Reprinted from the German Journal of Psychiatry · http://www.gjpsy.uni-goettingen.de · ISSN 1433-1055

Posttraumatic Stress Disorder Can Easily Be Faked, but Faking Can Be Detected in Most Cases

Thomas Merten[1], Robert Lorenz[2], and Sam Schlatow[2]

[1]Vivantes Netzwerk für Gesundheit, Klinikum im Friedrichshain, Berlin, Germany
[2]Humboldt University Berlin, Institute of Psychology, Berlin, Germany

Corresponding author: Thomas Merten, PhD, Vivantes Klinikum im Friedrichshain, Dept. of Neurology, Landsberger Allee 49, D-10249 Berlin, Germany; Email: thomas.merten@vivantes.de

Abstract

Background: Symptom overreporting and malingering are a topic of utmost importance in the field of forensic assessment in cases of claimed posttraumatic stress disorder (PTSD). Individuals may be coached regarding the symptoms of the condition they claim and regarding malingering detection strategies. Symptom validity tests (SVTs) should be as resistant against such knowledge as possible.
Method: Healthy, well-educated volunteers were instructed to simulate how a person in a forensic examination may respond in order to convincingly feign mental disorder after a traumatic event. Four groups of these experimental malingerers (n = 20, each) were given different scenarios. In a 2 × 2 experimental design, they received either specific PTSD symptom information or not, and either a warning against exaggerating or not.
Results: Using a regression based formula for the German version of the Impact of Event Scale–Revised, the majority of simulators were wrongly classified as PTSD patients; in performance tests, they also demonstrated cognitive impairment. However, 96 percent of the participants were correctly classified as malingerers when a multi-method approach of symptom validity assessment was used. The Structured Inventory for Malingered Symptomatology, the Morel Emotional Numbing Test, and the Word Memory Test performed well in identifying feigned PTSD, while the MMPI-2 Fake Bad Scale and the Reliable Digit Span did so to a lesser degree. Only three simulators who received symptom information and warning were able to pass all five symptom validity measures. Participants who received symptom information alone were not able to perform in a more convincing way.
Conclusion: The results demonstrate that PTSD symptoms can easily be presented by healthy adults. Within the framework of this experimental analog study it was shown that feigned PTSD can be detected in most cases if a multimethod approach to symptom validity assessment is employed. The usefulness of symptom validity testing in real-world forensic evaluations has been demonstrated by a number of other studies (German J Psychiatry 2010; 13: 140-149).

Keywords: Posttraumatic Stress Disorder, malingering, symptom validity testing, negative response bias, Morel Emotional Numbing Test, symptom overreporting

Received: 6.7.10
Revised version: 29.7.10
Published: 15.8.10

Introduction

With an increased awareness of the scope and scale on which medical conditions and mental or cognitive disorders are being feigned in medico-legal examinations (e.g., Larrabee, 2007; Mittenberg et al., 2002; Quezada-Ortega et al., 2006; Richman et al., 2005; Stevens et al., 2008), the discussion about fabricated or exaggerated claims of posttraumatic stress disorder (PTSD) has gained increased attention (Freeman et al., 2008; McNally, 2003; Rosen, 2004a; Taylor et al., 2007). In light of such spectacular early cases like the sinking of the fish processing vessel Aleutian Enterprise on March 22, 1990 and the subsequent misdiagnosis of PTSD in a considerable number of survivors who had been coached by their attorneys, Rosen (2004b, 2006) has repeatedly called attention to the fact that PTSD base rate estimates may be grossly distorted; as a consequence, the whole scientific data base on PTSD may be contaminated by non-authentic cases. Claims of PTSD in forensic examinations or in other contexts where primary or

secondary gain is at stake should not be taken at face value. The main reason why PTSD appears to be particularly prone to symptom exaggeration and malingering can be seen in the fact that the diagnosis is often based on patients' self report. In contrast to traumatic brain injury, more objective measures like MRI scans or documented loss of consciousness are not available for the diagnosis of PTSD.

How much symptom report and patient behaviour may be distorted when secondary gain is at stake, like in the Aleutian Enterprise sinking, was best "illustrated by a survivor who explained that he did not know why he went to see a psychologist except that his attorney wanted him to be evaluated. He said he wanted to go back to fishing but his attorney told him it would look better if he did not work" (Rosen, 1995, p. 84). Hickling et al. (2002) demonstrated in a simulator study how easy it was to fool experienced clinical evaluators with symptoms of posttraumatic stress disorder (PTSD) if the diagnosis was based on interview alone.

The subsequent development and refinement of special techniques and tests to detect distortion in medico-legal examinations has made the identification of false symptom reports easier and less prone to clinical judgment which may be considerably biased by the clinician's personal opinions and beliefs. In fact, symptom validity assessment according to modern methodological standards is now expected to be an integral part of neuropsychological evaluations (Bush et al., 2005). Apart from symptom validity tests proper which aim at identifying fabricated or exaggerated cognitive impairment, a number of psychological symptom validity measures have been developed.

Some of them, like the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; Widows & Smith, 2005), are instruments specially designed for the detection of negative response bias, while others are scales imbedded into more complex personality tests, like the Minnesota Multiphasic Personality Inventory-2 (MMPI-2). Among the more recently developed MMPI-2 scales thought to be sensitive to feigned psychopathology are the Infrequency-Psychopathology Scale *F(p)* (Arbisi & Ben-Porath, 1995), the Fake Bad Scale FBS (Lees-Haley et al., 1991), and the Response Bias Scale RBS (Gervais et al., 2007). In a recent review, Singh et al. (2007) have found that "malingering of psychiatric disorders is perhaps commoner than previously considered and is particularly difficult to detect" (p. 131); they recommended instruments such as those mentioned here as a useful adjunct to support detection of malingering.

