely injured or traumatised individuals are usually motivated to provide strong reports about their symptoms, considering their opportunity to obtain certain incentives.

Second, and related to the previous issue, participants were pre-screened using self-report questionnaires. This approach might have led to the inaccurate classification of participants into groups. For example, when we used the SRSI to screen for over-reporting of symptoms of aversive experiences, we had a high rate of false positive results (Chapter 4). This means that the participants in the high impact group were misclassified as simulators who over-reported the impact of their personal history. However, without ground truth, there is no way of determining whether these participants with a highly stressful history were misclassified as over-reporters, or whether they were actually over-reporting during the pre-screening, leading them to be misclassified as honest participants in the pre-study phase. The limitation of relying on self-reports applies to the majority of studies in psychology, and it reflects the core issues in symptom validity assessment in practice. Self-reports are often misleading, thus, drawing conclusions on patients' self-reports can lead to poor decisions. For example, a claim for financial compensation might be declined based on the person's symptom report, although, they are genuinely suffering from, for instance, PTSD. Therefore, in the evaluation of symptom reports, practitioners are well advised to gather information from other sources (e.g., family members, work), and should apply a multi-method approach for symptom assessment. Regarding symptom validity research, scientists should become more creative and use other methods for ascertaining a correct classification of participants. For example, we controlled the participants' group membership by inducing physical symptoms in one group of participants using a demanding exercise (Study 3, Chapter 2), or, in another study, we exposed participants to an aversive scene via a Virtual Reality paradigm (Chapter 5).

Third, we conducted the majority of studies using the online tool Qualtrics in order to distribute our questions to the participants. Research shows that participants responding to the online questionnaires sometimes respond in an inattentive way (e.g., Fleischer et al., 2015). There are a few different ways of dealing with the issue of inattentive responding. For example, including items that serve as a check for consistency in reporting, as included in the SRSI, can also screen for inattentive responders (Merten et al., 2016). Inattentive responding might also be related to the lack of incentives offered for participating in studies. Previous studies showed that once participants are paid or offered compensation for joining the study, the frequency of reluctant, passive responders is low (Roivainen, Veijola, & Miettunen, 2016). It is consequently a reasonable concern whether participants would provide the same responses to an online study as they would if they were tested in person. To be certain about whether our results were influenced by using the online tool, we also examined each

method in the lab settings. The studies on the VA or SRSI, in a lab and online, produced similar results with regard to number of details, length of statements, levels of symptom endorsement, and the detection rates, suggesting an independency from the actual method of data collection (i.e., online versus face to face). Nevertheless, it is still possible that the conditions in which a study is conducted can significantly influence its outcomes. Thus, these conditions should always be clearly specified.

The last, and possibly the most important issue concerns the restrictions associated with the simulation design applied in our research. Investigating malingering in an artificial environment eliminates two important features of this phenomenon: the deceit and the intent of a deceiver (Merten, 2017; Niesten, et al., 2017; Nies & Sweet, 1994). All the participants who were asked to mangle were instructed to feign certain conditions and were offered an incentive to do so. Thus, instead of showing an actual deceptive behaviour, our participants were offered an incentive to comply with our instructions (i.e., malingering-simulation paradox, Rogers & Cruise, 1998). In other words, participants were asked to be obedient, rather than to be deceptive. Furthermore, we cannot certainly know whether the participants were willing to comply with their roles or that they were interested in the incentive (Merten et al., 2017; Nies & Sweet, 1994). We asked participants to grade their motivation to follow and comply with the instructions in all of the studies. However, participants might have provided socially desirable responses. The simulation design is adequate for validating measures used in symptom validity assessment. However, it does not allow a closer examination of underpinnings of malingering (Niesten et al., 2017; Merten et al., 2017). Thus, Niesten and her colleagues (2017) have proposed a new methodology for investigating malingering that may be associated with greater ecological validity. These researchers first pre-tested their participants for any somatic symptoms, and included participants who did not have any current symptoms. Then, the participants received an instruction to perform a long and dull task. During the task, the experimenters informed their participants that they could end the session early if they experienced any somatic symptoms. The task itself was not designed to provoke any somatic symptoms, thus, if a participant reported such complaints, the experimenters knew that the person was feigning. As such, the group assignment to feigners and truth tellers was not a priori set by researchers, rather the participants themselves chose whether to fabricate symptoms or to proceed with the task (remaining truthful). This shift from complying with the instructions of experimenters to choosing whether or not to feign has, partially, bridged the gap between real-life situations and malingering research. This new design might be an avenue worth following in future research in symptom validity assessment. However, the simulation design is widely used in other deception detection fields of research, such as lie-detection, and it presents a valuable method for investigating deceptive behaviours. The majority of studies within the context of lie-detection rely on the mock-crime paradigms, in which

participants are asked to commit a crime, and to later deceive others about performing that action. In this scenario, it might be possible to adjust the methodology, and induce the intent to carry out a mock-crime among participants, instead of asking them to do so. Nevertheless, the simulation design is the only appropriate method in research investigating more severe offences. For example, one cannot expect that researchers should try to induce the intent to execute a terroristic attack in their participants (e.g., Vrij, Granhag, Mann, & Leal, 2011), rather than just asking them to imagine they are a terrorist.

Future Avenues in Symptom Validity Assessment Research

In order to further explore and facilitate the VA in the malingering context, future studies should consider incorporating some sort of warning or misdirection such that participants will think there could be consequences for reporting false verifiable information. This way, the study design would better simulate the real life high-risk situations (e.g., insurance claims), in which people would be more reluctant to fabricate checkable information. For example, previous research showed that displaying a warning that suggests that some type of a lie-detection scale will be administered later during testing, considerably decreased faking good (positive response style) among candidates in personnel selection assessment (Schrader & Osburn, 1977).

Additionally, investigating symptoms which are less frequent than headaches or back pain (e.g., hearing voices), might result in different findings with respect to the number of non-verifiable details and lengthy statements. Because people are less likely to have personal experience with hearing voices, for example, their descriptions might be brief and less detailed than reports of people who genuinely suffer from such complaints. The influence of personality traits on the verifiability of statements also requires further exploration. Earlier research showed that higher levels of fantasy proneness, alexithymia, and dissociation significantly influence the clarity and credibility of a person's report (e.g., Peace & Masliuk, 2011). It is worth investigating whether the verifiability of statements would differ between individuals with high, medium, or low levels of fantasy proneness or with different levels of other personality traits (e.g., alexithymia; Boskovic, Reymarks, & Merckelbach, 2018). Such investigation might be able to reveal whether the production of verifiable details, for instance, can be linked to a certain personality traits.

Another potential avenue is to test the VA with sexual abuse reports. CBCA has been of assistance to practitioners in assessing the credibility of sexual reports for years, and is still widely used (see Ernberg, Magnusson, Landström, & Tidefors, 2018). However, the VA might be a better alternative for this task. As previously discussed, CBCA includes 19 criteria for coding, one of which is the quantity of details, which serves as an indicator for truthful reports. Our findings in the malingering context, however, challenge this assumption, showing that deceptive participants

provided more details about being exposed to an aversive event than honest participants. In the light of these results, one might anticipate that people who falsely report being victims of sexual abuse provide lengthier and more detailed statements than do real victims. In such cases, the utility of CBCA might be undermined. Furthermore, analysing the content of reports of sexual assault in terms of their verifiability might be a more reliable method than using the CBCA. Research showed that the inter-rater agreement for total CBCA scores is good, but that it significantly oscillates between the 19 criteria (Vrij, 2008). On the other hand, it was shown that the use of CBCA total scores should be avoided (Hauch et al., 2017). One might assume that coders agreeing on numerous criteria (e.g., unusual details) is a more challenging task than agreeing on whether or not details could, in principle, be checked (verified) or not. Furthermore, other studies indicate that CBCA coders are vulnerable to contextual bias, showing that coders' assessment of credibility is heavily influenced by, for instance, their view on a person's morality or other personality traits (Bogaard et al., 2013). However, the contextual information, such as information about evaluatees' personality characteristics, should not affect the raters' perceptions of whether or not a detail is verifiable. An additional argument that supports the use of the VA over CBCA in practice is that the coding of (non)verifiable details is more efficient and less time consuming than coding 19 different criteria for details. Future research should undertake investigation of coding reliability using both CBCA and the VA both between coders and between different labs.

Regarding future investigation of the SRSI, inclusion of various clinical samples would provide more information concerning the appropriate cutoff points, especially in patients with physical symptoms (Boskovic, Peters, & Merckelbach, 2018). Our research showed that the standard cutoff point of nine endorsed pseudosymptoms might lead to a low detection rate of people simulating physical complaints such as pain. Furthermore, testing the SRSI in patients who are instructed to underreport or to exaggerate their complaints might provide more precise information about the sensitivity of the SRSI to distorted response styles and genuine symptomatology. However, this is not without ethical concerns. Instructing patients to exaggerate their health problems could increase the severity of their symptom reports over time (i.e., the residual effect of feigning; Merckelbach, Jelicic, & Pieters, 2010). The residual effect of feigning has been found among (healthy) undergraduate students who, after playing a role of a feigner, continued endorsing more symptoms, although they were no longer instructed to fabricate health complaints (Merckelbach et al., 2010). In the case of patients, the residual effect of feigning might lead to significantly more intense symptoms.

More generally, future studies should examine the stability of fabricated symptoms. This issue is important because of the widespread belief that fluctuations and inconsistencies within symptom reports over time might be a cue to malingering (e.g., Blaney, 2014). However, previous studies

investigating trauma-related symptoms showed that those instructed to feign such complaints were more uniform in their reports (Peace, Porter, & Cook, 2010). It might well be that malingerers are aware of this consistency-veracity misbelief, thus, intentionally producing a consistent story over time. Similar findings were shown in the lie-detection context, indicating that liars try to ensure consistency when reporting their alibis (Vrij, Mann, Leal, & Granhag, 2010). In contrast, the state of people truly suffering from symptoms may oscillate over a period of time causing naturally-occurring inconsistency (Taylor, Frueh, & Asmundson, 2007). Our preliminary findings concerning the stability of common, most often reported symptoms (see Petrie et al., 2014) shows that, during a five-day observation period, those who malingered reported more consistent and exaggerated symptom intensity scores than truth tellers (Boskovic, Zwaan, & Merckelbach, 2018). The myth that inconsistency in the way a person reports their symptoms over time signals deception is still held by practitioners in symptom validity assessment (e.g., Kessler et al., 2017) and by professionals evaluating the credibility of asylum seekers reports (Herlihy, Scragg, & Turner, 2002). Thus, it is important that researchers share their findings with practitioners, whose decision making could significantly improve by dismissing the consistency-focused evaluation.

Practical Implications

Across this thesis, the utility of three methods for detecting fabricated symptoms is examined. Our result indicated that two methods, the Verifiability Approach (VA), and the Self-Report Symptom Inventory (SRSI) might be of assistance to practitioners conducting symptom validity assessment. In contrast, our studies on the utility of the Modified Stroop Task (MST) did not provide any support for its use in practice.

Given that the clinical interview is a mandatory part of symptom validity assessment, using methods of verbal credibility assessment could moderately help decision making about examinees' statements. However, employment of the VA would also require certain adjustments to be made in the interview procedure, such as audiotaping and transcribing the interview, or asking a patient to write down a report. This might seem unusual to practitioners who are used to relying on their own notes taken from the clinical assessment. Yet, one needs to be aware of the heuristics and biases that certainly are reflected in those notes and the interpretation of them, thus, the actual patients' reports may be a more reliable source of information. The main advantages of applying the VA are that it does not require any additional equipment, and that practitioners do not need to perform any sort of follow-up to the details that seem checkable (Nahari, 2018). Therefore, implementing the VA as an additional check to the self-reports would not disturb the current protocol of symptom validity assessment. Simply counting the details that fall into the two categories, verifiable and non-verifiable, and looking into the ratio between them may provide an indication of a patient's veracity.

Nevertheless, one has to keep in mind the shortcomings of this approach. Mainly, the low production of verifiable details among people with genuine complaints, and the VA's lack of sensitivity to malingerers' strategy in providing false verifiable information. Therefore, finding the most suitable way of implementing the VA within symptom validity assessment needs further work.

Recently, the use of the SVTs (i.e., Symptom Validity Tests) as an official part of the symptom assessment procedure, received strong support from associations of practitioners (see Bush, Heilbronner, & Ruff, 2014; Chafetz et al., 2015). We believe that our findings regarding the SRSI favor the implementation of this new measure in symptom validity assessment, although its low sensitivity to genuine symptomatology and fabricated physical complaints requires further examination.

Finally, practitioners' agreement on the plausibility of symptoms, whether bizarre or common symptoms, supports the use of certain Western measures (i.e., the SIMS) with non-Western samples. This is important – particularly in light of the assessment of asylum seekers (see van der Heide et al., 2017). The cultural differences between an asylum seeker and an evaluator, such as language and differences in describing and understanding symptoms, may significantly influence the way the asylum seeker responds to the given symptom measures (e.g., Nijdam-Jones & Rosenfeld, 2017). However, practitioners' similar views on what is considered to be an implausible complaint suggest that odd responding to symptom measures might be, independently of culture, indicative of over-reporting. Nevertheless, practitioners should always be aware of cultural differences, and should try to design the diagnostic assessment in such way that it provides the most valid answers to the question of symptom validity.

Overall, our findings encourage future investigation and implementation of a multi-method approach, including certain features of the VA (namely the non-verifiable details and the length) and the SRSI, in symptom validity assessment. We also indicate the importance of noticing the hyperbolism as malingerers' dominant reporting strategy.

Conclusion

The main purpose of this thesis was to investigate the utility of a multi-method approach in detecting fabricated symptoms. The three main goals were to 1) contribute to the field of symptom validity assessment by investigating a novel method for assessment of self-reports, the Verifiability Approach (VA); 2) critically examine a controversial method in symptom validity assessment, the Modified Stroop Task (MST); and 3) provide an initial systematic investigation of a newly developed self-report measure of over-reporting, the Self-Report Symptom Inventory (SRSI).

This research represents the first attempt to explore whether a lie-detection tool, the Verifiability Approach, could be applied to the field of symptom validity assessment. We found that evaluating the (non)verifiability of statements might add to symptom validity assessment. When asked to spontaneously provide a symptom report, malingerers generated long and non-verifiable statements. People who genuinely suffered from a physical condition, in contrast, briefly elaborated on the topic, but included a relatively high ratio of information that in principle could be checked. The findings suggest that describing a purely internal experience (e.g., symptoms), or an external event (e.g., exposure) leads to a similar reporting patterns among malingerers. Malingerers' dominant strategy was to mainly focus on elaborating their subjective experience (e.g., feelings), to provide information that is inaccessible to verification.

Our investigation of the controversial task, the Modified Stroop task, yielded unfavourable findings regarding its utility in symptom validity assessment. Previous studies also indicated the unreliability (e.g., Buckley et al., 2002; Buckley et al., 2003), and the small effects of the Modified Stroop task, that, if present at all, proved difficult to replicate (Kimble et al., 2009). Taken together with our findings, we concluded that this reaction-time task should not serve as a detection tool for fabricated symptoms.

Our research also provided a systematic investigation of the SRSR, a measure that is yet to be welcomed by the professionals in this field. With regard to symptom endorsement, we showed that malingerers extensively over-reported their complaints, by claiming to suffer from both genuine symptoms and bizarre, implausible symptoms. Truth tellers, instead, had a high endorsement of plausible symptoms, but a relatively low endorsement of pseudosymptoms. Including a mix of plausible, genuine symptoms and implausible statements, as the SRSI does, seems to be a promising method for detecting malingerers.

Another favourable finding for implementation of the SRSI is that practitioners across the world have a uniform understanding of the plausibility of certain complaints. However, considering practitioners' inability to differentiate between dissociative and pseudosymptoms, it is likely that

people who report dissociative symptoms might be mistakenly detected as a symptom over-reporter, and vice versa. This issue warrants attention and further examination.

Overall, across the eight studies presented in this thesis, all of the three main goals were achieved. Our findings suggest that malingerers' dominant reporting strategy might be explained as *hyperbolism*, and that implementing a multi-method approach, including certain features of the VA and the SRSI, would be the most efficient way of detecting it. However, this conclusion needs future inspection. The current thesis provides a solid foundation and encouragement for future investigation of two novel methods for detecting fabricated symptoms, the VA and the SRSI.