Morel (1998a,b) developed an instrument specifically designed to identify faked PTSD. The Morel Emotional Numbing Test (MENT) was designed in a way that virtually any adult, unless suffering from severe specific neurocognitive disorders related to word processing, visual acuity, spatial neglect or face processing would be able to complete the task with 90 to 100 percent accuracy. However, patients who were intentionally trying to convince the examiner of the presence of PTSD symptoms may well distort their performance and score above a cutoff for low effort.

A number of studies have demonstrated that negative response bias in PTSD claimants can be identified if validity measures are included in the assessment (e.g., Freeman et al., 2008; Frueh et al., 2000; Greiffenstein et al., 2004; for a recent review of the relevant literature, see Rubenzer, 2009). All assessment approaches outlined above may bear some potential for the detection of feigned PTSD because claimants may follow very different scenarios or strategies of false symptom presentation. Thus, in a recent prospective study with 61 civil forensic patients with claimed trauma symptoms, Merten et al. (2009) found evidence for possible negative response bias in as much as 70 percent of the claimants when a multi-method approach was applied. Results of three stand-alone symptom validity tests were available for all patients (MENT, SIMS, and WMT), 25 percent of the participants scored positive on all three tests. Moreover, high probability of negative response bias was associated with symptom overreporting and demonstration of cognitive deficits in performance tests. The results indicated that substantial rates of uncooperativeness must be expected in civil forensic patients with claimed PTSD.

The current study aimed at following this line of research and investigating how far a variety of symptom validity measures are sensitive to detect feigned PTSD symptoms using a different design. For this purpose, an analog study with experimental simulators was designed. Experimental malingering studies with participants simulating feigned mental or cognitive disorders are one of the indispensable methodological approaches to modern malingering research. A related research question of equal importance which was to be addressed in the study was to investigate the influence of symptom information and coaching, which may both be found among claimants.

Symptom information and coaching have, for some time, been an important issue to take into account with malingering research. In fact, a survey by Wetter and Corrigan (1995) with 70 practicing attorneys and 150 law students demonstrated that the majority of respondents believed that an attorney should discuss with his or her clients what psychological testing involves. A substantial proportion of the respondents believed they should inform a client of validity scales on a psychological test. A survey by Essig et al. (2001) showed that attorneys typically spend up to an hour preparing their clients for neuropsychological evaluations and commonly cover test content, detection of malingering, and brain injury symptoms. Lees-Haley (1997) suggested that "several converging lines of evidence point to the conclusion that attorneys are influencing data relied upon by psychological experts" (p. 321). This author also pointed to the importance of research on methods for correcting for such influences. Thus, an instrument's resistance against coaching has been proposed to be a quality marker for symptom validity tests (Hartman, 2002). In the meantime, an important body of research literature on coaching has been accumulated; several summary reports are currently available (Blaskewitz et al., 2007; Gorny & Merten, 2005; Suhr & Gunstad, 2007).

For the current study, two factors were subject to systematical variation in order to study the potential of different symptom validity measures to detect feigned PTSD. These factors were (1) symptom information given to the participants instructing them about characteristics of posttraumatic stress, (2) a warning that specific validity measures could be

employed in the assessment procedure and, consequently, participants should not exaggerate false symptom report in order to avoid detection. A similar 2 x 2 design was described by Dunn et al. (2003) with brain injury as the condition to be faked by the simulators.

The main hypothesis according to previous research (cf., in particular, Gorny & Merten, 2005) was that mere symptom information should not enhance the participants' ability to present PTSD symptoms in a believable way (that is without being detected by symptom validity indicators). However, warning or the combination of information and warning may be more effective, but most or all experimental malingerers would be successfully identified if a multi-method approach to malingering detection is used. Moreover, in accordance with previous results it can be expected that a number of participants may be able to identify symptom validity tests as such and, yet, they will not be successful in passing them (e.g., Gorny & Merten, 2005; Merten et al., 2005).

Method

Research Participants

A group of 80 well-educated healthy adults was investigated, all of them being native speakers of German. They were either undergraduate first year students of psychology (n = 29; 36%) or recruited as volunteers in the social networks of the two experimenters (S. S. and R. L.). Students and professionals from psychology-related fields were excluded, except for the first year students. The sample was chosen to be homogeneous in terms of educational background, so inclusion was restricted to well-educated adults, with at least university entrance level (so called Abitur).

The total group consisted of 40 males and 40 females with a mean age of 26.4 years (SD = 8.9; range: 19 to 53 years) and a mean of 15.8 years of education (SD = 2.5). 58 participants (73%) were university undergraduates while 13 held a university degree. Post-hoc analyses showed that there were no systematic differences in terms of age, gender or education between the four experimental groups.

Procedure and Instruments

According to a pre-established design, participants were assigned to one of four experimental groups. All groups were experimental malingerers. They all received a basic scenario designed to acquaint them with the role of a person who had suffered an armed robbery on the way home from work. The detailed scenario described the hold-up and the shock the person suffered so he or she decided not to go to work for a couple of days. Some weeks later, on the background of high work load and frustrating working conditions, the participant learned from a newspaper article about posttraumatic stress disorder and depression as common sequelae of traumatic events. According to German labor law, attacks which happen on the way from or to work are eligible for workers compensation payment. So, on the basis of a conscious decision not to waive that opportunity, he or she reported the claim to the insurer and was sent to an independent psychological assessment. In order to get compensation, the participant had to demonstrate in a convincing manner that significant symptoms were persisting.

Groups 2 and 4 obtained, in addition to that basic scenario, more detailed information about sequelae of posttraumatic stress (about half a printed page). For groups 3 and 4, the scenario ended with an element of warning. The participants were informed about the inclusion of validity measures in the assessment and warned not to exaggerate symptom report or else they might fail and their report would be judged to be implausible. On the basis of this factor design, the effects of both symptom information and warning could be examined.

In order to study specific PTSD-like symptom report as well as generalization to other domains, both self-report instruments and performance tests were included in the experiment. The following standardized instruments were administered in the order given here:

(1) The WAIS-III subtest Digit Span Forward and Backward (Wechsler, 1997). As a symptom validity index, the Reliable Digit Span (RDS) was computed (Greiffenstein et al., 1994).

(2) The oral version of the WMT (Green, 2003), Immediate Recognition (IR) subtest. The examiner reads aloud twenty word pairs, such as "rain – snow" and "animal – mouse". Immediate Recognition consists of choosing one of the words in the list (e.g., "rain") when given a pair of words containing one list word and one non-list, foil word (e.g., "rain" and "storm").

(3) The Trail Making Test (TMT: Reitan, 1992). This is a test of attention and executive functions which is commonly used in neuropsychological assessment. Because of largely negative results in previous studies (e.g., Martin et al., 2003; Merten et al., 2005), the ratio TMT-B : TMT-A was not used as an indicator of low effort in this study although it had been proposed to constitute a measure indicative for malingering (Ruffulo et al., 2000).

(4) The German version of the Impact of Event Scale–Revised (Maercker & Schützwohl, 1998; Weiss & Marmar, 1996). The authors of the German adaptation maintained the original IES response format ascribing 0, 1, 3, and 5 points to the four possible responses "not at all", "rarely", "sometimes" and "often". Thus, the total scores and the score of three symptom clusters (Intrusion, Hyperarousal, and Avoidance) can not be compared to IES-R studies based on the original English-language version. Instead, Maercker and Schützwohl (1998) developed a formula based on a logistic regression, with a resulting sensitivity of .70 to .76 and a specificity of .88 to .89. The formula was:

*Diagnostic Value X = -0.02 * Intrusion + 0.07 * Avoidance + 0.15 * Hyperarousal – 4.36*

Probable PTSD was classified if the diagnostic value was above 0.

(5) A set of 200 items which constitute the raw version of a questionnaire for discriminating plausible psychopathology from exaggerated symptom report. For validation purposes only, the items of the Fake Bad Scale (FBS) developed by Lees-Haley et al. (1991) were dispersed throughout the questionnaire. Only the FBS scores were included in the current analyses while the questionnaire scale scores themselves have not yet been validated.

If necessary, the administration of the questionnaire was interrupted about thirty minutes after the end of the WMT-IR trial and was resumed after the following WMT subtests. In cases where the questionnaire was completed earlier (i.e., before the WMT-DR was scheduled), items from the Standard Progressive Matrices (SPM: Raven, 1956) were given in order to fill the time gap. The SPM results were not included in the analyses. This procedure was thought to guarantee the implementation of the time regimen which is warranted for the WMT and which should strictly be followed in this kind of test-related research.

(6) 30 minutes after the completion of the WMT-IR trial, the WMT Delayed Recognition (DR), Multiple Choice Recognition (MCR), Paired Associates (PA), and Free Recall (FR) subtests were given. DR and IR are the same type of recognition task (see above) but using different foil words for the DR trial. In the MCR trial, the first word of each pair from the original learning list is presented. The participant has to choose the word that originally came with it in the list, from a selection of eight words. In the Paired Associates subtest, the person is told the first word and is asked to say the second word (e.g., "rain", to which a correct response would be "snow"). In the Free Recall subtest, the person is asked to repeat as many words as possible from the original word list.

(7) The German adaptation of the Morel Emotional Numbing Test (MENT: Morel, 1998a). This is a test specifically designed for detecting feigned posttraumatic stress disorder. The concept of the test refers to the PTSD symptom of emotional numbing. Patients with false claims of PTSD may display implausible difficulties in the perception of emotions. The test has demonstrated its usefulness in a number of studies which have recently been summarized by Morel and Shepherd (2008).

(8) The Structured Inventory of Malingered Symptomatology (SIMS: Widows & Smith, 2005; German version: Cima et al., 2003). This is a 75-item questionnaire developed to assess patients' endorsement of unlikely, bizarre or very rare symptoms which may be perceived by respondents as belonging to known syndromes. Apart from a total score, scores for five subscales (symptom domains) can be obtained: Low Intelligence (LI), Affective Disorders (AF), Neurological Impairment (N), Psychosis (P), and Amnestic Disorders (AM).

(9) 20 minutes after the completion of the WMT-FR trial, the WMT Long Delayed Free Recall (LDFR) subtest was given. Like in the FR trial, the person is asked to say as many words from the original word list as possible. Again, SPM items were given but not included in the analyses if a participant finished the previous task before the starting point.

For the current study, the cutoffs proposed in the original publications of the symptom validity scales were used (Green, 2003; Morel, 1998a; Cima et al., 2003; Lees-Haley et al., 1991; Greiffenstein et al., 1994). While different cutoffs may be used depending on referral background, group membership or statistical considerations, it was beyond the scope of this study to investigate optimal cutoff employment.

To verify role understanding a multiple-choice questionnaire was developed as a pre-experimental manipulation check. Apart from ensuring role commitment, the questionnaire was conceived to be an additional part of the scenario, information or warning because basic information was repeated there. All questions were related to the scenario which had been read by the participants beforehand. If participants committed more than one error in the pre-experimental check, they had to study the scenario again carefully, and the questionnaire was given for a second time. This, in fact, was necessary in 26 cases.