REFERENCES

- Adamou, M. C., & Hale, A. S. (2003). PTSD and the law of psychiatric injury in England and Wales: finally coming closer? *Journal of the American Academy of Psychiatry and the Law Online*, 31, 327-332.
- Akehurst, L., Easton, S., Fuller, E., Drane, G., Kuzmin, K., & Litchfield, S. (2017). An evaluation of a new tool to aid judgments of credibility in the medico-legal setting. *Legal and Criminological Psychology*, 20, 68-80. <https://doi.org/10.1111/lcrp.12079>
- Ali, S., Jabeen, S., & Alam, F. (2015). Multimodal Approach to Identifying Malingered Posttraumatic Stress Disorder: A Review. *Innovations in Clinical Neuroscience*, 12, 12-20.
- Allcott, D., Anderson, S., Friedland, D., Leng, N., Gross, M., Skelton-Robinson, M., & Weller, M. (2014). How do experts reporting for the legal process validate symptoms? The results of a survey. *Medicine, Science and the Law*, 54, 68-73. <https://doi.org/10.1093/arclin/act073>
- Amado, B. G., Arce, R., Farina, F., & Vilari-o, M. (2016). Criteria-Based Content Analysis (CBCA) reality criteria in adults: A meta-analytic review. *International Journal of Clinical and Health Psychology*, 16, 201-210. <https://doi.org/10.1016/j.ijchp.2016.01.002>
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text revision). Washington, DC: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (DSM-5®). Washington, DC: Author. <https://doi.org/10.1176/appi.books.9780890425596>
- Anestis, J. C., Gottfried, E. D., & Joiner, T. E. (2015). The utility of MMPI-2-RF Substantive Scales in prediction of negative treatment outcomes in a community mental health center. *Assessment*, 22, 23-35. <https://doi.org/10.1177/1073191114536771>
- Ardolf, B. R., Denney, R. L., & Houston, C. M. (2007). Base rates of negative response bias and malingered neurocognitive dysfunction among criminal defendants referred for neuropsychological evaluation. *The Clinical Neuropsychologist*, 21, 899-916.
- Armistead-Jehle, P. (2010). Symptom validity test performance in US veterans referred for evaluation of mild TBI. *Applied Neuropsychology*, 17, 52-59.
- Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, 38, 23-32. [https://doi.org/10.1016/0022-3999\(94\)90005-1](https://doi.org/10.1016/0022-3999(94)90005-1)
- Ball, J. R., Mitchell, P. B., Touyz, S. W., Griffiths, R. A., & Beumont, P. J. (2004). Clinical utility of the modified Stroop task as a treatment outcome measure: Questions raised. *Clinical Psychologist*, 8, 76-80.
- Bass, C., & Halligan, P. W. (2007). Illness related deception: social or psychiatric problem?. *Journal of the Royal Society of Medicine*, 100, 81-84. <https://doi.org/10.1177/014107680710000223>
- Bauer, K. (2004). Covert video surveillance of parents suspected of child abuse: the British experience and alternative approaches. *Theoretical medicine and bioethics*, 25, 311-327.
- 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.
- 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. [https://doi.org/10.1016/S0887-6185\(01\)00055-X](https://doi.org/10.1016/S0887-6185(01)00055-X)
- Berger, O., McNiel, D. E., & Binder, R. L. (2012). PTSD as a criminal defense: a review of case law. *Journal of the American Academy of Psychiatry and the Law Online*, 40, 509-521.
- Bernardy, N. C., & Friedman, M. J. (2015). Psychopharmacological strategies in the management of posttraumatic stress disorder (PTSD): What have we learned?. *Current Psychiatry Reports*, 17, 20. <https://doi.org/10.1007/s11920-015-0564-2>
- Bernstein, E. M., & Putnam, F. W. (1986). Development, reliability, and validity of a dissociation scale. *The Journal of Nervous and Mental Disease*, 174, 727-735.
- Berry, D. T., & Nelson, N. W. (2010). DSM-5 and malingering: A modest proposal. *Psychological Injury and Law*, 3, 295-303. <https://doi.org/10.1007/s12207-010-9087-7>

- Ben-Porath, Y. S. (2012). *Interpreting the MMPI-2-RF*. Minneapolis, MN: University of Minnesota Press.
- Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal*, 5, 404–417. doi:10.1016/j.spinee.2004.11.016
- Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist*, 15, 19–45. <https://doi.org/10.1076/clin.15.1.19.1907>
- Bielecki, M., Popiel, A., Zawadzki, B., & Sedek, G. (2017). Age As moderator of emotional stroop task performance in posttraumatic stress disorder (PTSD). *Frontiers in Psychology*, 8, 1614. <https://doi.org/10.3389/fpsyg.2017.01614>
- Blandon-Gitlin, I., Pezdek, K., Lindsay, D. S., & Hagen, L. (2009). Criteria-based content analysis of true and suggested accounts of events. *Applied Cognitive Psychology*, 23, 901–917. doi:10.1002/acp.1504
- Blaney, P. H. (2015). Medical model of mental disorders. *Encyclopedia of Clinical Psychology*. <https://doi.org/10.1002/9781118625392.wbecp382>
- Bogaard, G., Meijer, E. H., Vrij, A., Broers, N. J., & Merckelbach, H. (2014). Contextual bias in verbal credibility assessment: Criteria-based content analysis, reality monitoring and scientific content analysis. *Applied Cognitive Psychology*, 28, 79-90. <https://doi.org/10.1002/acp.2959>
- Bonanno, G. A. (2005). Resilience in the face of potential trauma. *Current directions in psychological science*, 14, 135-138.
- Boskovic, I., Biermans, A., Merten, T., Jelcic, 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*. <https://doi.org/10.3389/fpsyg.2018.01195>
- Boskovic, I., Bogaard, G., Merckelbach, H., Vrij, A., & Hope, L. (2017). The Verifiability Approach to detection of malingered physical symptoms. *Psychology, Crime & Law*, 23, 717-729.
- Boskovic, I., Hope, L., Ost, J., Orthey, R., & Merckelbach, H. (2017). *Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task*. Manuscript under review.
- Boskovic, I., Merckelbach, H., Merten, T., Hope, L., & Jelcic, M. (2018). *The Self-Report Symptom Inventory as a screen for over-reporting: An explorative study with instructed simulants*. Accepted manuscript pending revision.
- Boskovic, I., Peters, M., Merckelbach, H. (2018). *Exposure narratives and symptom reports among patients and feigners*. Ongoing research.
- Boskovic, I., Reymarks, A., & Merckelbach, H. (2018). *Telling a good story 2.0.: Fantasy proneness and the Verifiability Approach*. Manuscript in preparation.
- Boskovic, I., van der Heide, D., Hope, L., Merckelbach, H., & Jelcic, M. (2017). Plausibility judgments of atypical symptoms across cultures: An explorative study among western and non-western experts. *Psychological Injury and Law*, 10, 274-281.
- Boskovic, I., Zwaan, L., & Merckelbach, H. (2018). *Stability of common symptoms among truth tellers and feigners*. Manuscript in preparation.
- Boyer, M., & Dickerson, M. (2003). Attentional bias and addictive behaviour: automaticity in a gambling-specific modified Stroop task. *Addiction*, 98, 61-70.
- Buckley, T. C., Blanchard, E. B., & Hickling, E. J. (2002). Automatic and strategic processing of threat stimuli: A comparison between PTSD, panic disorder, and nonanxiety controls. *Cognitive Therapy and Research*, 26, 97-115. <https://doi.org/10.1023/A:1013897805918>
- 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. <https://doi.org/10.1023/A:1022063412056>

- Burges, C., & McMillan, T. M. (2001). The ability of naive participants to report symptoms of post-traumatic stress disorder. *British Journal of Clinical Psychology*, 40, 209-214. <https://doi.org/10.1348/014466501163544>
- Burkett, B. G., Whitley, G., & Press, V. (1998). *Stolen valor*. Dallas, TX: Verity Press Publishing.
- Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law*, 7, 197-205. <https://doi.org/10.1007/s12207-014-9198-7>
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2. manual for administration and scoring*. Minneapolis MN: University of Minnesota Press.
- Cannon, B. J. (2003). An Emotional Stroop Effect to Malingering Related Words. *Perceptual and Motor Skills*, 96, 827-834. <https://doi.org/10.2466/pms.2003.96.3.827>
- Chafetz, M. D. (2008). Malingering on the social security disability consultative exam: Predictors and base rates. *The Clinical Neuropsychologist*, 22, 529-546. <https://doi.org/10.1080/13854040701346104>
- Chafetz, M. D., Abrahams, J. P., & Kohlmaier, J. (2007). Malingering on the Social Security disability consultative exam: A new rating scale. *Archives of Clinical Neuropsychology*, 22, 1-14.
- Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., ... & Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, 29, 723-740. <https://doi.org/10.1080/13854046.2015.1099738>
- Chafetz, M., & Underhill, J. (2013). Estimated costs of malingered disability. *Archives of Clinical Neuropsychology*, 28, 633-639. <https://doi.org/10.1093/arclin/act038>
- Charles, C., Gafni, A., Whelan, T., & O'Brien, M. A. (2006). Cultural influences on the physician–patient encounter: the case of shared treatment decision-making. *Patient Education and Counseling*, 63, 262-267. <https://doi.org/10.1016/j.pec.2006.06.018>
- Cima, M., Merckelbach, H., Hollnack, S., Butt, C., Kremer, K., Schellbach-Matties, R., & Muris, P. (2003). The other side of malingering: supernormality. *The Clinical Neuropsychologist*, 17, 235-243. <https://doi.org/10.1076/clin.17.2.235.16507>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, Z. E., & Appelbaum, P. S. (2016). Experience and opinions of forensic psychiatrists regarding PTSD in criminal cases. *Journal of the American Academy of Psychiatry and the Law Online*, 44, 41-52.
- Constans, J. I., Kimbrell, T. A., Nanney, J., 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. <https://doi.org/10.1037/a0035100>
- Conroy, M. A., & Kwartner, P. P. (2006). The definition of malingering. *Applied Psychology in Criminal Justice*, 2, 30-51.
- Correa, A. A., & Rogers, R. (2010). Cross-cultural applications of the PAI. *Clinical Applications of the Personality Assessment Inventory*, 135-148.
- Creamer, M., Bell, R., & Failla, S. (2003). Psychometric properties of the impact of event scale—revised. *Behaviour Research and Therapy*, 41, 1489-1496.
- Crichton, A. H., Wygant, D. B., Applegate, K. C., Umlauf, R. L., & Granacher, R. P. (2014). Can brief measures effectively screen for pain and somatic malingering? Examination of the modified somatic perception questionnaire and pain disability index. *The Spine Journal*, 14, 2042–2050.
- Culhane, S. E., Hosch, H. M., & Kehn, A. (2008). Alibi generation: Data from US Hispanics and US non-Hispanic whites. *Journal of Ethnicity in Criminal Justice*, 6, 177–199.
- Cummings, C. M., Caporino, N. E., & Kendall, P. C. (2014). Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychological Bulletin*, 140, 816-845. <http://dx.doi.org/10.1037/a0034733>

- Dandachi-FitzGerald, B., & Merckelbach, H. (2013). Feigning a memory deficit: The medical symptom validity test as an example. *Journal of Experimental Psychopathology*, 4, 46–63. Doi:10.5127/jep.025511
- Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28, 771–783. <https://doi.org/10.1093/arclin/act073>
- De Beurs, E., 2011. *Brief Symptom Inventory (BSI) en BSI 18. Handleiding 2011*. Leiden, The Netherlands: PITS.
- Department of Veterans Affairs Office of Inspector General. (2005). *Review of State Variances in VA Disability Compensation Payments*. Washington: Department of Veterans Affairs Office of Inspector General.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, 129, 74. <https://doi.org/10.1037/0033-2909.129.1.74>
- Derogatis, L. R. (2001). *BSI 18, Brief Symptom Inventory 18: Administration, scoring and procedures manual*. Minneapolis, Minn: NCS Pearson, Incorporated.
- Dibbets, P., & Schulte-Ostermann, M. A. (2015). Virtual reality, real emotions: a novel analogue for the assessment of risk factors of post-traumatic stress disorder. *Frontiers in psychology*, 6, 681. doi: 10.3389/fpsyg.2015.00681
- Douglas-Hamilton, J. (2016). *The Truth about Rudolf Hess*. Frontline Books.
- Dückers, M. L., Alisic, E., & Brewin, C. R. (2016). A vulnerability paradox in the cross-national prevalence of post-traumatic stress disorder. *The British Journal of Psychiatry*, 209, 300–305. <https://doi.org/10.1192/bjp.bp.115.176628>
- Edens, J. F., Poythress, N. G., & Watkins-Clay, M. M. (2007). Detection of malingering in psychiatric unit and general population prison inmates: A comparison of the PAI, SIMS, and SIRS. *Journal of Personality Assessment*, 88, 33–42.
- Eekhout, I., Reijnen, A., Vermetten, E., & Geuze, E. (2016). Post-traumatic stress symptoms 5 years after military deployment to Afghanistan: an observational cohort study. *The Lancet Psychiatry*, 3, 58–64.
- Elal, G., & Slade, P. (2005). Traumatic exposure severity scale (TESS): A measure of exposure to major disasters. *Journal of Traumatic Stress*, 18, 213–220.
- Elhai, J. D., Flitter, J. M. K., Gold, S. N., & Sellers, A. H. (2001). Identifying subtypes of women survivors of childhood sexual abuse: An MMPI-2 cluster analysis. *Journal of Traumatic Stress*, 14, 157–175.
- 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.
- Ernberg, E., Magnusson, M., Landström, S., & Tidefors, I. (2018). Court evaluations of young children's testimony in child sexual abuse cases. *Legal and Criminological Psychology*. <https://doi.org/10.1111/lcrp.12124>
- Ferrari, R., Constantoyannis, C., & Papadakis, N. (2001). Cross-cultural study of symptom expectation following minor head injury in Canada and Greece. *Clinical Neurology and Neurosurgery*, 103, 254–259. [https://doi.org/10.1016/S0303-8467\(01\)00161-5](https://doi.org/10.1016/S0303-8467(01)00161-5)
- Fishbain, D. A., Cutler, R., Rosomoff, H. L., & Rosomoff, R. S. (1999). Chronic pain disability exaggeration/malingering and submaximal effort research. *The Clinical Journal of Pain*, 15, 244–274. ISSN: 0749-8047.
- Fisher, P. L., & Durham, R. C. (1999). Recovery rates in generalized anxiety disorder following psychological therapy: an analysis of clinically significant change in the STAI-T across outcome studies since 1999. *Psychological Medicine*, 29, 1425–1434.
- Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology*, 8, 196–202. <https://doi.org/10.1017/iop.2015.25>