As an incentive, participants of all groups were told that the most convincing of them would win a sum of 50 Euros (equivalent to about $70). For fairness, this sum was awarded to one participant of each group.

As a post-experimental manipulation check to evaluate role commitment after the experiment, a separate set of questions was presented to the participants. Finally, the participants of all groups were asked whether, in their opinion, tests had been included which might have served to detect symptom exaggeration or malingering. If this was confirmed, they were asked to specify the instrument.

Results

The raw scores of the four groups obtained for the different variables are contained in Table 1. For the naïve group (no symptom information, no warning), psychopathology is reported extensively in the IES-R, but also neuropsychological test scores are lower than expected for healthy adults (cf. TMT norms in Tombaugh, 2004; German digit span norms in Härting et al., 2000). IES-R scale scores in all three subscales (Intrusion, Avoidance, Hyperarousal) and for all groups were well above those published by Maercker and Schützwohl (1998) for a group of former political prisoners in East Germany and a group of crime victims. Using the regression-based formula proposed by those authors, 78 percent of all participants would have been wrongly classified as suffering from PTSD (95% of naïve, 80% of informed, 65% of warned and 70% of informed and warned participants).

The elevated scores in all SIMS subscales, which were found, in particular, for the naïve group and for the informed group, point to a general tendency for overgeneralizing symptom report when psychopathology is feigned. Among the subscales, Affective Disorders, Neurological Impairment, and Amnestic Disorders appeared to be particularly prone to overgeneralization.

Table 1. Raw Scores of the Test Variables for the Different Experimental Groups.

Test Variables	Group 1 Naïve		Group 2 Informed		Group 3 Warning		Group 4 Info + Warning		2 x 2 ANOVA	Main Effect Information	Main Effect Warning	Interaction
	m	SD	*m*	SD	*m*	SD	*m*	SD	*F*(*df*=3)	*F*(*df*=1)	*F*(*df*=1)	*F*(*df*=1)
Symptom Validity Measures												
Word Memory Test-IR (%)	78.88	17.73	74.88	18.31	77.25	15.28	81.30	12.18	0.56			
Word Memory Test-DR (%)	74.80	21.12	65.88	18.18	77.80	15.74	79.80	14.10	2.47+			
Word Memory Test-Co (%)	70.88	19.15	67.88	16.00	70.30	16.46	74.25	16.18	0.48			
Reliable Digit Span	6.80	2.17	6.65	2.70	7.30	1.49	8.25	2.22	2.18+			
Emotional Numbing Test	13.60	10.24	22.15	7.86	14.35	9.26	11.80	8.20	5.24*	2.25	5.77*	7.71*
SIMS Total Score	29.60	9.48	30.55	10.56	26.55	10.34	22.95	7.40	2.59+			
SIMS Neurological	6.10	2.95	4.80	3.02	4.75	2.15	4.15	2.13	2.00			
SIMS Affective Disorders	9.20	2.14	10.10	2.13	9.25	2.12	8.95	1.96	1.15			
SIMS Psychosis	3.35	2.11	3.60	3.50	2.40	2.23	1.75	1.83	2.35+			
SIMS Low Intelligence	2.95	1.99	4.10	3.31	3.50	3.05	2.60	2.33	1.16			
SIMS Amnestic Disorders	8.00	3.04	7.95	3.97	6.65	4.22	5.50	3.22	2.14			
Fake Bad Scale	21.80	2.65	20.95	4.05	21.25	3.61	20.95	4.36	0.23			
Neuropsychological Tests												
Digit Span Forward	6.70	2.76	6.60	2.87	7.20	2.24	8.60	2.82	2.36+			
Digit Span Backward	4.80	1.77	4.25	2.02	5.30	1.78	6.20	1.88	3.94*	0.18	8.63*	3.02+
Word Memory Test–MC	57.50	25.47	47.75	20.74	55.25	21.67	62.25	19.63	1.51			
Word Memory Test–PA	57.25	26.73	51.75	19.42	57.25	16.26	66.00	18.18	1.65			
Word Memory Test–DFR	38.00	19.43	33.25	11.36	37.13	17.36	43.63	16.75	1.35			
Word Memory Test–LDFR	37.38	18.42	31.88	13.40	34.50	20.43	43.75	17.40	1.68			
Trail Making Test A	44.45	17.66	58.95	58,80	39.20	13.45	33.25	9.47	2.39+			
Trail Making Test B	81.95	28.58	88.40	52.64	76.90	20.01	65.10	17.31	1.81			
Symptom Scales												
Impact of Event Scale	78.85	10.14	78.10	12.49	73.90	11.42	73.65	10.62	1.19			
IES-R Intrusion	26.10	5.85	27.40	5.60	25.65	4.80	24.05	5.54	1.28			
IES-R Avoidance	24.85	6.67	23.90	5.88	23.35	5.48	25.00	5.47	0.36			
IES-R Hyperarousal	27.90	4.95	26.80	5.36	24.90	5.30	24.60	4.19	2.00			

Notes: SIMS Structured Inventory of Malingered Symptomatology, IES-R Impact of Event Scale–Revised, IR Immediate Recognition, DR Delayed Recognition, Co Consistency, MC Multiple-Choice, PA Paired Associates, DFR Delayed Free Recall, LDFR Long Delayed Free Recall.
+ $p < 0.10$; * $p < 0.05$

Table 2. Percentage of Participants Correctly Classified as Suspect for Faking Psychopathology (*n* = 20, per group). Percentage of Total Passes in the Symptom Validity Measures. Group 1: Naïve; Group 2: Informed; Group 3: Warned; Group 4: Informed and Warned.

Instrument	Group				Total Sensitivity
	1	2	3	4	
Word Memory Test	75	85	75	75	78
Reliable Digit Span	65	60	60	35	55
Morel Emotional Numbing Test	70	95	75	65	76
SIMS – Total Score	90	90	85	70	84
Fake Bad Scale	75	65	70	65	69
Number of Passes					
In All Five Criteria	0	0	0	15	4
In 4 Criteria	10	0	5	0	4
In 3 Criteria	10	10	15	15	13
In 2 Criteria	15	20	15	15	16
In One of the Criteria	25	35	40	40	35
In None of the Criteria	40	35	25	15	29

Separate analyses of variance (ANOVAs) were performed to test the effects of symptom information (yes/no) and warning not to exaggerate symptom report (yes/no). As can be seen from the trends in the results, symptom information apparently had little influence on test scores. For the WMT and the TMT variables, slightly worse results were obtained for informed participants, but these differences were all non-significant. The ANOVAs showed trends ($p < .10$), but significant results were observed only for the MENT and WAIS-III Digit Span Backward. Only for these two instruments, main effects and interaction were computed. For both variables, significant effects of warning were obtained, with better (i.e., less extreme) results for participants who received a warning. Moreover, a significant interaction was obtained for the MENT (Figure 1). On the MENT, informed and warned participants committed only half as many errors as informed participants. For this instrument, information without warning resulted in a deterioration of test performance.

While a trend was obtained for the SIMS to show a somewhat reduced potential to identify malingering in informed and warned participants, the same was not true for the MMPI-2 Fake Bad Scale which, in turn, appeared to be resistant to both information and warning.

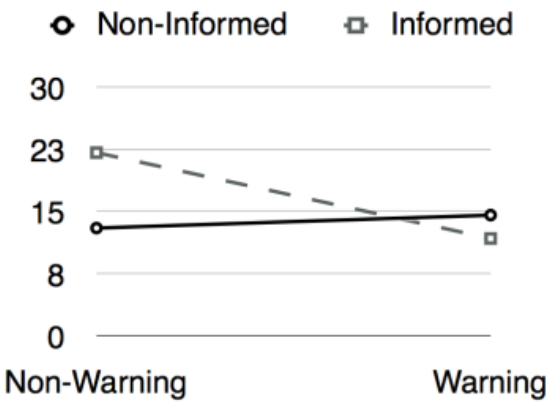

Figure 1. Results of the Morel Emotional Numbing Test (error scores) for four different simulator groups

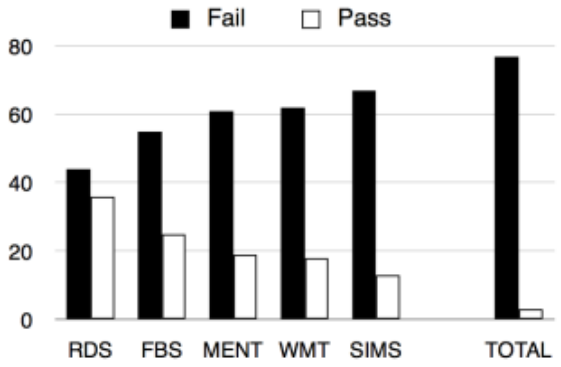

Figure 2. Number of passes and fails for the symptom validity measures. In the context of this study, the final decision *fail* was made if at least one of the measures was a fail

These main results can also be found on the level of passes and fails in the symptom validity instruments (Table 2, Figure 2). The total sensitivity over all four groups is highest for the SIMS with 84 percent of correct identifications; the lowest sensitivity to detect feigning was obtained for the RDS. 96 percent of the experimental malingerers would have been detected by at least one SVT, while 93 percent failed in at least two of the instruments. A relatively high percentage of experimental malingerers (36%) were identified by all five measures. Those three participants who passed all measures were all members of the warned and informed group.

Most participants (95%) thought that there were tests in the battery which aimed to identify symptom fabrication or exaggeration. In particular, 22 of them (28%) correctly identified the WMT as such an instrument, 26 (33%) the MENT, 13 (16%) the SIMS, and 24 (30%) the questionnaire which included the FBS. Also, Digit Span was named by 8 participants (10%), or the Trail Making Test by 10 (13%). Even the IES-R was thought to be a malingering measure, but it was named only once.

Discussion

The study investigated a number of German-language adaptations of a PTSD symptom report questionnaire, neuropsychological tests, and symptom validity tests in the context of claimed PTSD. First of all, healthy participants from the normal population are well capable of reporting psychopathological symptoms, which is not surprising at all. Most symptom report questionnaires, including PTSD scales, rely on patients' self-report and do not contain validity scales. Any person can endorse symptoms *ad libitum*. Casual evidence shows that some professionals tend to ascribe higher scores on self-report scales to higher degrees of psychopathology (or even: to more typical or more authentic psychopathology). This practice may be significantly flawed. As can be seen from the present results, naïve simulators appear to be prone to overreporting symptoms and exaggerating neurocognitive impairment in a particularly strong way, and symptom information does not help them to reduce the amount of exaggeration. Significant and sometimes even extreme elevations across assessment measures are known to occur in some groups of real-world claimants as well as in experimental malingerers (e.g., Merten et al., 2007; Peters et al., 2006; Tolin et al., 1996; for a more general discussion on the limitations of symptom report measures: Williamson, 2007). Expected compensation appears to constitute a main factor for symptom overreporting (Tolin et al., 2004). Thus, contrary to conventional wisdom, reported extreme psychopathology (which often does not correspond to the clinical presentation and everyday functioning of forensic patients) may rather be a signal for negative response bias than for psychopathology itself. This view, however, poses the serious question of the validity of symptom report scales in any context where secondary or primary gain is an issue.