- Freeman, T., Powell, M., & Kimbrell, T. (2008). Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. *Psychiatry Research*, 158, 374-380. <https://doi.org/10.1016/j.psychres.2007.04.002>
- Friel, A., White, T., & Hull, A. (2008). Posttraumatic stress disorder and criminal responsibility. *The Journal of Forensic Psychiatry & Psychology*, 19, 64-85. <https://doi.org/10.1080/14789940701594736>
- Frueh, B. C., Hamner, M. B., Cahill, S. P., Gold, P. B., & Hamlin, K. L. (2000). Apparent symptom overreporting in combat veterans evaluated for PTSD. *Clinical Psychology Review*, 20, 853-885. [https://doi.org/10.1016/S0272-7358\(99\)00015-X](https://doi.org/10.1016/S0272-7358(99)00015-X)
- Gacono, C. B., Meloy, J. R., Sheppard, K., Speth, E., & Roske, A. (1995). A clinical investigation of malingering and psychopathy in hospitalized insanity acquittees. *Journal of the American Academy of Psychiatry and the Law Online*, 23, 387-397.
- Giger, P., & Merten, T. (2013). Swiss population-based reference data for six symptom validity tests. *Clínica y Salud*, 24, 153-159. doi:10.1016/S1130-5274(13)70016-1
- Giger, P., & Merten, T. (in press). Equivalence of the German and the French versions of the Self-Report Symptom Inventory. *Swiss Journal of Psychology*.
- Giosan, C., Glovsky, V., & Haslam, N. (2001). The lay concept of 'mental disorder': a cross-cultural study. *Transcultural Psychiatry*, 38, 317-332. <https://doi.org/10.1177/136346150103800303>
- Gnisci, A., Caso, L., & Vrij, A. (2010). Have you made up your story? The effect of suspicion and liars' strategies on reality monitoring. *Applied Cognitive Psychology*, 24, 762-773.
- Göbber, J., Petermann, F., Piegza, M., & Kobelt, A. (2012). Beschwerdevalidierung bei Rehabilitanden mit Migrationshintergrund in der Psychosomatik [Symptom validation in patients with migration background in psychosomatic medicine]. *Rehabilitation*, 51, 356-364.
- Granhag, P. A., Andersson, L. O., Strömwall, L. A., & Hartwig, M. (2004). Imprisoned knowledge: Criminals' beliefs about deception. *Legal and Criminological Psychology*, 9, 103-119.
- Green, S. B. & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (fourth edition). New Jersey: Pearson
- Greiffenstein, M. F., & Baker, W. J. (2006). Miller was (mostly) right: Head injury severity inversely related to simulation. *Legal and Criminological Psychology*, 11, 131-145. <https://doi.org/10.1348/135532505X49828>
- Greve, K. W., Bianchini, K. J., & Brewer, S. T. (2013). The assessment of performance and self-report validity in persons claiming pain-related disability. *The Clinical Neuropsychologist*, 27, 108-137. doi:10.1080/13854046.2012.739646
- Greve, K. W., Bianchini, K. J., Etherton, J. L., Meyers, J. E., Curtis, K. L., & Ord, J. S. (2010). The Reliable Digit Span test in chronic pain: Classification accuracy in detecting malingered pain-related disability. *The Clinical Neuropsychologist*, 24, 137-152. <https://doi.org/10.1080/13854040902927546>
- Greve, K. W., Ord, J. S., Bianchini, K. J., & Curtis, K. L. (2009). Prevalence of malingering in patients with chronic pain referred for psychologic evaluation in a medico-legal context. *Archives of Physical Medicine and Rehabilitation*, 90, 1117-1126. doi:10.1016/j.apmr.2009.01.018
- Guriel, J., & Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A critical review. *Clinical Psychology Review*, 23, 881-904. <https://doi.org/10.1016/j.cpr.2003.07.001>
- Guriel, J., & Fremouw, W. (2003). Assessing malingered posttraumatic stress disorder: A critical review. *Clinical Psychology Review*, 23, 881-904.
- Hall, R. C., & Hall, R. C. (2007). Detection of malingered PTSD: an overview of clinical, psychometric, and physiological assessment: where do we stand? *Journal of Forensic Sciences*, 52, 717-725.
- Hall, R., Hall, R., Chapman, M. (2006). Effects of terrorist attacks on the elderly: Part 2. Posttraumatic stress, acute stress, and affective disorders. *Clinical Geriatrics*, 14, 17-24

- Hamilton, J. C., & Feldman, M. D. (2001). Chest pain in patients who are malingering. In J. W. Hurst & D. C. Morris (Eds.), *Chest pain* (pp. 443–456). Armonk, NY: Futura Publishing.
- Hare, R. D. (2003). *The Psychopathy Checklist-Revised*. Toronto, ON.
- Hartshorne, H., & May, M. A. (1928). *Studies in deceit*. New York: Macmillan.
- Hartwig, M., Anders Granhag, P., & Strömwall, L. A. (2007). Guilty and innocent suspects' strategies during police interrogations. *Psychology, Crime & Law*, 13, 213–227.
- Harvey, A. C., Vrij, A., Nahari, G., & Ludwig, K. (2017). Applying the Verifiability Approach to insurance claims settings: Exploring the effect of the information protocol. *Legal and Criminological Psychology*, 22, 47–59. <https://doi.org/10.1111/lcrp.12092>
- Harzing, A. W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management*, 6, 243–266. <https://doi.org/10.1093/arclin/act073>
- Hauch, V., Sporer, S. L., Masip, J., & Blandón-Gitlin, I. (2017). Can credibility criteria be assessed reliably? A meta-analysis of criteria-based content analysis. *Psychological Assessment*, 29, 819–823. <https://doi.org/10.1037/pas0000426>
- Hauch, V., Sporer, S. L., Michael, S., & Meissner, C. A. (2016). Does training improve the detection of deception? A meta-analysis. *Communication Research*, 43, 283–343. <http://dx.doi.org/10.1177/0093650214534974>
- Hausotter, W., & Schouler-Ocak, M. (2007). *Begutachtung von Menschen mit Migrationshintergrund und Arbeitnehmern nichtdeutscher Herkunft: Unter medizinischen und psychologischen Aspekten (1. Aufl.)*. München: Elsevier Urban & Fischer.
- Heilbrunner, R. L. (Ed.). (2005). *Forensic Neuropsychology Casebook*. Guilford Press.
- Helvoort, D., Merckelbach, H., & Merten (under review). The Self-Report Symptom Inventory (SRS): Validation in forensic inpatients with severe psychopathology.
- Herlihy, J., Scragg, P., & Turner, S. (2002). Discrepancies in autobiographical memories—implications for the assessment of asylum seekers: repeated interviews study. *Bmj*, 324, 324–327.
- Henderson, A. S., Byrne, D. G. and Duncan-Jones, P. (1981). *Neurosis and the social environment*. New York: Academic Press.
- Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Thousand Oaks, CA: Sage.
- Hiscock, M., & Hiscock, D. (1989). Refining the forced-choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, 11, 967–974.
- Hosch, H. M., Culhane, S. E., Tubb, V. A., & Granillo, E. A. (2011). Town vs. gown: A direct comparison of community residents and student mock jurors. *Behavioral Sciences & The Law*, 29, 452–466.
- Huntington, P. S. (1993). The clash of civilizations. *Foreign Affairs*, 72, 22–49.
- Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, 30, 473–496.
- Iverson, G. L., & Binder, L. M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, 15, 829–858.
- Jackson, J. C., Sinnott, P. L., Marx, B. P., Murdoch, M., Sayer, N. A., Alvarez, J. M., Greevy, R. M., Schnurr, P. P., Friedman, M. J., Shane, A. C., Owen, R. R., Keane, T. M., & Speroff, T. (2011). Variation in practices and attitudes of clinicians assessing PTSD-related disability among veterans. *Journal of Traumatic Stress*, 24, 609–613. <https://doi.org/10.1002/jts.20688>
- Jelicic, M., Merckelbach, H., & Boskovic, I. (2018). Seven myths about malingering. In Otgaar, H., & Howe, M. L. (Eds.) (2018). *Finding the Truth in the Courtroom: Dealing with Deception, Lies, and Memories*. Oxford University Press.
- Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61, 760–771.
- Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88, 67. <https://doi.org/10.1023/A:1022063412056>

- Johnson, R. C., Edman, J. L., & Danko, G. P. (1995). Self reported negative experiences and dissociation. *Personality and Individual Differences*, 18, 793-795.
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36, 264-277. <https://doi.org/10.1177/0022022104272905>
- Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36, 264-277.
- Jones, S. M. (2017). Dissimulation strategies on standard neuropsychological tests: A qualitative investigation. *Brain injury*, 31, 1131-1141. <https://doi.org/10.1080/02699052.2017.1283444>
- Jupe, L. M., Leal, S., Vrij, A., & Nahari, G. (2017). Applying the Verifiability Approach in an international airport setting. *Psychology, Crime & Law*, 23, 812-825.
- 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.
- Keesler, M. E., McClung, K., Meredith-Duliba, T., Williams, K., & Swirsky-Sacchetti, T. (2017). Red flags in the clinical interview may forecast invalid neuropsychological testing. *The Clinical Neuropsychologist*, 31, 619-631.
- Keogh, E., Bond, F. W., French, C. C., Richards, A., & Davis, R. E. (2004). Test anxiety, susceptibility to distraction and examination performance. *Anxiety, Stress, and Coping*, 17, 241-252. [Doi:10.1080/10615300410001703472](https://doi.org/10.1080/10615300410001703472)
- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, 34, 119-138. <https://doi.org/10.1146/annurev-publhealth-031912-114409>
- 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. <https://doi.org/10.1016/j.janxdis.2009.02.002>
- Kitchen, R. (2003). Investigating benefit fraud and illness deception in the United Kingdom. In P. W., Halligan, C., Bass, & D. A., Oakley (Eds.), *Malingering and illness deception*. New York, NY: Oxford University Press.
- Kleinman, A., & Cohen, A. (1997). Psychiatry's global challenge. *Scientific American*, 276, 86-91. <http://www.jstor.org/stable/24993661>
- Knoll, J., & Resnick, P. J. (2006). The detection of malingered post-traumatic stress disorder. *Psychiatric Clinics*, 29, 629-647.
- 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.
- Köhnken, G. (2004). Statement validity analysis and the 'detection of the truth'. In Graham, P-A. & Stromwall, L.A (Eds). (2004). *The Detection of Deception in Forensic Contexts* (pp. 41-63). Cambridge University Press. <https://doi.org/10.1017/CBO9780511490071.003>
- Kolk, A. M., Hanewald, G. J., Schagen, S., & van Wijk, C. M. G. (2003). A symptom perception approach to common physical symptoms. *Social Science & Medicine*, 57, 2343-2354. [doi:10.1016/S0277-9536\(02\)00451-3](https://doi.org/10.1016/S0277-9536(02)00451-3)
- Kooiman, C. G., Spinhoven, P., & Trijsburg, R. W. (2002). The assessment of alexithymia: a critical review of the literature and a psychometric study of the Toronto Alexithymia Scale-20. *Journal of psychosomatic research*, 53, 1083-1090.
- Kramer, D. A., & Goldman, M. S. (2003). Using a modified Stroop task to implicitly discern the cognitive organization of alcohol expectancies. *Journal of Abnormal Psychology*, 112, 171.
- Kucyi, A., & Davis, K. D. (2015). The dynamic pain connectome. *Trends in Neurosciences*, 38, 86-95. <https://doi.org/10.1016/j.tins.2014.11.006>
- Kucyi, A., Scheinman, A., & Defrin, R. (2015). Distinguishing feigned from sincere performance in psychophysical pain testing. *The Journal of Pain*, 16, 1044-1053.

- Kuperman, V. (2006). Narratives of psychiatric malingering in works of fiction. *Medical Humanities*, 32, 67–72.
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410-425. <https://doi.org/10.1076/clin.17.3.410.18089>
- Larrabee, G.J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 625-631.
- Lawson, D. (2006). Test Anxiety: a test of attentional bias. *Electronic Theses and Dissertations*, paper 38, 1-123. Retrieved 1st February 2017, from: <https://library.umaine.edu/theses/pdf/LawsonDJ2006.pdf>
- Leal, S., Vrij, A., Warmelink, L., Vernham, Z., & Fisher, R. P. (2015). You cannot hide your telephone lies: Providing a model statement as an aid to detect deception in insurance telephone calls. *Legal and Criminological Psychology*, 20, 129–146. doi:10.1111/lcrp.12017
- Lees–Haley, P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, 53, 745-755. [https://doi.org/10.1002/\(SICI\)1097-4679\(199711\)53:7<745::AID-JCLP13>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-4679(199711)53:7<745::AID-JCLP13>3.0.CO;2-L)
- Leins, D. A., Fisher, R. P., & Ross, S. J. (2013). Exploring liars' strategies for creating deceptive reports. *Legal and Criminological Psychology*, 18, 141-151.
- Levashina, J., Weekley, J. A., Roulin, N., & Hauck, E. (2014). Using Blatant Extreme Responding for Detecting Faking in High-stakes Selection: Construct validity, relationship with general mental ability, and subgroup differences. *International Journal of Selection and Assessment*, 22, 371-383.
- Lippa, S. M. (2017). Performance validity testing in neuropsychology: a clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, 1-31.
- Lovett, M. (2005). A strategy-based interpretation of Stroop. *Cognitive Science*, 29, 493-524. Doi: 10.1207/s15516709cog0000_24
- MacLeod, C. & Rutherford, E.M. (1992). Anxiety and the selective processing of emotional information. *Behaviour Research and Therapy*, 30, 479-495. [Doi:10.1016/0005-7967\(92\)90032-C](https://doi.org/10.1016/0005-7967(92)90032-C)
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95, 15-20.
- MacNeil, B. M., & Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences*, 41, 641-651. <https://doi.org/10.1016/j.paid.2006.03.004>
- Malcore, S. A., Schutte, C., Van Dyke, S. A., & Axelrod, B. N. (2015). The development of a reduced-item Structured Inventory of Malingered Symptomatology (SIMS). *Psychological Injury and Law*, 8, 95-99.
- Mangiulli, I., Lanciano, T., Jelacic, M., van Oorsouw, K., Battista, F., & Curci, A. (2018). Can implicit measures detect source information in crime-related amnesia?. *Memory*, 1-11. <https://doi.org/10.1037/a0035100>
- Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist*, 29, 741-776.
- Masip, J., Sporer, S. L., Garrido, E., & Herrero, C. (2005). The detection of deception with the reality monitoring approach: A review of the empirical evidence. *Psychology, Crime & Law*, 11, 99-122. <https://doi.org/10.1080/10683160410001726356>
- Mathews, A., & MacLeod, C. (1985). Selective processing of threat cues in anxiety states. *Behaviour Research and Therapy*, 23, 563-569. [https://doi.org/10.1016/0005-7967\(85\)90104-4](https://doi.org/10.1016/0005-7967(85)90104-4)
- 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
- McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist*, 23, 1050-1066. <https://doi.org/10.1080/13854040802665790>

- McDermott, B. E., & Feldman, M. D. (2007). Malingering in the medical setting. *Psychiatric Clinics of North America*, 30, 645–662. doi:10.1016/j.psc.2007.07.007
- McDermott, B. E., Dualan, I. V., & Scott, C. L. (2013). Malingering in the correctional system: does incentive affect prevalence?. *International Journal of Law and Psychiatry*, 36, 287-292. <https://doi.org/10.1016/j.ijlp.2013.04.013>
- McGrath, R. E., Mitchell, M., Kim, B. H., & Hough, L. (2010). Evidence for response bias as a source of error variance in applied assessment. *Psychological Bulletin*, 136, 450. <https://doi.org/10.1037/a0019216>
- 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.
- Meijer, R., Vries, R., Bruggen, V. (2011). An evaluation of the Brief Symptom Inventory–18 using item response theory: which items are most strongly related to psychological distress. *Psychological Assessment*, 23, 193-202. Doi: 10.1037/a0021292.
- Merckelbach, H. (2004). Telling a good story: fantasy proneness and the quality of fabricated memories. *Personality and Individual Differences*, 37, 1371-1382.
- Merckelbach, H., & Smith, G. P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, 18, 145-152.
- Merckelbach, H., à Campo, J., Hardy, S., & Giesbrecht, T. (2005). Dissociation and fantasy proneness in psychiatric patients: a preliminary study. *Comprehensive Psychiatry*, 46, 181-185. <https://doi.org/10.1016/j.comppsy.2004.08.001>
- Merckelbach, H., Boskovic, I., Pesy, D., Dalsklev, M., & Lynn, S. J. (2017). Symptom overreporting and dissociative experiences: a qualitative review. *Consciousness and Cognition*, 49, 132-144. <https://doi.org/10.1016/j.concog.2017.01.007>
- Merckelbach, H., Giesbrecht, T., van Heugten-van der Kloet, D., Jong, J. D., Meyer, T., & Rietman, K. (2015). The overlap between dissociative symptoms and symptom over-reporting. *The European Journal of Psychiatry*, 29, 165-172.
- 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.
- Merckelbach, H., Jelicic, M., & Pieters, M. (2011). The residual effect of feigning: How intentional faking may evolve into a less conscious form of symptom reporting. *Journal of Clinical and Experimental Neuropsychology*, 33, 131-139. <https://doi.org/10.1080/13803395.2010.495055>
- Merckelbach, H., Langeland, W., de Vries, G., & Draijer, N. (2014). Symptom overreporting obscures the dose–response relationship between trauma severity and symptoms. *Psychiatry Research*, 217, 215-219. <https://doi.org/10.1016/j.psychres.2014.03.018>
- Merckelbach, H., Muris, P., Horselenberg, R., & Stougie, S. (2000). Dissociative experiences, response bias, and fantasy proneness in college students. *Personality and Individual Differences*, 28, 49-58. [https://doi.org/10.1016/S0191-8869\(99\)00079-3](https://doi.org/10.1016/S0191-8869(99)00079-3)
- Merckelbach, H., Prins, C., Boskovic, I., Niesten, I., & À Campo, J. (2018). Alexithymia as a potential source of symptom over-reporting: An exploratory study in forensic patients and non-forensic participants. *Scandinavian Journal of Psychology*. <https://doi.org/10.1111/sjop.12427>
- Merckelbach, H., Smeets, T., & Jelicic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *The Journal of Forensic Psychiatry & Psychology*, 20, 378-386. <https://doi.org/10.1080/14789940802456686>
- Merten, T., Friedel, E., & Stevens, A. (2007) Die Authentizität der Beschwerdenschilderung in der neurologisch-psychiatrischen Begutachtung: eine Untersuchung mit dem Strukturierten Fragebogen Simulierter Symptome [Authenticity of symptom reports in independent neurological and psychiatric