Moyer et al. (2002) found in another experimental malingering study that participants who tried to fake PTSD after a severe motor vehicle accident were not able to benefit from knowledge of PTSD to produce more accurate symptom profiles. The most important result of that study was that subjects who had additional information about PTSD symptoms were more likely to be detected as faking using the MMPI-2 *F* scale. A general trend in the coaching literature appears to indicate that symptom information alone does not improve the ability of patients to convincingly malinger psychological or cognitive impairment (e.g., Dunn et al., 1998; Gorny & Merten, 2005; Wetter et al., 1994).

Even in a regression-based formula developed by Maercker and Schützwohl (1998) to discriminate patients with PTSD from non-PTSD patients, 78 percent of the experimental malingerers were wrongly classified as PTSD patients, with the highest rate for naïve malingerers. This also demonstrates that symptom report is highly vulnerable to false claims of psychopathology and that even the use of symptom profiles does not prevent false-positive patient identifications. In a similar vein, McGuire (2002) has found the IES to be vulnerable to symptom magnification and manipulation. In his famous report on the Aleutian Enterprise sinking, Rosen (1995) illustrated the ill effect of symptom information on practical psychological assessment by a psychologist's statement: "Given that the symptoms were almost writing themselves out of the book, it didn't seem as if I had a difficult diagnostic picture in front of me." (p. 84)

As became clear from both the demonstrated reduced neuropsychological abilities as well as from SIMS response patterns, there was a general trend for symptom overgeneralization. Symptom demonstration as well as symptom report was not restricted to PTSD symptoms only.

The problem of malingered psychopathology can be solved with the use of symptom validity assessment methods. It has often been demonstrated that using modern cognitive and psychological SVTs can contribute to identifying a large percentage of patients with fabricated or exaggerated symptomatology, and multi-method approaches appear to be particularly efficient (cf. Bush et al., 2005; Merten et al., 2009). In the present study, a high number of experimental malingerers were correctly identified using such an approach. SVTs appear to show differential value in this respect, but sensitivity to detect feigning may, of course, vary according to the specific target symptomatology (e.g., faking specific mental or cognitive symptoms such as anxiety, depression or memory impairment vs. overgeneralizing symptom claims) and to individual strategies of false symptom report. In the current study, the SIMS appeared to be particularly useful in identifying false PTSD claims, but WMT and MENT were also able to detect more than three quarters of the total group. While the development of sophisticated methods to reveal negative response bias has led to claimants being increasingly better prepared for these methods by their attorneys, mere symptom information was not very useful. The combination of information and warning may significantly enhance the ability of claimants to avoid being detected. However, with the limited group size investigated here, this general trend in the results was only significant for the MENT.

These results point to a dilemma which is present in real forensic evaluations. Apart from being coached by third parties, patients are increasingly able to find information about assessment techniques, using modern media (Bauer & McCaffrey, 2006). Knowledge about SVTs potentially threatens the effectiveness of instruments and assessment procedure. Consequently, a good test should be resistant against coaching. Hartman (2002) proposed this as one quality marker for SVTs, and Morel and Marshman (2008) have judged it to be the most difficult to quantify and the most difficult of Hartman's criteria to meet.

The present study sought to follow major methodological demands of modern coaching studies (such as pre- and post-experimental manipulation checks, sample homogeneity, positive incentive for successful feigning), yet some limitations are inherent in that kind of research. A major problem for any analog study is the potential lack of external validity. The participants as well as the context of analog studies will always be different from those found in real-world forensic evaluations. Participants who are asked to invent or exaggerate symptoms in an experiment, even when playing a role as part of a well-described scenario involving a financial claim, do not actually have such a claim and, therefore, might not behave exactly as actual claimants do. Forensic patients who exaggerate symptom report could, in principle, endorse more

or less symptoms than the exaggerators in a simulation study. In contrast, internal validity of analog studies is usually high because they allow for a thorough control of experimental conditions. Consequently, Rogers (1997) suggested that both analog studies and known-groups designs should be performed in order to validate tests or procedures for use in forensic contexts.

Results which concern the Fake Bad Scale have to be dealt with cautiously because the items were not given in the context of the full MMPI-2 but dispersed in a set of other questionnaire items. For reasons of limits to the total time available for the experiment, it was not possible to give the full MMPI-2. While this change in item context is supposed not to change scale validity dramatically, this assumption has not been checked empirically, so caution is warranted.

Limitations also arise from the fact that no honest-response, full-effort control group was investigated. This limitation had beforehand been accepted by the authors in order not to decrease the sample size of the subgroups and also not to weaken statistical power by increasing the number of post-hoc simultaneous group comparisons. Rather, it was found that the behaviour of full-effort controls in the instruments used had been sufficiently investigated in previous studies (e.g., Cima et al., 2003; Merten et al., 2004; Morel, 1998a; Tydecks et al., 2006). Also, the present study aimed at investigating the influence of different instructional sets, not at distinguishing experimental simulators from full-effort controls. A more serious limitation may be seen in the fact that no patient control group was studied. When doing this kind of research with real patients with PTSD diagnosis, we are confronted with a number of challenges of a particular kind. Typical samples of PTSD patients (as they have been studied in the research literature so far) comprise an unknown number of patients with secondary gain who may be malingering or exaggerating psychopathology, and a careful check of negative response bias is usually not demanded in clinical studies until now. In this vein, we must assume that scientific data on the frequency and sequelae of traumatic life events is biased to an unknown extent (Rosen, 2004a; Rubenzer, 2005).