- examinations: A study with the Structured Inventory of Malingered Symptomatology]. *Praxis der Rechtspsychologie*, 17, 140–154.
- Merten, T., Giger, P., Merckelbach, H., & Stevens, A. (in press). *Self-Report Symptom Inventory (SRSI). Handbuch* [SRSI professional manual]. Göttingen, Germany: Hogrefe.
- Merten, T. & Merckelbach, H. (2013). Forced-choice tests as single-case experiments in the differential diagnosis of intentional symptom distortion. *Journal of Experimental Psychopathology*, 4, 20-37.
- Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law*, 6, 122-137. <https://doi.org/10.1007/s12207-013-9155-x>
- Merten, T. (2017). Logical paradoxes and paradoxical constellations in medicolegal assessment. *Psychological Injury and Law*, 10, 264 – 273. <https://doi.org/10.1007/s12207-017-9297-3>
- Merten, T., & Rogers, R. (2017). An international perspective on feigned mental disabilities: conceptual issues and continuing controversies. *Behavioral Sciences & the Law*, 35, 97-112. <https://doi.org/10.1002/bsl.2274>
- Merten, T., Merckelbach, H., Giger, P., & Stevens, A. (2016). The Self-Report Symptom Inventory (SRSI): A new instrument for the assessment of distorted symptom endorsement. *Psychological Injury and Law*, 9, 102-111. <https://doi.org/10.1007/s12207-016-9257-3>
- Merten, T., Thies, E., Schneider, K., & Stevens, A. (2009). Symptom validity testing in claimants with alleged posttraumatic stress disorder: Comparing the Morel Emotional Numbing Test, the Structured Inventory of Malingered Symptomatology, and the Word Memory Test. *Psychological Injury and Law*, 2, 284-293.
- Meyer, J. F., Faust, K. A., Faust, D., Baker, A. M., & Cook, N. E. (2013). Careless and random responding on clinical and research measures in the addictions: A concerning problem and investigation of their detection. *International Journal of Mental Health and Addiction*, 11, 292-306. <https://doi.org/10.1007/s11469-012-9410-5>
- Miller, H. A. (2001). *M-FAST: Miller-Forensic Assessment of Symptoms Test professional manual*. Odessa, FL: Psychological Assessment Resources, Inc
- Miller, L. S., Boyd, M. C., Cohn, A., Wilson, J. S., & McFarland, M. (2006, February). Prevalence of sub-optimal effort in disability applicants. *Poster presented in 34th annual meeting of the International Neuropsychological Society, Boston, US*.
- Miranda, H., Gold, J. E., Gore, R., & Punnett, L. (2006). Recall of prior musculoskeletal pain. *Scandinavian Journal of Work, Environment & Health*, 32, 294-299.
- Mittenberg, W., Patton, C., Canyock, E. M., & Condit, D. C. (2002). Base rates of malingering and symptom exaggeration. *Journal of Clinical and Experimental Neuropsychology*, 24, 1094-1102. <https://doi.org/10.1076/jcen.24.8.1094.8379>
- Mogg, K., & Marden, B. (1990). Processing of emotional information in anxious subjects. *British Journal of Clinical Psychology*, 29, 227-229. Doi: 10.1111/j.2044-8260.1990.tb00874.x
- 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.
- Morel, K. R., & Marshman, K. C. (2008). Critiquing symptom validity tests for posttraumatic stress disorder: A modification of Hartman's criteria. *Journal of Anxiety Disorders*, 22, 1542-1550. <https://doi.org/10.1016/j.janxdis.2008.03.008>
- Morey, L. C. (2007). *The Personality Assessment Inventory professional manual* (2nd ed.). Odessa, FL: Psychological Assessment Resources.
- Moul, D. E., Nofzinger, E. A., Pilkonis, P. A., Houck, P. R., Miewald, J. M., & Buysse, D. J. (2002). Symptom reports in severe chronic insomnia. *Sleep*, 25, 548-558. <https://doi.org/10.1093/sleep/25.5.548>
- Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences*, 32, 337-348.
- Musso, M. W., & Gouvier, W. D. (2012). “Why is this so hard?”: A review of detection of malingered ADHD in college students. *Journal of Attention Disorders*, 18, 186-201. Doi: 10.1177/1087054712441970

- Nahari, G. (2018). The Applicability of the Verifiability Approach to the Real World. In Rosenfeld, P. J. (Ed.) (2018). *Detecting Concealed Information and Deception* (pp. 329-351). London, UK: Elsevier Inc. <https://doi.org/10.1016/B978-0-12-812729-2.00014-8>
- Nahari, G., & Vrij, A. (2014). Can I borrow your alibi? The applicability of the Verifiability Approach to the case of an alibi witness. *Journal of Applied Research in Memory and Cognition*, 3, 89-94. Doi:10.1016/j.jarmac.2014.04.005
- Nahari, G., Leal, S., Vrij, A., Warmelink, L., & Vernham, Z. (2014). Did somebody see it? Applying the Verifiability Approach to insurance claim interviews. *Journal of Investigative Psychology and Offender Profiling*, 11, 237-243. Doi:10.1002/jip.1417
- Nahari, G., Vrij, A., & Fisher, R. P. (2012). Does the truth come out in the writing? Scan as a lie detection tool. *Law and Human Behavior*, 36, 68-76. <https://doi.org/10.1037/h0093965>
- Nahari, G., Vrij, A., & Fisher, R. P. (2014a). Exploiting liars' verbal strategies by examining the verifiability of details. *Legal and Criminological Psychology*, 19, 227-239. <https://doi.org/10.1111/j.2044-8333.2012.02069.x>
- Nahari, G., Vrij, A., & Fisher, R. P. (2014b). The Verifiability Approach: Countermeasures facilitate its ability to discriminate between truths and lies. *Applied Cognitive Psychology*, 28, 122-128. <https://doi.org/10.1002/acp.2974>
- Nelson, J. M., Lindstrom, W., & Foels, P. A. (2014). Test anxiety and college students with attention deficit hyperactivity disorder. *Journal of Psychoeducational Assessment*, 32, 548-557. Doi:10.1177/0734282914521978
- Nies, K. J., & Sweet, J. J. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology*, 9, 501-552. <https://doi.org/10.1093/arclin/9.6.501>
- Niesten, I. J., Nentjes, L., Merckelbach, H., & Bernstein, D. P. (2015). Antisocial features and "faking bad": A critical note. *International Journal of Law and Psychiatry*, 41, 34-42. <https://doi.org/10.1016/j.ijlp.2015.03.005>
- Niesten, I., Merckelbach, H., van Impelen, A., Jelicic, M., Manderson, A., & Cheng, M. (2017). A lab model for symptom exaggeration: What do we need? *Journal of Experimental Psychopathology*, 8, 55-75. <http://doi/10.5127/jep.051815>
- Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment*, 29, 1321. <https://doi.org/10.1037/pas0000438>
- Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2017). A cross-cultural analysis of the Test of Memory Malingering among Latin American Spanish-speaking adults. *Law and Human Behavior*, 41, 422-428. <http://doi:10.1037/lhb0000250>
- Oberlander, V., Naefgen, C., Koppehele-Gossel, J., Quinten, L., Banse, R., & Schmidt, A. F. (2016). Validity of content-based techniques to distinguish true and fabricated statements: A meta-analysis. *Law and Human Behavior*, 40, 440-457.
- Pankratz, L., Fausti, S.A., & Peed, S. (1975). Case study: A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology*, 43, 421-422.
- Paulhus, D. L. (2011). Overclaiming on Personality Questionnaires. In M. Ziegler, C. MacCann & R. D. Roberts (Eds.) (2011). *New Perspectives on Faking in Personality Assessment* (pp. 151- 164). New York, NY: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195387476.003.0045>
- 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. <https://doi.org/10.1007/s12207-011-9102-7>
- Peace, K. A., & Porter, S. (2011). Remembrance of lies past: A comparison of the features and consistency of truthful and fabricated trauma narratives. *Applied Cognitive Psychology*, 25, 414-423.

- Peace, K. A., Porter, S., & Cook, B. L. (2010). Investigating differences in truthful and fabricated symptoms of traumatic stress over time. *Psychological Injury and Law*, 3, 118-129. <https://doi.org/10.1007/s12207-010-9078-8>
- Peace, K., & Richards, V. (2014). Faking it: incentives and malingered PTSD. *Journal of Criminal Psychology*, 4, 19-32.
- Petrie, K. J., Faasse, K., Crichton, F., & Grey, A. (2014). How common are symptoms? Evidence from a New Zealand national telephone survey. *BMJ Open*, 4, e005374. doi:10.1136/bmjopen-2014-005374
- Pollock, P. H., Quigley, B., Worley, K. O., & Bashford, C. (1997). Feigned mental disorder in prisoners referred to forensic mental health services. *Journal of psychiatric and mental health nursing*, 4, 9-15. <https://doi.org/10.1111/j.1365-2850.1997.tb00171.x>
- Porter, S., Peace, K. A., & Emmett, K. A. (2007). You protest too much, methinks: Investigating the features of truthful and fabricated reports of traumatic experiences. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, 39, 79-91.
- Porter, S., Yuille, J. C., & Lehman, D. R. (1999). The nature of real, implanted, and fabricated memories for emotional childhood events: implications for the recovered memory debate. *Law and Human Behavior*, 23, 517- 537.
- Poyner, G. (2010). Psychological evaluations of veterans claiming PTSD disability with the Department of Veterans Affairs: A clinician's viewpoint. *Psychological Injury and Law*, 3, 130-132. <https://doi.org/10.1007/s12207-010-9076-x>
- Price, S. A., & Hanson, R. K. (2007). A modified Stroop task with sexual offenders: Replication of a study. *Journal of Sexual Aggression*, 13, 203-216.
- Purisch, A., & Sbordone, R. (1997). Forensic neuropsychology: Clinical issues and practice. *The Neuropsychology Handbook*, 2, 309-356.
- Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, 31, 206-230. <https://doi.org/10.1093/arclin/act073>
- Rassin, E., Boskovic, I., & Merckelbach, H. (2018). Posttraumatic Stress Disorder and Diminished Criminal Responsibility as "New Evidence" in Criminal Revision Procedures. *Journal of Forensic Sciences*, 63, 1911-1913. <https://doi.org/10.1111/1556-4029.13789>
- Resnick, P. J., & Knoll, J. (2005). Faking it. How to detect malingered psychosis. *Current Psychiatry*, 4, 13-25.
- Resnick, P. J., West, S. G., & Wooley, C. N. (2018). Malingering Posttraumatic Disorders. In Rogers, R., & Bender, S.D. (Eds.), *Clinical Assessment of Malingering and Deception* (4th Edition), pp. 22-61. The Guilford Press, New York, NY.
- Resnick, P. J. West., S., & Payne, JW (2008). Malingering of post-traumatic disorders. In (Ed.) Rogers, R. (Ed.) (2008). *Clinical Assessment of Malingering and Deception* (pp. 109-127). New York, NY: Guilford Press.
- Richards, A., French, C. C., Johnson, W., Naparstek, J., & Williams, J. (1992). Effects of mood manipulation and anxiety on performance of an emotional Stroop task. *British Journal of Psychology*, 83, 479-491. Doi:10.1111/j.2044-8295.1992.tb02454.x
- Rienstra, A., Groot, P. F., Spaan, P. E., Majoie, C. B., Nederveen, A. J., Walstra, G. J. M., de Jonghe, J. F. M., van Gool, W. A., Olabarriaga, S. D., Korkhov, V. V., & Schmand, B. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*, 35, 59-70. <https://doi.org/10.1080/13803395.2012.751361>
- Rissmiller, D. A., Steer, R. A., Friedman, M., & Demercurio, R. (1999). Prevalence of malingering in suicidal psychiatric inpatients: a replication. *Psychological reports*, 84, 726-730. <https://doi.org/10.2466/pr0.1999.84.3.726>

- Rogers, R. (1990). Models of feigned mental illness. *Professional Psychology: Research and Practice*, 21, 182. <https://doi.org/10.1037/0735-7028.21.3.182>
- Rogers, R. (Ed.) (2008). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: The Guilford Press.
- Rogers, R. (2018). Researching response styles. In R. Rogers & S. D. Bender (eds.), *Clinical assessment of malingering and deception*, 4th ed. (pp. 592-614). New York: Guilford.
- Rogers, R., & Bender, S. D. (2018). *Clinical Assessment of Malingering and Deception* (4th edition). New York, NY: Guilford Press.
- Rogers, R., & Cruise, K. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, 22, 273-285.
- Rogers, R., & Shuman, D. W. (2005). Miranda and beyond: Competencies related to police investigations. In Rogers, R., & Shuman, D. W. (Eds.) (2006). *Fundamentals of forensic practice: Mental health and criminal law* (pp. 113-149). Boston, MA: Springer.
- Rogers, R., Bender, S. D. (2003). Evaluation of malingering and deception. In Goldstein, A. M., Weiner, I. B. (Eds.) (2003). *Handbook of Psychology: Forensic Psychology* (pp. 109-129). New York, NY: Wiley. <https://doi.org/10.1002/0471264385.wei1107>
- Rogers, R., Sewell, K. W., & Salekin, R. T. (1994). A meta-analysis of malingering on the MMPI-2. *Assessment*, 1, 227-237.
- Roivainen, E., Veijola, J., & Miettunen, J. (2016). Careless responses in survey data and the validity of a screening instrument. *Nordic Psychology*, 68, 114-123. <https://doi.org/10.1080/19012276.2015.1071202>
- Rosen, G. M. (2004). Litigation and reported rates of *posttraumatic stress disorder*. *Personality and Individual Differences*, 36, 1291-1294.
- Rosen, G. M., & Phillips, W. R. (2004). A cautionary lesson from simulated patients. *Journal of the American Academy of Psychiatry and the Law Online*, 32, 132-133.
- Rosenfeld, B., Sands, S. A., & Van Gorp, W. G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, 15, 349-359. <https://doi.org/10.1093/arclin/15.4.349>
- Rosenthal, R. (1994). Parametric measures of effect size. In Cooper, H., & Hedges, L. V. (Eds.). *The handbook of research synthesis* (page 231-244). New York: Russell Sage Foundation.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86, 638-641.
- Ruff, R. M., Klopfer, J., & Blank, J. (2016). Homogenous base rates for malingering in neuropsychological examination of litigants. *Psychological Injury and Law*, 9, 143-153.
- Rubenzon, S. (2009). Posttraumatic stress disorder: Assessing response style and malingering. *Psychological Injury and Law*, 2, 114-142.
- Santamaría, P., Ramírez, P. C., & Ordi, H. G. (2013). Prevalencia de simulación en incapacidad temporal: percepción de los profesionales de la salud. *Clínica y Salud*, 24, 139-151.
- Savitsky, K., & Gilovich, T. (2003). The illusion of transparency and the alleviation of speech anxiety. *Journal of Experimental Social Psychology*, 39, 618-625. [https://doi.org/10.1016/S0022-1031\(03\)00056-8](https://doi.org/10.1016/S0022-1031(03)00056-8)
- Schmand, B., Lindeboom, J., Schagen, S., Heijt, R., Koene, T., & Hamburger, H. L. (1998). Cognitive complaints in patients after whiplash injury: the impact of malingering. *Journal of Neurology, Neurosurgery & Psychiatry*, 64, 339-343.
- Schoenberg, M. R., Dorr, D., & Morgan, C. D. (2003). The ability of the Millon Clinical Multiaxial Inventory – third edition to detect malingering. *Psychological Assessment*, 15, 198-204. doi:10.1037/1040-3590.15.2.198
- Schrader, A. D., & Osburn, H. G. (1977). Biodata faking: Effects of induced subtlety and position specificity. *Personnel Psychology*, 30, 395-404.

- Scott, C., & McDermott, B. (2011). Malingering. In Buchanan, A., & Norko, M. A. (Eds.), *The Psychiatric Report: Principles and Practice of Forensic Writing*. New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9780511895074.021>
- 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. <https://doi.org/10.1016/j.cbpra.2008.01.003>
- Simons, J. S., Garrison, J. R., & Johnson, M. K. (2017). Brain mechanisms of reality monitoring. *Trends in cognitive sciences*, 21, 462-473. <https://doi.org/10.1016/j.tics.2017.03.012>
- Sleep, C. E., Petty, J. A., & Wygant, D. B. (2015). Framing the results: Assessment of response bias through select self-report measures in psychological injury evaluations. *Psychological Injury and Law*, 8, 27-39. doi:10.1007/s12207-015-9219-1
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545-561. [https://doi.org/10.1076/1385-4046\(199911\)13:04;1-Y;FT545](https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545)
- Smith, G. P. (2008). Brief screening measures for the detection of feigned psychopathology. In R. Rogers (Ed.), *Clinical Assessment of Malingering and Deception* (3rd ed.) (pp. 323-342). New York: Guilford.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law Online*, 25, 183-189.
- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35, 50-61. <https://doi.org/10.1177/0022022103260380>
- 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.
- Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology*, 26, 774-789. Doi:10.1093/arclin/acr066
- Spielberger, C. D. (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research*. New York: Academic Press.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. D. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Staniloiu, A., & Markowitsch, H. J. (2014). Dissociative amnesia. *The Lancet Psychiatry*, 1, 226-241. [https://doi.org/10.1016/S2215-0366\(14\)70279-2](https://doi.org/10.1016/S2215-0366(14)70279-2)
- Stefanovics, E., He, H., Ofori-Atta, A., Cavalcanti, M. T., Neto, H. R., Makanjuola, V., Ighodaro, A., Leddy, M., & Rosenheck, R. (2016). Cross-national analysis of beliefs and attitude toward mental illness among medical professionals from five countries. *Psychiatric Quarterly*, 87, 63-73. Doi:10.1007/s11126-015-9363-5
- Steller M. (1989) Recent Developments in Statement Analysis. In: Yuille J.C. (Eds) (1989) *Credibility Assessment*. Nato Science (Series D: Behavioural and Social Sciences), vol 47. Springer, Dordrecht. https://doi.org/10.1007/978-94-015-7856-1_8
- Steller, M., & Koehnken, G. (1989). Criteria-based statement analysis. In D. C. Raskin (Ed.), *Psychological methods in criminal investigation and evidence* (pp. 217-245). New York, NY, US: Springer Publishing Co.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662
- Sullivan, B. K., May, K., & Galbally, L. (2007). Symptom exaggeration by college adults in attention-deficit hyperactivity disorder and learning disorder assessments. *Applied Neuropsychology*, 14, 189-207.
- Sullivan, K., Lange, R.T., & Dawes, S. (2005). Methods of detecting malingering and estimated symptoms exaggeration base rates in Australia. *Journal of Forensic Neuropsychology*, 4, 44-70.

- Szogi, E. G., & Sullivan, K. A. (2018). Malingered Posttraumatic Stress Disorder (PTSD) and the Effect of Direct Versus Indirect Trauma Exposure on Symptom Profiles and Detectability. *Psychological Injury and Law*, 1-11.
- Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, 16, 495-505. <https://doi.org/10.1076/clin.16.4.495.13909>
- Taylor, S., Frueh, B. C., & Asmundson, G. J. (2007). Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: Methods, obstacles, and recommendations. *Journal of Anxiety Disorders*, 21, 22-41. <https://doi.org/10.1016/j.janxdis.2006.03.016>
- Terheggen, M. A., Stroebe, M. S., & Kleber, R. J. (2001). Western conceptualizations and eastern experience: A cross-cultural study of traumatic stress reactions among Tibetan refugees in India. *Journal of Traumatic Stress*, 14, 391-403.
- Thakker, J., & Ward, T. (1998). Culture and classification: The cross-cultural application of the DSM-IV. *Clinical Psychology Review*, 18, 501-529.
- Thomas, A. (2001). Factitious and malingered dissociative identity disorder: Clinical features observed in 18 cases. *Journal of Trauma & Dissociation*, 2, 59-77. https://doi.org/10.1300/J229v02n04_04
- Thomas, T. A., & Fremouw, W. J. (2009). The Stroop interference effect and spontaneous recall in true and malingered motor-vehicle accident-related PTSD. *The Journal of Forensic Psychiatry & Psychology*, 20, 936-949. <https://doi.org/10.1080/14789940903174162>
- Thomson, P., Keehn, E. B., & Gumpel, T. P. (2009). Generators and interpreters in a performing arts population: Dissociation, trauma, fantasy proneness, and affective states. *Creativity Research Journal*, 21, 72-91.
- Tracy, D. K., & Rix, K. J. (2017). Malingering mental disorders: clinical assessment. *BJPsych Advances*, 23, 27-35. <https://doi.org/10.1192/apt.bp.116.015958>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.
- van Dam, D., Ehring, T., Vedel, E., & Emmelkamp, P. M. (2013). Screening for posttraumatic stress disorder in civilian substance use disorder patients: cross-validation of the Jellinek-PTSD screening questionnaire. *Journal of Substance Abuse Treatment*, 44, 126-131.
- Van der Heide, D. H., & Merckelbach, H. (2017). Validity of symptom reports of asylum seekers in a psychiatric hospital: A descriptive study. *Journal of the Academy of Psychiatry and the Law*. <http://dx.doi.org/10.1016/j.ijlp.20.2016.05.007>
- van der Heide, D., Boskovic, I., & Merckelbach, H. (2017). Standard Symptom Inventories for Asylum Seekers in a Psychiatric Hospital: Limited Utility Due to Poor Symptom Validity. *Psychological Injury and Law*, 10, 358-367.
- Van Egmond, J., & Kummeling, I. (2002). A blind spot for secondary gain affecting therapy outcomes. *European Psychiatry*, 17, 46-54. [https://doi.org/10.1016/S0924-9338\(02\)00622-3](https://doi.org/10.1016/S0924-9338(02)00622-3)
- Van Herk, H., Poortinga, Y. H., & Verhallen, T. M. (2004). Response styles in rating scales: Evidence of method bias in data from six EU countries. *Journal of Cross-Cultural Psychology*, 35, 346-360. <https://doi.org/10.1177/0022022104264126>
- van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, 28, 1336-1365. <https://doi.org/10.1080/13854046.2014.984763>
- van Impelen, A., Merckelbach, H., Jelicic, M., Niesten, I. J., & à Campo, J. (2017). Differentiating Factitious from Malingered Symptomatology: the Development of a Psychometric Approach. *Psychological Injury and Law*, 1-17. <https://doi.org/10.1007/s12207-017-9301y>

- van Impelen, A., Merckelbach, H., Niesten, I. J., Jelicic, M., Huhnt, B., & Campo, J. A. (2017). Biased Symptom Reporting and Antisocial Behaviour in Forensic Samples: A Weak Link. *Psychiatry, Psychology and Law*, 24, 530-548. <https://doi.org/10.1080/13218719.2016.1256017>
- van Wijk, C. M. G., Huisman, H., & Kolk, A. M. (1999). Gender differences in physical symptoms and illness behavior. *Social Science & Medicine*, 49, 1061–1074. doi:10.1016/S0277-9536(99)00196-3
- Verbrugge, L. M., & Ascione, F. J. (1987). Exploring the iceberg: Common symptoms and how people care for them. *Medical Care*, 25, 539-569.
- Vickery, C. D., Berry, D. T. R., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology*, 19, 37–48.
- Vilar-López, R., Santiago-Ramajo, S., Gómez-Río, M., Verdejo-García, A., Llamas, J. M., & Pérez-García, M. (2007). Detection of malingering in a Spanish population using three specific malingering tests. *Archives of Clinical Neuropsychology*, 22, 379-388. Doi:10.1016/j.acn.2007.01.012
- Vrij, A. (2000). *Detecting lies and deceit: the psychology of lying and implications for professional practice*. (Wiley series in psychology of crime, policing and law). Chichester: Wiley.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons; Chichester, UK.
- Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection: A plea to change police practice. *Criminal Justice and Behavior*, 35, 1323-1336. <https://doi.org/10.1177/0093854808321530>
- Vrij, A., Fisher, R., Mann, S., & Leal, S. (2008). A cognitive load approach to lie detection. *Journal of Investigative Psychology and Offender Profiling*, 5, 39-43.
- Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Lying about flying: The first experiment to detect false intent. *Psychology, Crime & Law*, 17, 611-620. <https://doi.org/10.1080/10683160903418213>
- 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. <https://doi.org/10.1007/s10979-007-9103-y>
- Vrij, A., Mann, S., Leal, S., and Granhag, P. A. (2010c). Getting into the minds of pairs of liars and truth-tellers: an examination of their strategies. *Frontiers in Psychology*, 3, 17–22. doi: 10.2174/1874917801003010017
- Vrij, A., Nahari, G., Isitt, R., & Leal, S. (2016). Using the verifiability lie detection approach in an insurance claim setting. *Journal of Investigative Psychology and Offender Profiling*, 13, 183-197. <https://doi.org/10.1002/jip.1458>
- Waller, N. G., Putnam, F. W., & Carlson, E. B. (1996) Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. *Psychological Methods*, 1, 300-321.
- Waller, N., Putnam, F. W., & Carlson, E. B. (1996). Types of dissociation and dissociative types: A taxometric analysis of dissociative experiences. *Psychological methods*, 1, 300.
- Walters, G. D., Rogers, R., Berry, D. T., Miller, H. A., Duncan, S. A., McCusker, P. J., ... & Granacher Jr, R. P. (2008). Malingering as a categorical or dimensional construct: The latent structure of feigned psychopathology as measured by the SIRS and MMPI-2. *Psychological Assessment*, 20, 238. <https://doi.org/10.1037/1040-3590.20.3.238>
- 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* (2nd ed., p. 168-189). New York: Guilford Press.
- Weiss, K. J., & Van, L. D. (2017). Liability for Diagnosing Malingering. *The Journal of the American Academy of Psychiatry and the Law*, 45, 339-347.
- Weiss, R. A., & Rosenfeld, B. (2012). Navigating cross-cultural issues in forensic assessment: Recommendations for practice. *Professional Psychology: Research and Practice*, 43, 234. <https://doi.org/10.1037/a0025850>

- Weiss, R. A., & Rosenfeld, B. (2012). Navigating cross-cultural issues in forensic assessment: Recommendations for practice. *Professional Psychology: Research and Practice*, 43, 234-240.
- Widows, M. R., & Smith, G. P. (2005). *Structured Inventory of Malingered Symptomatology professional manual*. Odessa: Psychological Assessment Resources.
- Wierzbicki, M., & Tyson, C. M. (2007). A summary of evaluations for learning and attention problems at a university training clinic. *Journal of Postsecondary Education and Disability*, 20, 16-27.
- Williams, J. M. (1997). G., Watts, FN, MacLeod, C., Mathews. *Cognitive Psychology and Emotional Disorders* (2nd ed.). Chichester: John Wiley.
- Wilson, S. C., & Barber, T. X. (1983). The fantasy-prone personality: Implications for understanding imagery, hypnosis, and parapsychological phenomena. In Sheikh, A. A. (Ed.), *Imagery: Current Theory, Research, and Application* (p. 340-387). New York, New York: Wiley.
- Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., Olesen, J., Allgulander, C., Alonso, J., Faravelli, C., Fratiglioni, L., Jennum, P., Lieb, R., Maercker, A., van Os, J., Preisig, M., Salvador-Carulla, L., Simon, R., Stenhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21, 655-679.
- Worley, C. B., Feldman, M. D., & Hamilton, J. C. (2015). The case of factitious disorder versus malingering. *Psychiatric Times*, 26-28.
- Yeo, L., Goh, V., Liem, G. (2016). School-based intervention for test anxiety. *Child & Youth Care Forum*, 45, 1-17.
- Young, G. (2014). *Malingering, feigning, and response bias in psychiatric/psychological injury: implications for practice and court*. New York: Springer.
- Young, G. (2015). Malingering in forensic disability-related assessments: Prevalence 15±15%. *Psychological Injury and Law*, 8, 188-199. <https://doi.org/10.1007/s12207-015-9232-4>
- Young, G. (2016). PTSD in court I: Introducing PTSD for court. *International Journal of Law and Psychiatry*, 49, 238-258. <https://doi.org/10.1016/j.ijlp.2016.10.012>
- Ziegler, M. (2015). "F*** You, I Won't Do What You Told Me!"—Response Biases as Threats to Psychological Assessment. *European Journal of Psychological Assessment*, 31, 153-158. <https://doi.org/10.1027/1015-5759/a000292>
- Zimmerman, M. (2003). What should the standard of care for psychiatric diagnostic evaluations be?. *The Journal of Nervous and Mental Disease*, 191, 281-286. <https://doi.org/10.1097/01.NMD.0000066149.40946.FA>

SUMMARY

SUMMARY

Malingering is defined as the intentional false presentation of symptoms driven by an external incentive, which can be financial (e.g., compensation), legal (e.g., diminished criminal responsibility), or involve other types of personal gain (e.g., to obtain medication). Malingering occurs on a non-trivial scale in both civil and criminal contexts, and has serious legal and social ramifications. The aim of this thesis was to examine whether a multi-method assessment of symptoms might enhance the detection of symptom fabrication, and whether an assessor's cultural background impacted their perceptions of malingering.

In this thesis, three different methods for detecting malingering were tested across a total of eight studies. Specifically, we tested a novel lie-detection tool, the Verifiability Approach (VA) in order to investigate whether the VA could contribute to the credibility assessment of physical symptom reports. The logic behind the VA is that people who are telling the truth should produce more verifiable details - information that can, in principle, be checked – compared with liars. Liars, in contrast, tend to avoid verifiable information and report non-verifiable details. Next, we critically examined the utility of a controversial task, the Modified Stroop task (MST), a reaction time measure of attentional bias among patients, in the detection of malingering. The MST involves the presentation of disorder-related and neutral words that are shown in different colors. The task is to color-name the words while disregarding their semantic meaning. The idea behind the MST is that genuine patients should show prolonged reaction times (RT) when presented with disorder-related, compared with neutral words (i.e., the MST effect). We also applied a newly developed measure of over-reporting, the Self-Report Symptom Inventory (SRSI), in order to investigate symptom endorsement among experimental malingerers. The SRSI includes two scales of symptoms: genuine (plausible) symptoms and pseudosymptoms (unlikely complaints). The rationale behind the SRSI is that genuine patients will endorse more of genuine and fewer of the pseudosymptoms, whereas malingerers will overendorse both types of symptoms. Finally, we investigated whether the cross-cultural background of practitioners influences their view on exaggerated symptoms.

In the first experimental chapter (Chapter 2), we report the findings of research designed to test the VA with respect to the detection of fabricated symptom statements. We investigated the extent to which people with genuine symptoms compared with instructed malingerers differed in the provision of checkable details. In Study 1, we examined statements of students genuinely suffering from various physical symptoms (e.g., headache, backpain), and students instructed to malingering such experiences. We found that malingerers, relative to truth tellers, produced longer statements that contained fewer verifiable details. In Study 2, we repeated the task, but participants were informed that their statements would be inspected for verifiable information. Providing this additional

information to participants led to non-significant differences between instructed malingerers and truth tellers in terms of both verifiable and non-verifiable information, and the overall length of statements. In Study 3, we experimentally induced symptoms (physical exercise) in one group of students (truth tellers), whereas two other groups received instructions to mangle having been engaged in physical exercise. Participants were not informed about the type of information they should provide. The results confirmed our findings from Study 1. We observed longer statements containing more non-verifiable information for both malinger groups relative to the control group. In other words, an extensive amount of non-verifiable details was indicative of fabricated symptom reports in both Study 1 and Study 3.

In Chapter 3, we tested whether a reaction time task, the MST, used in combination with the SRSI, might assist in the detection of malingered anxiety-related symptoms. In Study 4, we focused on test-anxiety relying on a within-subject design. Students who were not suffering from test anxiety were asked to first genuinely respond to the MST task and the SRSI, and seven days later they were instructed to mangle having test-anxiety and to repeat the tasks. We found that students in the latter session produced the MST effect typically found in genuine test-anxiety patients, while the MST effect did not emerge in the first session. Participants in the second session overendorsed genuine symptoms and pseudosymptoms related to anxiety, compared with the first session, which led to a detection rate of 77% of test anxiety malingerers. In Study 5, we investigated PTSD-related symptoms, and included three groups of participants: 1) participants with current high impact aversive experiences, 2) participants with low impact aversive experiences, and 3) actors, who also had a low impact history but were asked to simulate being under the effects of a high impact of aversive experience. The MST effect did not emerge in any of the groups. However, the actors produced longer response latencies than both high and low impact groups. Actors also overendorsed items of the SRSI, thus 89% of these malingerers were successfully detected as such. Problematically, however, 27% of the honest group were also classified as malingerers.