The present results may add to the ongoing debate about PTSD as a disorder that can easily be feigned. The widespread availability of information about PTSD in the public sphere (including the Internet) as well as a the uncritical use of symptom report scales without checking for symptom validity may add to a high number of false diagnoses. While the DSM-IV (American Psychiatric Association, 1995) states that malingering must be excluded when making the PTSD diagnosis, this demand is often neglected in clinical work as well as in the forensic arena. When modern approaches to the assessment of symptom validity are used, a substantial number of false claims can be identified by the forensic or clinical expert. Taken together and beyond the data presented here, the recommendations to be made from previous research can be summarized in the following way: Experts should be aware that PTSD may currently be grossly overdiagnosed. The diagnosis of PTSD must never be based on questionnaire data alone. While questionnaires can provide valuable additional information, they can and must never replace thorough clinical evaluation because they are prone to over- or underreporting. Thus, questionnaire results per se are not evidence for the presence or absence of symptoms. Without a thorough review of the credibility of such self-reports by empirically validated control scales or symptom validity measures the value of questionnaire results is highly limited. PTSD can easily be malingered. It takes an experienced examiner and a thorough review of all available information (including a clinical evaluation by the examiner himself) to check the plausibility of the diagnosis.

Disclosure of Interests

The study did not receive third party funding. The first author has participated in the development of the German versions of the Word Memory Test and the Morel Emotional Numbing Test, with no financial benefit. He has also arranged German adaptations of other symptom validity tests.

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fourth Edition. DSM-IV. International version with ICD-10 codes. Washington, DC: Author 2005

Arbisi PA, Ben-Porath YS. An MMPI-2 infrequency scale for use with psychopathological populations: The Infrequency–Psychopathology scale, *F(p)*. Psychol Assess 1995;7:424-431

Bauer L, McCaffrey RJ. Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? Arch Clin Neuropsychol 2006;21:121–126

Blaskewitz N, Gorny I, Merten T, Kathmann N. Anleitung zum nicht-authentischen Testverhalten: Coaching-Praktiken und ihre Folgen [Instruction for non-authentic test behaviour: Test coaching and its consequences]. Praxis Rechtspsychol 2007;17:108-124

Bush SS, Ruff RM, Tröster AI, Barth JT, Koffler SP, Pliskin NH, Reynolds CR, Silver CH. Symptom validity assessment: Practice issues and medical necessity. Arch Clin Neuropsychol 2005;20:419-426

Cima M, Hollnack S, Kremer K, Knauer E, Schellbach-Matties R, Klein B, Merckelbach H. "Strukturierter Fragebogen Simulierter Symptome". Die deutsche Version des "Structured Inventory of Malingered Symptomatology: SIMS" [The German version of the Structured Inventory of Malingered Symptomatology SIMS]. Nervenarzt 2003;74:977-986

Dunn TM, Shear PK, Howe S, Ris MD. Detecting neuropsychological malingering: Effects of coaching and information. Arch Clin Neuropsychol 2003;13:121-134

Essig S M, Mittenberg W, Petersen RS, Strauman S, Cooper JT. Practices in forensic neuropsychology: Perspectives of neuropsychologists and trial attorneys. Arch Clin Neuropsychol 2001;16:271-291

Freeman T, Powell M, Kimbrell T. Measuring symptom exaggeration in veterans with chronic posttraumatic stress disorder. Psychiatr Res 2008;158:374-380

Frueh BC, Hamner MB, Cahill SP, Gold PB, Hamlin KL. Apparent symptom overreporting in combat veterans evaluated for PTSD. Clin Psychol Rev 2000;20:853-885

Gervais RO, Ben-Porath YS, Wygant DB, Green P. Development and validation of a Response Bias Scale (RBS) for the MMPI-2. Assess 2007;14:196-208

Gorny I, Merten T. Symptom information – warning – coaching: How do they affect successful feigning in neuropsychological assessment? J Forens Neuropsychol 2005;4(4):71-97

Green P. Green's Word Memory Test for Windows: User's Manual. Edmonton, Canada: Green's Publishing 2003

Greiffenstein MF, Baker WJ, Axelrod B, Peck EA, Gervais R. The Fake Bad Scale and MMPI-2 *F*-family in detection of implausible psychological trauma claims. Clin Neuropsychol 2004;18:573–590

Greiffenstein MF, Baker WJ, Gola T. Validation of malingered amnesia measures with a large clinical sample. Psychol Assess 1994;6:218-224

Härting C, Markowitsch HJ, Neufeld H, Calabrese P, Deisinger K, Kessler J (eds). WMS-R. Wechsler Gedächtnistest – Revidierte Fassung [Wechsler Memory Scale–Revised]. Bern, Switzerland: Huber 2000

Hartman DE. The unexamined lie is a lie worth fibbing. Neuropsychological malingering and the Word Memory Test. Arch Clin Neuropsychol 2002;17:709-714

Hickling EJ, Blanchard EB, Mundy E, Galovski TE. Detection of malingered MVA related posttraumatic stress disorder: An investigation of the ability to detect professional actors by experienced clinicians, psychological tests, and psychophysiological assessment. J Forens Psychol Pract 2002;2:33–54

Larrabee GJ. Introduction: Malingering, research designs, and base rates. In Larrabee GJ (ed), Assessment of malingered neuropsychological deficits (pp. 3-13). Oxford: Oxford University Press 2007

Lees-Haley PR. Attorneys influence expert evidence in forensic psychological and neuropsychological cases. Assess 1997;4:321-324