In Chapter 4, we focused on the SRSI alone, and its utility for detecting malingered physical (pain-related) and psychological (anxiety-related) symptoms (Study 6). In a between subjects design we included an honest comparison group and two groups of instructed malingerers (pain and anxiety symptoms). The malingerers of pain and malingerers of anxiety endorsed more genuine symptoms and pseudosymptoms than participants in the honest group. Also, both malingerer groups overendorsed symptoms corresponding to their alleged conditions. The detection rates reached 48% for simulators of pain, and 73% for simulators of anxiety, suggesting that the SRSI has a low sensitivity to simulated physical complaints.

In Chapter 5 (Study 7), using a combination of the VA and SRSI, we wanted to examine malingerers' strategies in fabricating exposure narratives and symptom reports in a PTSD-related condition. In a between subjects design, we experimentally induced the PTSD-like symptoms in honest comparison group, using a Virtual Reality (VR) paradigm. The other group was not exposed to the VR, but was instructed to simulate that experience (malingerers). We applied the VA to examine the veracity of their exposure narratives, and our findings were consistent with the general results reported in Chapter 2. Malingerers produced longer statements containing more non-verifiable details, whereas the honest group included a higher proportion of verifiable information. The quality of the symptom reports was investigated using the subscales of the SRSI describing genuine and pseudosymptoms pertaining to anxiety and PTSD complaints. Malingerers endorsed more both genuine symptoms and pseudosymptoms than the honest group. Applied together, the features of the VA and the SRSI subscales correctly classified 76% of participants, which supports a multi-method approach within the symptom validity assessment.

Finally, in Chapter 6 (Study 8), we investigated whether the cultural background of practitioners affected their plausibility judgments concerning various symptoms. We included Western and non-Western practitioners, and presented them with a mix of atypical symptoms, dissociative symptoms, and every-day complaints. Their task was to rate the plausibility of each symptom in terms of exaggerated or authentic. There were no significant differences between culturally diverse practitioners in how they judged the plausibility of various symptoms. All practitioners rated atypical and dissociative items as significantly less authentic than every-day complaints, but they did not distinguish between atypical symptoms and dissociative symptoms.

In sum, our results indicated that, with certain adjustments, the VA might contribute to symptom validity assessment. The MST, however, was shown to be an unreliable detection tool, the use of which should be avoided to detect symptom fabrication. The SRSI appears to be a promising method for detecting symptom over-reporting. Furthermore, combining the VA and the SRSI led to adequate detection rates of instructed malingerers, indicating the benefits of combining different methods in symptom validity assessment. However, further research is necessary to establish the generalizability of our findings to different samples, such as patients.

ACKNOWLEDGMENTS

The last three years have been everything one researcher could wish for. Working in two foreign countries and to be surrounded with by far smarter and more creative people than myself, I can only explain as a dream job and ego's worst nightmare. For giving me the ticket for this roller-coaster of excitement, fun, achievements, and rejections, I first thank the House of Legal Psychology program and its founders. Also, my deepest gratitude for everything I have learned and became as an academic goes to the people who have supervised me over the years of my education. To my professors Zdenka Novovic, Ljiljana Mihic, and Miklos Biro, who spotted and tolerated my curiosity and temper since the beginning of my studies. Thank you for teaching me the first steps of research. Also, my thanks go to Vladan, Brana, and Tamara and the War Trauma Center, who gave me room to practice those steps. Marko Jelicic, thank you for all the support, laughter, and for trying to learn our language – Veliko hvala! Lorraine Hope, you have been always just one email away, and you have busted the myth how the busiest researchers cannot be incredibly cool. Thank you. James Ost, you have my deep respect for keeping the rock 'n roll vibe in academia. To you, Harald, I owe so much more than words can express. I know that your ego doesn't need any further boosting, but your enthusiasm is something that cannot be found even among us, beginners. The “sprezzatura” with which you give amazing insights and guidance can only be known to those who had the luck to meet you. You have been more than anything I could wish for in a supervisor. Dank U wel.

My Dutch/Belgian pack, Bart and Karen, I was so lucky to have found you, and I am thankful for that every day. Love you! To, Sasa, my dearest friend, my rock, I cannot thank you enough for everything. Neopisivo ti hvala, bratic. Valentina, my Sunshine, I have missed you every day, but I knew that you were always there for me - Volim te, Sunce. Prpa, that message from you got me here – hvala! To Nina, you proved that I could actually love an American, but I am sure I adore your Swedish side. Kathy, my Latina, thank you for the Baileys nights and all the talks, Veronica needed it. David, my buddy, the Portsmouth experience bonded us for life. To Robin, thanks for the patience, friendship, and all the help. To the rest of the crew and all of my roommates, Shiri (my Mocha), Alex, Nikky, Kate, Pimpa, Eliza, Elly, Conny, Ana, Glynis, Bri, Sergii, Alfons, thanks for the support.

Mojoj porodici, mami i Maki, hvala na podrsci i razumevanju mog puta, iako me je odveo daleko od vas. Bokaco, tebi dugujem toliko toga, zauvek ti hvala.

My *principessa*, my amore, my buddy, my biggest support, Ivan, I do not know how I would do this without you. Ti ringrazio infinitamente per ogni singolo giorno speso assieme e per quanto abbiamo costruito fin' ora. Sono davvero orgogliosa di te, Amore mio. Ti amo indescrivibilmente, grazie di esistere.

To Zizu, my Sun, Moon, and Stars, who made writing this thesis a hundred times more challenging, but who also made me a thousand times better person. Also, he is probably the only one who won't pretend to have read this thesis – Avav you.

Finally, thanks also go to my younger self who believed that she could get here one day. I doubted you so many times, but I guess I was wrong. Here you go.

DISSEMINATION

Study 1 and Study 2 reported in Chapter 2 are published as one paper in *Psychology, Crime, and Law*, while Study 3 is published separately in *Psychiatry, Psychology, and Law*. The findings of these studies were also presented in a form of a poster during the 5th Symptom Validity Conference in Basel, Switzerland in July 2017, and as oral presentation during the *European Association for Psychology and Law conference in July 2016*. Study 4 in Chapter 3 is published in *Frontiers in Psychology*, while the Study 5 paper is currently under revision for publication at the *Journal of Behavior Therapy and Experimental Psychiatry*. The findings of the Study 4 were presented during the *European Association for Psychology and Law conference in May 2017*. Chapter 4 covers one study about SRSI, and that manuscript is currently in press at the *European Journal of Psychological Assessment*. The study reported in Chapter 5 is in press at the *Legal and Criminological Psychology*. These findings were also presented at the *European Association for Psychology and Law conference* in June 2018. Chapter 6 has already been published in *Psychological Injury and Law* (see below), and presented as a poster during the 5th Symptom Validity Conference in Basel, Switzerland in July 2017.

Publications

Boskovic, I. (2018). Your sample might be malingering: Students and motivation to malingering.

Submitted manuscript.

Boskovic, I., & Merckelbach, H. (2018). Fake Posttraumatic Stress Disorder (PTSD) costs real money. *The Inquisitive Mind Magazine*, 4/2018.

Boskovic, I., Biermans, A., Merten, T., Jelacic, 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*. In press.

Boskovic, I., Bogaard, G., Merckelbach, H., Vrij, A., & Hope, L. (2017). The Verifiability Approach to detection of malingered physical symptoms. *Psychology, Crime & Law*, 23, 717-729.

Boskovic, I., Dibbets, P., Hope, L., & Jelacic, M. (in press). Verify the scene, report the symptoms: Testing the Verifiability Approach and the SRSI in detection of fabricated PTSD. *Legal and Criminological Psychology*.

Boskovic, I., Hope, L., Ost, J., Orthey, R., & Merckelbach, H. (2018). Detecting feigned high impact experiences: A symptom over-report questionnaire outperforms the emotional Stroop task. Accepted manuscript pending revision.

- Boskovic, I., Merckelbach, H., Merten, T., Hope, L., & Jellicic, M.** (in press). The Self-Report Symptom Inventory as a screen for over-reporting: An explorative study with instructed simulants. *European Journal of Psychological Assessment*.
- Boskovic, I., Tejada Gallardo, C., Vrij, A., Hope, L., & Merckelbach, H.** (2018). Verifiability on the Run: An Experimental Study on the Verifiability Approach to Malingered Symptoms. *Psychiatry, Psychology, and Law*. In press.
- Boskovic, I., van der Heide, D., Hope, L., Merckelbach, H., & Jellicic, M.** (2017). Plausibility judgments of atypical symptoms across cultures: An explorative study among Western and non-Western experts. *Psychological Injury and Law*, 10, 274–281.
- Jellicic, M., Merckelbach, H., & **Bošković, I.** (2017). Seven myths about feigning. In H. Otgaar & M.L. Howe (Eds.). *Finding the truth in the courtroom: Handling deception, lies, and memories*. Oxford University Press.
- Merckelbach, H., Dalsklev, M., van Helvoort, D., **Boskovic, I.**, & Otgaar, H. (2018). Symptom self-reports are susceptible to misinformation. *Psychology of Consciousness: Theory, Research, and Practice*. Doi 10.1037/cns0000159
- Merckelbach, H., **Boskovic, I.**, Pesy, D., Dalsklev, M., & Lynn, S. (2017). Symptoms overreporting and dissociative experiences: A qualitative review. *Consciousness and Cognition*, 49, 132-144.
- Merckelbach, H., Prins, C., **Boskovic, I.**, Niesten, I. & à Campo, J. (2018). Alexithymia as a potential source of symptom over-reporting: An exploratory study in forensic patients and non-forensic participants. *Scandinavian Journal of Psychology*. Doi 10.1111/sjop.12427
- Rassin, E., **Boskovic, I.**, & Merckelbach, H. (2018). PTSD and diminished criminal responsibility as “new evidence” in criminal revision procedures. *Journal of Forensic Science*, 63, 1911-1913. <https://doi.org/10.1111/1556-4029.13789>
- van der Heide, D., **Boskovic, I.**, & Merckelbach, H. (2017). Standard Symptom Inventories for Asylum Seekers in a Psychiatric Hospital: Limited Utility Due to Poor Symptom Validity. *Psychological Injury and Law*, 10, 358-367.

Manuscripts in preparation

- Boskovic, I., Zwaan, L., & Merckelbach, H.** (2018). Stability of common symptoms among truth tellers and feigners. Manuscript in preparation.
- Boskovic, I., Reymarks, A., & Merckelbach, H.** (2018). Telling a good story 2.0.: Fantasy proneness and the Verifiability Approach. Manuscript in preparation.

Ongoing research

- Boskovic, I., Peters, M., Merckelbach, H.** (2018). Exposure narratives and symptom reports among patients and feigners. Ongoing research.

Conference Oral presentations

Boskovic, I. (2018, October). *When honesty is not the best policy: Symptom validity testing and quality of narratives in fabricated symptoms*. Plenary speaker, European Union of Medicine in Assurance and Social Security (EUMAAS), Maastricht, the Netherlands.

Boskovic, I., (2018, June). *Verify the scene, report the symptoms: Testing the verifiability Approach and the SRSI in false PTSD claims*. Presentation, EAPL, Turku, Finland.

Boskovic, I., (2017, July). *The Verifiability Approach to malingered physical symptoms*. Presentation, 5th Symptom Validity Conference, Basel, Switzerland.

Boskovic, I., (2017, May). *Stroop performance and symptom endorsement in feigning test anxiety*. Presentation, EAPL, Mechelen, Belgium.

Conference Poster presentation

Boskovic, I., (2017, July). *Odd stays odd: Cross-cultural stability of atypical symptoms*. Poster presentation, 5th Symptom Validity Conference, Basel, Switzerland.

Boskovic, I., (2016, July). *The Verifiability Approach to Malingering Physical Symptoms: Just Let Them Talk*. Poster presentation at European Association of Psychology and Law (EAPL), Toulouse, France.

CURRICULUM VITAE

Irena Boskovic was born on the 2nd of May 1989, in Ivanjica, Serbia. She finished the Bachelor psychology studies at Faculty of Philosophy, University of Novi Sad, after which she continued with the Masters in Clinical psychology at the same institution. During and after her studies, Irena was a member of the clinical psychology research group, and a volunteer in the War Trauma Center in Novi Sad. She obtained basic training in Gestalt and Cognitive Behavioral therapy. In 2015, she was accepted to the House of Legal Psychology, the Erasmus Mundus Joint Doctorate program, at both Maastricht University and the University of Portsmouth. During her doctoral studies, Irena was mainly interested in investigating ways to detect those who fabricate symptoms in order to receive benefits (i.e., malingering). After three years, Irena successfully defended her doctorate thesis at the University of Portsmouth in September 2018. Until now, she has published more than 10 scientific articles, outcomes of her Ph.D. studies as well as collaborative projects, and a book chapter. Irena is also a columnist for university paper, *Observant*. Yet, Irena's most important accomplishment is being the owner of a black Labrador named *Zizu*, who often malingers, as depicted on the cover of this thesis.

VALORIZATION ADDENDUM

*It's hard to explain to people what the significance of an invention is, so it's hard to get funding.
The first thing they say is that it can't be done. Then they say, "You didn't do it right."
Then, when you've done it, they finally say, "Well, it was obvious anyway."*

Robert S. Ledley

Today, research is appreciated as much as it is funded. Money became the way we valorize everything that surrounds us. Thus, it is not surprising that this applies to research as well. However, the whole doctrine comes with the risk of fixating the field of forensic psychology on one to two popular topics that are being funded for a time being, until a new focus theme appears. Thinking outside of the “funding box” may be more challenging, but also more important than ever.

The focus of this thesis - malingering lays between the clinical and forensic/legal psychology, bridging the existing gap. Although it is reasonable to expect that such position of malingering research creates more funding opportunities, the truth is that it does not. Even finding the appropriate publication outlets seems like a challenge. While clinical journals do not appreciate the idea of fabricated symptoms investigation, the legal/forensic journals also find their reasons to see this topic as less important than, for instance, memory research. Therefore, if funding is the currency of valorization, this chapter will be a brief one.

If, however, the practical utility and the impact on practitioners in many different fields are considered, the value of malingering research is multidimensional, bearing relevance to medical, psychological, and even law professionals. The financial, ethical, and social consequences of malingering are profound, and affect all of the previously listed lines of work. Thus, research that provides tools and insights that help practitioners to prevent spending resources on those who do not actually need them is of critical importance in a society in which financial health care means are limited. Furthermore, malingering research, despite the widespread ill found skepticism of some clinicians (e.g., Jackson et al., 2011), aims to protect and secure genuine patients.

This thesis, in particular, goes a step further than mainstream investigations in the field, in that it touches upon the main cues to detection of malingering using a multi-method approach. The research evaluated a verbal credibility method that was never applied before in this line of research, whereas the reaction time measures were often propagated based on misleading findings. The dominant approach in symptom validity assessment, symptom validity tests, was also enriched by a new measure, for which this thesis provides a solid foundation for further investigation.

The results of this work were presented at relevant legal/forensic psychology conferences and academic organizations, as well as at conferences for medical practitioners and insurance companies experts. Not surprisingly, presenting this line of research always received more attention and appreciation from practitioners than from fellow academics. This leads me to the conclusion: People

who fund the research in Europe are usually not the end users of it, because if they were, the topic of malingering would finally be as supported as it is needed.

APPENDICES

APPENDIX 1

CHAPTER 2

STUDY 1 & 2

Supplemental table 2.1

Inter-rater agreement for four groups of details in Study 1 and 2.

| Group of details | Intraclass correlation coefficient | |
|------------------|------------------------------------|--------|
| | Study 1 | Study2 |
| Perceptual | .95 | .81 |
| Spatial | .63 | .84 |
| Temporal | .96 | .83 |
| Descriptive | .84 | .94 |

Supplemental table 2.2.

Number (%) of participants in both groups providing different types of details.

| Type of details | Study 1 | | Study 2 | |
|---------------------------|----------------------------|------------------------|----------------------------|------------------------|
| | Truth tellers <i>n</i> (%) | Malingers <i>n</i> (%) | Truth tellers <i>n</i> (%) | Malingers <i>n</i> (%) |
| <i>Verifiable details</i> | | | | |
| - Perceptual | 0 (0%)* | 2 (2.4%)* | 1 (2.7%)* | 7 (10.4%)* |
| - Spatial | 1 (2.4%)* | 1 (1.2%)* | 10 (26.3%) | 8 (11.9%) |
| - Temporal | 2 (4.9%)* | 1 (1.2%)* | 11 (28.9%) | 20 (29.8%) |
| - Descriptive | 10 (24.4%) | 10 (11.9%) | 18 (48.6%) | 41 (61.2%) |
| <i>Non-verifiable</i> | | | | |
| - Perceptual | 32 (78%) | 79 (94%) | 31 (83.8%) | 65 (97.0%) |
| - Spatial | 22 (53.7%) | 35 (41.7%) | 30 (81.1%) | 38 (56.7%) |
| - Temporal | 35 (85.4%) | 76 (90.5%) | 33 (89.2%) | 59 (88.1%) |
| - Descriptive | 39 (95.1%) | 84 (100%) | 37 (100%) | 66 (98.5%) |
| Total | 41 (100%) | 84 (100%) | 37 (100%) | 67 (100%) |

Notes: * Excluded from further *t*-test analysis because of the small sample size

Results

Exploring the potential differences between groups in Study 1 for separate categories of details, we found that truth tellers reported significantly less perceptual non-verifiable details than malingers, $t(123) = 4.30, p = .001, d = .86$. In Study 2, truth tellers produced significantly more spatial verifiable, $t(103) = 2.75, p = .01, d = .48^{25}$, and spatial non-

²⁵ Because of the low number of participants in both groups, we also calculated a Mann-Whitney U test (U test = 1052.00, $z = 2.04, p = .02$), which also yielded a significant result.

Appendix

verifiable details than malingerers, $t(103) = 2.67, p = .01, d = .51$. Thus, truth tellers generated more checkable and non-checkable information about locations or spatial arrangement of people and objects than malingerers. Group differences with regard to the other details categories did not reach significance.

Supplemental table 2.3.

Group differences in number of perceptual, spatial, temporal, and descriptive details (verifiable and non-verifiable) in Study 1 and 2.

| | | Study 1 | | | Study 2 | | |
|-------------------------------|---------------|--------------|------------|----------|---------------|------------|----------|
| Details | Group | M (SD) | <i>t/U</i> | <i>p</i> | M (SD) | <i>t/U</i> | <i>p</i> |
| <i>Verifiable details</i> | | | | | | | |
| Perceptual | Truth tellers | .00(.00) | .86 | .39 | .02 (.16) | 1172.0 | .07 |
| | Malingers | .10(.79) | | | .15 (.47) | | |
| Spatial | Truth tellers | .02(.15) | .52 | .60 | .87(2.07) | 2.75* | .01 |
| | Malingers | .01(.11) | | | .15(.44) | | |
| Temporal | Truth tellers | .07(.34) | 1.48 | .14 | 1.10 (3.36) | 1.13 | .26 |
| | Malingers | .01(.10) | | | .60 (1.13) | | |
| Descriptive | Truth tellers | .83(1.96) | 1.91 | .03 | 6.26 (10.52) | .60 | .55 |
| | Malingers | .32(1.01) | | | 5.58 (7.64) | | |
| <i>Non-verifiable details</i> | | | | | | | |
| Perceptual | Truth tellers | 4.22(3.63) | 4.30** | .00 | 7.00 (6.23) | 1.37 | .17 |
| | Malingers | 8.09(5.18) | | | 8.73 (6.17) | | |
| Spatial | Truth tellers | .78(.88) | .70 | .48 | 2.79(3.04) | 2.67** | .00 |
| | Malingers | 1.15(3.35) | | | 1.52(1.82) | | |
| Temporal | Truth tellers | 3.17(3.12) | 1.48 | .14 | 10.34 (9.32) | 1.36 | .18 |
| | Malingers | 4.20(3.87) | | | 8.14 (7.06) | | |
| Descriptive | Truth tellers | 10.66(6.14) | 1.95 | .05 | 28.79 (17.81) | .38 | .70 |
| | Malingers | 14.85(13.01) | | | 26.63 (17.62) | | |

Notes: Alpha values corrected $\alpha = .02$; * $p < .02$; ** $p < .00$;

Summary of findings on separate categories of details

In study 1, we found that malingerers' deception strategy was to generate significantly longer statements than truth tellers, and there were indications that truth tellers and malingerers differed with regard to certain categories of details. Malingerers' reports were filled with descriptions of subjective perceptual experiences (e.g., "I felt very sharp pain"; "In the front part of my head"; "I was thinking this will never pass"). This runs counter to the widespread assumption that statements that are rich in details are more likely to be truthful (Johnson, 2006). One could argue that malingerers are aware of this misbelief and exploit it.

In the study 2, despite non-significant differences between groups in length and number of non-verifiable and verifiable, differences in separate categories of details were still present. Specifically, verifiable spatial details, such as "My mom entered my room", or "I drove to the city center with my sister", were more prevalent in the statements of truth tellers than in those of malingerers. The same was true for non-verifiable spatial details: Relative to malingerers, truth tellers reported more information about the location and spatial arrangements of the objects, even when these details could not be verified ("I stayed in my bed", "I sat in the chair", "Went downstairs to the kitchen" etc.). Our finding is reminiscent of the work of Leins and colleagues (2011), who showed that truth tellers retrieved more spatial information with higher consistency. Liars, on the other hand, tried to avoid mentioning spatial information because imagined event often lack spatial details (Leins, Fisher, Vrij, Leal, & Mann, 2011).

STUDY 3



Picture 2.1.

Stairs at the Faculty of Psychology and Neuroscience at Maastricht University

Instructions for participants

Instruction truth tellers

Welcome to this experiment about perception of internal sensations. During this brief experiment, you will be given two tasks. For the first task, you will have to go down to the ground floor and back to the third floor, twice, walking as fast as you can (running). After the task, you will be given time to rest. We would like you to go to the library, restaurant, have a walk, doing whatever you want. Please return after 30 min. You will be then given paper and pen in order for you to write statements about your sensations and experience during the experiment. You may write not only about what you felt, thought and saw during the exercise, but also about the immediate period thereafter and what you did during the break. It is important that you explain your sensations and your experience with as many details as possible. If you succeed in convincing researchers that you suffered from a high level of physical distress because of this experiment you will be rewarded a bonus credit or an additional €7.5 voucher. All in all, the experiment will last for about 1 hour. You are able to stop during the experiment at any time. If you have any question, please let us know.

Instruction naïve malingerers

Welcome to this experiment about perception of internal sensations. During this brief experiment, you will be given two tasks. First task is to imagine that you just had an exercise and that you run the stairs from the third floor to the ground floor and back, twice, after which you had a 30 minutes break, in which you could go to the library, restaurant, have a walk, whatever you wanted. Imagine all the sensations you felt and action you took after the exercise. It is crucial to try to imagine that experience with as many details as possible. The second task is to write everything about the exercise (what you felt, thought and saw), as well about the period after the exercise and what did you do and where did you go. Try to convince us that you really had that experience 30 minutes ago. So, try to write a detailed statement about your exercise (running up and down the stairs), about sensations you feel as a consequence of that, and the period after the exercise. It is important that you explain your sensations and your experience with as many details as possible. If you succeed in convincing researchers that you suffered from a high level of physical distress because of this experiment you will be rewarded a bonus credit or an additional €7.5 voucher. All in all, the experiment will last for about 1 hour. You are able to stop during the experiment at any time. If you have any question, please let us know.

Instruction coached malingerers

Welcome to this experiment about perception of internal sensations. You are the second group in this study. The first group had to do physical exercise before this part of experiment, and it included fast walking (running) the stairs from third to the ground floor and back, twice, as fast as possible. After the exercise they had 30 minutes break during which they had freedom to go and do whatever they want (for example, go to library, restaurant, have a walk, doing whatever they wanted). Basically, your main task will be to write a statement which will convince us that you are a part of the first group. To do so you will be given two tasks. The first task is to recall the last time you had an intense physical exercise (running, walking the stairs) and all the sensations you felt then and thoughts you had. It isn't important when that experience has happened, but it is crucial to try to recall that memory with as many details as possible. The second task is to write everything you can remember about your sensations and thoughts but as if they are consequences of the exercise you just had 30 minutes ago, walking up and down the stairs from zero to the third floor and back, as fast as possible. So, try to write a detailed statement about your previous sensations as if you just had them as a result of this specific activity, even though you didn't have that experience. You should write not only about your sensations and thoughts during the exercise, but also describe the immediate period thereafter and what did you do and where did you go. Again, it is important to explain your sensations and your experience with as many details as possible. If you succeed in convincing researchers that you suffered from high level of physical distress because of this experiment you will be rewarded bonus credit or an additional €7.5 voucher. All in all, the experiment will last for about 1 hour. You are able to stop during the experiment at any time.

APPENDIX 2

CHAPTER 3

STUDY 4

Supplemental Table 3.1.

List of words used in the Stroop task. English translation with the original (Dutch) version in parenthesis.

| Type of words English (and Dutch translation) | | |
|--------------------------------------------------|---------------------------|--------------------------|
| Neutral | Anxiety – related | Test anxiety – related |
| Chair (Stoel) | Illness (Ziekte) | Time (Tijd) |
| Umbrella (Paraplu) | Injury (Verwonding) | Test (Toets) |
| Solid (Massief) | Debts (Schulden) | Panic (Paniek) |
| Sitting (Zitten) | Nervous (Nerveus) | Worry (Bezorgd) |
| Walk (Wandeling) | Death (Dood) | Exam (Examen) |
| Hammer (Hamer) | Heartbeat (Hartslag) | Performance (Prestatie) |
| Pencil (Potlood) | Unemployed (Werkeloos) | Frustration (Frustratie) |
| Potato (Aardappel) | Failure (Mislukking) | Doubt (Twijfel) |
| Locker (Kluis) | Abandoned (Achtergelaten) | Confusion (Verwarring) |
| Shoe (Schoen) | Trembling (Huilveren) | Challenge (Uitdaging) |
| Sock (Sok) | Crazy (Gek) | Incomplete (Incompleet) |
| Water (Water) | Stroke (Beroerte) | Incorrect (Incorrect) |

Vignette (English translation)

XX is 21 years old. S/he is a student at Maastricht University. XX lives in Maastricht in a student house. XX's mother is currently living in another city. XX's father died in an accident when XX was 10 years old. XX's mother suffers from medical problems since then, so XX had to take care of her. XX started studies last year and came here after a caretaker had been arranged for mother. Currently XX is failing the year. XX suffers from high test anxiety, and this presents a huge obstacle for her/him. Therefore, XX was not able to do the tests in the previous period. XX has always had some problems with situations in which s/he was evaluated or judged, especially in test situations. However, this got worse over time, so now it is difficult for XX to take a test, because s/he starts panicking, sweating, and s/he fills like s/he is going to faint the minute the test starts. So, in this way XX failed a series of exams and (s)he has even running out all the resit possibilities. The only solution for XX to get an additional chance in retaking the exams is to go to the exam committee and tell them about the test anxiety problem. In order to succeed in getting a new chance for the tests, XX has to be very convincing in presenting the reasons for the privilege of extra resits.

Now, you are going to fill out the questionnaires about anxiety and related problems as if you were XX. Imagine that we are the exam committee and your questionnaires will be considered to be those of XX's. If you succeed in convincing us that your questionnaire scores point in the direction of severe test anxiety, you will participate in a lottery with a 1: 10 chance of winning an extra bonus of 20 euro's.

Now, we kindly ask you to read carefully the vignette again, and then start answering the questions as if you were XX.

Thank you and good luck!

Appendix

Supplemental Table 3.2.

Means and standard deviations across conditions on genuine symptoms and pseudosymptom subscales on SRSI.

| SRSI | | T1 <i>M(SD)</i> | T2 <i>M(SD)</i> | <i>Wilcoxon signed rank test Z</i> | <i>r</i> |
|----------------------|-------------------------|--------------------|--------------------|--------------------------------------------|----------|
| Credible symptoms | Cognitive | 1.27(1.67) | 6.18(2.36) | 3.93** | 0.59 |
| | Depressive | .73(.88) | 5.54(1.94) | 4.12** | 0.62 |
| | Pain | .82(1.10) | 3.32(2.64) | 3.60** | 0.54 |
| | Non-specific somatic | 2.09(2.16) | 8.14(2.00) | 4.02** | 0.60 |
| | PTSD / anxiety | 1.91(1.71) | 8.09(1.41) | 4.13** | 0.62 |
| Bogus symptoms | Cognitive/Memory | .36(.95) | 4.77(2.46) | 3.84** | 0.58 |
| | Neurological: Motor | .18(.50) | 1.59(1.30) | 3.45** | 0.52 |
| | Neurological: Sensory | .32(.72) | 2.86(2.33) | 3.69* | 0.55 |
| | Pain | .14(.35) | 1.54(2.40) | 2.55** | 0.38 |
| | Anxiety/Depression/PTSD | .23(.53) | 4.86(2.37) | 4.12** | 0.62 |
| Consistency check | | 3.59(1.47) | .82(1.01) | 3.97** | 0.60 |

Notes: * $p < .01$; ** $p < .001$; T1 = honest way of responding; T2 = feigning test anxiety.

STUDY 5

Supplemental table 3.3.

Pre-screening mean scores among eliminated participants (Impact of Events Scale – Revised (0-4), and PCL-5 score (0-4)).

| Eliminated participants | | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|----------------------------|
| | Impact groups (<i>n</i> = 29) | Actors (<i>n</i> = 21) |
| | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) |
| IES-R | 1.44 (.62) | 1.76 (.34) |
| • Intrusion | 1.47 (.96) | 1.87 (.53) |
| • Avoidance | 1.90 (.70) | 1.90 (.53) |
| • Hyperarousal | .78 (.87) | 1.45 (.67) |
| PCL-5 | 1.15 (.88) | 1.75 (.80) |
| <i>Note:</i> Participants were excluded based on: 1) Reported traumatic event (<i>n</i> = 3); 2) PLC-5 score above 3 (<i>n</i> = 4); and 3) The mean score on IES-R in general or/and on any of the three subscales above 1.5 (<i>n</i> = 43). Criteria 1 and 2 applied to the whole sample, while criterion 3 applied only to the Low impact and Actor group. | | |

Supplemental table 3.4.

Aversive Events List and endorsement frequencies across the whole sample (14 out of 18 listed events).

| Event | Frequency (%) |
|------------------------------------------------------------------------------------------------------------------------|-------------------|
| You moved (changed city or country or residence). | 17(26.6) |
| You ended an intimate relationship. | 8(12.5) |
| Someone close to you died (family member, or a close friend). | 8(12.5) |
| You ended a very important friendship. | 3(4.7) |
| You, a very close friend, or close family member had an accident that required emergency medical treatment. | 1(1.6) |
| You, a very close friend, or close family member has been hospitalized for a serious (life- threatening) illness. | 3(4.7) |
| You were fired from a job. | 1(1.6) |
| You had serious problems or disappointment at school or in an educational course (university, training program, etc.). | 3(4.7) |
| There was a significant reduction in your personal finances. | 4(6.3) |
| A pet (animal) to whom you were attached died, or got lost, or you had to give it away. | 2(3.1) |
| The behaviour of your family member was a significant problem for you. | 3(4.7) |
| You had unwanted fluctuation in your weight or your body size. | 1(1.6) |
| You had a serious conflict with a person close to you (a family member or a friend). | 5(7.8) |
| You were excluded from social events from your peers or close friends. | 5(7.8) |
| Total | 64(100.0) |

Appendix

Supplemental table 3.5.

Means and standard deviations of all groups on the Self-Report Symptom Inventory's plausible and pseudosymptom subscales.

| SRSI | Groups | | | χ^2 (2, <i>N</i> = 64) |
|-------------------------|---------------------------------|--------------------------------|----------------------------|-----------------------------|
| | High Impact (<i>n</i> = 22) | Low Impact (<i>n</i> = 24) | Actors (<i>n</i> = 18) | |
| | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | <i>M</i> (<i>SD</i>) | |
| Genuine PTSD/Anx/Dep | 4.55 (2.41) | 3.71 (1.99) | 9.34 (1.54) | 33.46* |
| Pseudo PTSD/Anx/Dep | 1.86 (2.15) | 1.04 (1.30) | 7.00 (2.50) | 34.68* |

Note: * $p < .01$; *Mann-Whitney U* tests indicated significant differences between Actors and both High and Low Impact groups ($ps < .001$), but no significant differences between the two Impact group ($ps > .05$).

The measure of fantasy proneness, the CEQ also correlated significantly and positively with the PTSD/Anxiety/Depression genuine and pseudosymptoms (Pearson's r of .53 and .48, $ps < .001$, respectively).

APPENDIX 3

CHAPTER 4

Scenario A

“Angst”

Stel je voor: je bent een jonge politiemedewerker. Je maakt lange dagen. En je bent zeer gemotiveerd. Je houdt van je werk. Dan is er die avond dat je dienst hebt en er een ernstig verkeersongeval gebeurt. Jij moet ter plekke de dingen regelen. Je ziet de verkeerslachtoffers, je hoort ze kermen van de pijn, je bent aangeslagen. In de dagen die volgen vraagt je baas niet een keer hoe het met je gaat. Je baas is sowieso een arrogante kerel. En dan realiseer je je dat je eigenlijk best wel eens een paar weken lekker thuis zou willen zitten. Zodat je eens een tijdje andere dingen kunt doen, die je ook leuk vindt: schilderen, muziek maken, oude vrienden bezoeken. Noem het een soort sabbatical. Je besluit naar de bedrijfsarts te gaan en te spelen dat je een ernstig trauma hebt overgehouden aan het verkeersongeval. De arts geeft je een aantal vragenlijsten. Je besluit die zo in te vullen dat de arts je een paar weken rust voorschrijft. Dus stel je voor dat jij die politiemedewerker bent. Vul met dat in het achterhoofd de vragenlijsten in. Probeer je rol zo overtuigend mogelijk te spelen.

“Anxiety”

Imagine this: you are a young police officer. You work hard. And you are very motivated. You love your job. Then there is this evening that you are on duty. There is this terrible traffic accident and it's your turn to go there and arrange everything. You see the victims, you hear them crying, you feel very sorry. In the days that follow, your boss is never asking you how you feel about what happened. Your boss is an arrogant guy anyway. And then you realize that you would love to stay for some weeks at home. So that you have the opportunity to do nice things: painting, playing music, visiting all friends. Call it a kind of sabbatical. You decided to see the police doctor and to simulate that you suffer from a trauma due to the accident that you witnessed. The doctor administers a couple of questionnaires to you. You decide to complete them in such way that the doctor will prescribe you some weeks of rest. So imagine that you are this police officer and with that in mind, complete the questionnaires. Try to be convincing.

Scenario B

“Pijn”

Stel je voor: Je bent student. Op een dag loop je over de zebra en word je aangereden door een oudere vent in een Mercedes Benz. Je valt op de grond. Je telefoon is kapot, je broek is kapot, maar voor de rest valt het mee. Maar de vent van de Mercedes blijkt een arrogante kerel. Hij vraagt niet hoe het met je gaat. Hij zegt dat je beter moet uitkijken. Je bedenkt dat je schadevergoeding van deze man wilt hebben. Je kan dat geld goed gebruiken. Dus haal je de politie erbij om een proces-verbaal van het ongeval op te maken en paar dagen later moet je naar de verzekeringsarts. Je besluit te spelen dat je ernstige pijnklachten hebt overgehouden aan het verkeersongeval. De arts geeft je een aantal vragenlijsten. Je besluit die zo in te vullen dat de arts wel tot de conclusie moet komen dat je iemand bent met ernstige pijnklachten. Dus stel je voor dat jij die student bent. Vul met dat in het achterhoofd de vragenlijsten in. Probeer je rol zo overtuigend mogelijk te spelen.

“Pain”

Imagine this: You are a student. One day you walk on the cross walk and a Mercedes Benz hits you. The driver is an older man. You fall on the ground. Your phone is damaged; your clothes are torn, but otherwise you are not hurt. However, the driver turns out to be an arrogant guy. He doesn't ask if you are well. He is just saying that you should have paid more attention. You decide that you want compensation money from this guy. So you insist that the police comes to write down an official report about the accident. A few days later you have to see the insurance doctor. You decide to simulate that you have serious pain complaints due to the accident. The doctor administers a couple of questionnaires to you. You decide to complete them in such way that the doctor will conclude that you are indeed a person with serious pain complaints. So imagine that you are this student and with that in mind, complete the questionnaires. Try to be convincing.

Appendix

Supplemental table 4.1.

Endorsement of genuine and pseudosymptoms on the SRSI subscales of honest comparison group, pain simulators, and anxiety simulators.

| SRSI scales | Honest comparison ¹ <i>M (SD)</i> | Pain simulators ² <i>M (SD)</i> | Anxiety simulators ³ <i>M (SD)</i> | <i>F</i> (2, 155) | Effect size (η^2) | <i>Cohen's d</i> |
|-------------------------|-------------------------------------------------|-----------------------------------------------|--------------------------------------------------|----------------------|--------------------------|-------------------------------------------------------------------|
| <i>n</i> | 51 | 54 | 53 | | | |
| Genuine symptoms | | | | | | |
| Cognitive | 2.04 (2.22) | 4.23 (2.97) | 6.06 (2.85) | 28.51 | .27 | 0.83 ^{1,2} 1.57 ^{1,3} 0.63 ^{2,3} |
| Depressive | 2.49 (2.04) | 4.48 (2.20) | 7.00 (2.51) | 51.91 | .40 | 0.94 ^{1,2} 1.97 ^{1,3} 1.07 ^{2,3} |
| Pain | 1.27 (1.13) | 8.59 (2.30) | 3.71 (3.29) | 125.14 | .62 | 4.04 ^{1,2} 0.99 ^{1,3} 1.72 ^{2,3} |
| Nonspecific somatic | 4.47 (2.75) | 7.57 (2.57) | 8.98 (2.14) | 44.14 | .36 | 1.16 ^{1,2} 1.83 ^{1,3} 0.60 ^{2,3} |
| PTSD/Anxiety | 2.06 (2.24) | 5.39 (2.71) | 8.02 (2.54) | 73.50 | .48 | 1.34 ^{1,2} 2.49 ^{1,3} 1.00 ^{2,3} |
| Total symptoms | 12.34 (7.56) | 30.26 (9.43) | 33.77 (10.66) | 78.73 | .50 | 2.09 ^{1,2} 2.32 ^{1,3} 0.35 ^{2,3} |
| Pseudosymptoms | | | | | | |
| Cognitive/Memory | .72 (1.10) | 2.03 (2.18) | 3.75 (2.56) | 28.49 | .27 | 0.76 ^{1,2} 1.54 ^{1,3} 0.72 ^{2,3} |
| Neurological: motor* | .27 (.57) | 2.28 (2.73) | 2.40 (2.99) | 13.03 | .14 | 1.02 ^{1,2} 0.98 ^{1,3} 0.04 ^{2,3} |
| Neurological: sensory* | .84 (1.12) | 1.76 (2.30) | 3.09 (2.89) | 13.29 | .15 | 0.51 ^{1,2} 1.03 ^{1,3} 0.51 ^{2,3} |
| Pain | .19 (.40) | 4.23 (2.95) | 1.94 (3.01) | 35.20 | .31 | 1.92 ^{1,2} 0.81 ^{1,3} 0.77 ^{2,3} |
| Anxiety/Depression | .64 (1.19) | 2.57 (2.65) | 6.19 (2.74) | 76.56 | .50 | 0.94 ^{1,2} 2.62 ^{1,3} 1.34 ^{2,3} |
| Total | 2.68 (2.90) | 12.87 (10.50) | 17.38 (12.00) | 33.08 | .30 | 1.33 ^{1,2} 1.68 ^{1,3} 0.40 ^{2,3} |

Note: All $ps < .001$; **Bonferroni post hoc* test indicated significant differences ($p < .05$) between all three groups, except for the neurological sensory symptoms subscale (honest group and pain simulators, $p = .144$), and for motor symptoms subscale (pain and anxiety simulators, $p = 1.00$). “^{1,2}” indicates contrast between the honest comparison group and pain simulators, “^{1,3}” indicates contrast between the honest comparison group and anxiety simulators, and “^{2,3}” indicates contrast between the pain simulators and anxiety simulators.

APPENDIX 4
CHAPTER 5

Newspaper Article

Two dead in crash between passenger train and car

In a crash between a passenger train and a car on Monday evening two people died. It concerns the passengers of the car, including one male (28) and a 4-month-old baby. There were around 100 passengers on the train. They remained unharmed. The exact cause of the accident remains unclear. Information released by the relatives showed that the man tried to send a picture of his young daughter, using his smartphone, right before the crash. It is possible that the man was distracted and

did not notice the railroad lights. The ongoing police investigation will have to offer conclusive information on this accident. "



APPENDIX 5

CHAPTER 6

Supplemental table 6.1.

The order of the items according to the mean plausibility score in Western and non-Western group (from 1 – definitely authentic to 63 – definitely exaggerated).

| Items | Western Rank | Western Mean Western SD | Non-Western Mean (SD) | Non-Western Rank |
|------------------|--------------|-------------------------|-----------------------|------------------|
| BSI-18 – Item 1 | 22 | 1.58 (.71) | 1.92 (.95) | 20 |
| BSI-18 – Item 2 | 6 | 1.59 (.76) | 1.48 (.80) | 1 |
| BSI-18 – Item 3 | 11 | 1.62 (.75) | 1.75 (1.06) | 13 |
| BSI-18 – Item 4 | 18 | 1.64 (.84) | 1.74 (1.02) | 7 |
| BSI-18 – Item 5 | 20 | 1.64 (.77) | 1.95 (1.02) | 23 |
| BSI-18 – Item 6 | 15 | 1.66 (.64) | 1.92 (1.09) | 21 |
| BSI-18 – Item 7 | 7 | 1.70 (.91) | 1.75 (1.06) | 14 |
| BSI-18 – Item 8 | 14 | 1.70 (.76) | 1.88 (1.21) | 18 |
| BSI-18 – Item 9 | 21 | 1.79 (.76) | 1.69 (.92) | 10 |
| BSI-18 – Item 10 | 4 | 1.82 (.92) | 1.62 (1.06) | 6 |
| BSI-18 – Item 11 | 8 | 1.84 (.76) | 1.56 (.90) | 2 |
| BSI-18 – Item 12 | 3 | 1.86 (.80) | 1.61 (1.01) | 3 |
| BSI-18 – Item 13 | 2 | 1.86 (.86) | 1.66 (1.25) | 7 |
| BSI-18 – Item 14 | 5 | 1.86 (.98) | 1.61 (.89) | 5 |
| BSI-18 – Item 15 | 13 | 1.89 (.97) | 1.84 (.99) | 17 |
| BSI-18 – Item 16 | 10 | 1.91 (1.05) | 1.67 (1.18) | 9 |
| BSI-18 – Item 17 | 16 | 1.93 (.97) | 1.92 (1.09) | 22 |
| BSI-18 – Item 18 | 1 | 1.98 (.92) | 1.61 (.82) | 4 |
| DES-T – Item 1 | 49 | 2.00 (1.04) | 2.28 (1.19) | 38 |
| DES-T – Item 2 | 48 | 2.04 (.97) | 2.40 (1.28) | 46 |
| DES-T – Item 3 | 43 | 2.04 (1.10) | 2.35 (1.23) | 43 |
| DES-T – Item 4 | 45 | 2.07 (.91) | 2.43 (1.25) | 48 |
| DES-T – Item 5 | 35 | 2.09 (.94) | 2.15 (1.38) | 34 |
| DES-T – Item 6 | 32 | 2.11 (.95) | 2.15 (1.11) | 33 |
| DES-T – Item 7 | 47 | 2.18 (.96) | 2.63 (1.49) | 56 |
| DES-T – Item 8 | 19 | 2.20 (.90) | 1.70 (1.15) | 11 |
| SIMS – Item 1 | 41 | 2.23 (1.10) | 2.52 (1.19) | 51 |
| SIMS – Item 2 | 12 | 2.30 (.95) | 1.77 (.93) | 16 |
| SIMS – Item 3 | 17 | 2.35 (1.10) | 2.06 (1.13) | 27 |
| SIMS – Item 4 | 42 | 2.36 (1.03) | 2.12 (1.20) | 29 |
| SIMS – Item 5 | 37 | 2.38 (1.14) | 2.05 (1.39) | 26 |
| SIMS – Item 6 | 26 | 2.41 (1.02) | 2.14 (1.11) | 30 |
| SIMS – Item 7 | 23 | 2.41 (1.09) | 1.90 (1.12) | 19 |
| SIMS – Item 8 | 36 | 2.41 (1.06) | 2.15 (1.25) | 32 |
| SIMS – Item 9 | 9 | 2.43 (1.08) | 1.67 (.97) | 8 |
| SIMS – Item 10 | 59 | 2.45 (1.19) | 3.02 (1.43) | 63 |
| *SIMS – Item 11 | 60 | 2.47 (1.16) | 2.47 (1.40) | 49 |
| *SIMS – Item 12 | 38 | 2.48 (1.04) | 2.22 (1.29) | 36 |
| SIMS – Item 13 | 25 | 2.55 (1.11) | 2.26 (1.28) | 37 |
| SIMS – Item 14 | 28 | 2.57 (1.16) | 2.31 (1.35) | 40 |
| SIMS – Item 15 | 57 | 2.58 (1.19) | 2.94 (1.25) | 62 |
| SIMS – Item 8 | 36 | 2.41 (1.06) | 2.15 (1.25) | 32 |
| SIMS – Item 9 | 9 | 2.43 (1.08) | 1.67 (.97) | 8 |
| SIMS – Item 10 | 59 | 2.45 (1.19) | 3.02 (1.43) | 63 |
| *SIMS – Item 11 | 60 | 2.47 (1.16) | 2.47 (1.40) | 49 |
| *SIMS – Item 12 | 38 | 2.48 (1.04) | 2.22 (1.29) | 36 |
| SIMS – Item 13 | 25 | 2.55 (1.11) | 2.26 (1.28) | 37 |
| SIMS – Item 14 | 28 | 2.57 (1.16) | 2.31 (1.35) | 40 |
| SIMS – Item 15 | 57 | 2.58 (1.19) | 2.94 (1.25) | 62 |
| SIMS – Item 16 | 30 | 2.60 (1.17) | 2.04 (1.19) | 25 |
| SIMS – Item 17 | 31 | 2.61 (1.04) | 2.07 (1.49) | 28 |
| SIMS – Item 18 | 63 | 2.61 (1.09) | 2.92 (1.36) | 61 |
| *SIMS – Item 19 | 55 | 2.62 (1.17) | 2.20 (1.17) | 35 |
| SIMS – Item 20 | 33 | 2.66 (1.03) | 2.15 (1.16) | 31 |
| SIMS – Item 21 | 24 | 2.71 (1.12) | 1.96 (1.18) | 24 |
| SIMS – Item 22 | 29 | 2.75 (1.10) | 2.31 (1.35) | 41 |
| SIMS – Item 23 | 53 | 2.78 (1.00) | 2.90 (1.24) | 60 |

Appendix

| | | | | |
|----------------|----|-------------|-------------|----|
| SIMS – Item 24 | 58 | 2.88 (1.25) | 2.73 (1.33) | 58 |
| SIMS – Item 25 | 39 | 2.88 (1.24) | 2.37 (1.40) | 44 |
| SIMS – Item 26 | 40 | 2.91 (1.07) | 2.53 (1.28) | 52 |
| SIMS – Item 27 | 54 | 2.98 (1.26) | 2.39 (1.30) | 45 |
| SIMS – Item 28 | 51 | 2.98 (1.18) | 2.49 (1.30) | 50 |
| SIMS – Item 29 | 61 | 3.02 (1.15) | 2.67 (1.37) | 57 |
| SIMS – Item 30 | 46 | 3.02 (1.21) | 2.30 (1.17) | 39 |
| SIMS – Item 31 | 27 | 3.09 (1.20) | 1.76 (1.19) | 15 |
| SIMS – Item 32 | 62 | 3.11 (1.11) | 2.85 (1.26) | 59 |
| SIMS – Item 33 | 44 | 3.13 (1.10) | 2.54 (1.30) | 53 |
| SIMS – Item 34 | 52 | 3.18 (1.18) | 2.56 (1.25) | 54 |
| SIMS – Item 35 | 38 | 3.18 (1.15) | 2.42 (1.21) | 47 |
| SIMS – Item 36 | 34 | 3.34 (1.13) | 2.31 (1.28) | 42 |
| SIMS – Item 37 | 56 | 3.43 (1.17) | 2.61 (1.34) | 55 |

Note: * Adjusted alpha < .02; SIMS - *Structured Inventory of Malingered Symptomatology* (SIMS; Smith & Burger, 1997), DES-T - *taxon items of Dissociative Experiences Scale* (DES; Bernstein & Putnam, 1986); BSI-18 - the *Brief Symptom Inventory-18* (BSI-18; Derogatis, 2001).