Lees-Haley PR, English LT, Glenn WJ. A Fake Bad Scale on the MMPI-2 for personal injury claimants. Psychol Rep 1991;68:208-210

Maercker A, Schützwohl M. Erfassung von psychischen Belastungsfolgen: Die Impact of Event Skala – revidierte Version [Assessment of post-traumatic stress reactions: The Impact of Event Scale–Revised]. Diagn 1998;44:130-141

Martin TA, Hoffman NM, Donders J. Clinical utility of the Trail Making Test ratio score. Appl Neuropsychol 2003;10:163-169

McGuire BE. Malingered post-traumatic stress symptoms on the Impact of Event Scale. Legal Criminol Psychol 2002;7:165-171

McNally RJ. Progress and controversy in the study of posttraumatic stress disorder. Ann Rev Psychol 2003;54:229-252

Merten T, Friedel E, Stevens A. Die Authentizität der Beschwerdenschilderung in der neurologisch-psychiatrischen Begutachtung: eine Untersuchung mit dem Strukturierten Fragebogen Simulierter Symptome [The authenticity of symptom report in independent medical examinations: An investigation with the Structured Inventory of Malingered Symptomatology]. Praxis Rechtspsychol 2007;17:140-154

Merten T, Green P, Henry M, Blaskewitz N, Brockhaus R. Analog validation of German-language symptom validity tests and the influence of coaching. Arch Clin Neuropsychol 2005;20:719-726

Merten T, Henry M, Hilsabeck R. Symptomvalidierungstests in der neuropsychologischen Diagnostik: eine Analogstudie [The use of symptom validity tests in neuropsychological assessment: An analogue study]. Z Neuropsychol 2004;15:81-90

Merten T, Thies E, Schneider K, Stevens A. 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. Psychol Inj Law 2009;2:284-293

Mittenberg W, Patton C, Canyock EM, Condit DC. Base rates of malingering and symptom exaggeration. J Clin Exp Psychol 2002;24:1094-1102

Morel KR. Manual for the Morel Emotional Numbing Test for posttraumatic stress disorder. Author 1998a

Morel KR. Development and preliminary validation of a forced-choice test of response bias for posttraumatic stress disorder. J Personal Assess 1998b;70:299-314

Morel KR, Marshman KC. Critiquing symptom validity tests for posttraumatic stress disorder: A modification of Hartman's criteria. J Anx Disorders 2008;22:1542-1550

Morel KR, Shepherd BE. Meta-Analysis of the Morel Emotional Numbing Test for PTSD: Comment on Singh, Avasthi, and Grover. Germ J Psychiatr 2008;11:128-131

Moyer DM, Burkhardt B, Gordon RM. Faking PTSD from a motor vehicle accident on the MMPI-2. Am J Forens Psychol 2002;20:81-89

Peters MJV, Jelicic M, Op Heij CDH, Merckelbach H. Detection of malingered burnout using standard clinical questionnaires. Germ J Psychiatr 2006;9:10-16

Quezada-Ortega MR, Razo-Mondragón JLP, Marín-Cotoñieto IA, Salinas-Tovar S, López-Rojas P. Simulación en trabajadores que solicitan pensión por invalidez laboral [Malingering in workers with compensation claims for work-related disability]. Gaceta Méd Méx 2006;142:109-112

Raven JC. Standard Progressive Matrices, Sets A, B, C, D, E. London: Lewis 1956

Reitan RM. Trail Making Test. Manual for Administration and Scoring. South Tucson, AZ: Reitan Neuropsychology Laboratory 1992

Richman J, Green P, Gervais R, Flaro L, Merten T, Brockhaus R, Ranks D. Objective tests of symptom exaggeration in independent medical examinations. J Occup Env Med 2006;48:303-311

Rogers R. Researching dissimulation. In Rogers R (ed), Clinical assessment of malingering and deception.

Second edition (pp. 398–426). New York: Guilford Press 1997

Rosen GM. The Aleutian Enterprise sinking and posttraumatic stress disorder: Misdiagnosis in clinical and forensic settings. Profess Psychol Res Pract 1995;26:82-87

Rosen GM (ed), Posttraumatic stress disorder: Issues and controversies. Chichester, England: Wiley 2004a

Rosen GM. Litigation and reported rates of posttraumatic stress disorder. Personal Indiv Diff 2004b;36:1291-1294

Rosen GM. DSM's cautionary guideline to rule out malingering can protect the PTSD data base. J Anx Disorders 2006;20:530-535

Rubenzer S. Malingering psychiatric disorders and cognitive impairment in personal injury settings. For the Defense 2005;April:4-6

Rubenzer S. Posttraumatic stress disorder: Assessing response style and malingering. Psychol Inj Law 2009;2:114-142

Ruffulo LF, Guilmette TJ, Willis WG. Comparison of time and error rates on the Trail Making Test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. Clin Neuropsychol 2000;14:223-230

Singh J, Avasthi A, Grover S. Malingering of psychiatric disorders: A review. Germ J Psychiatr 2007;10:126-132

Smith GP, Burger GK. Detection of malingering: Validation of the Structured Interview of Malingered Symptomatology (SIMS). J Am Acad Psychiatr Law 1997;25:183-189

Stevens A, Friedel E, Mehren G, Merten T. Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. Psychiatr Res 2008;157:191-200

Suhr JA, Gunstad J. Coaching and malingering: A review. In Larrabee GJ (ed), Assessment of malingered neuropsychological deficits (pp. 287-311). Oxford: Oxford University Press 2007

Taylor S, Frueh BC, Asmundson GJG. Detection and management of malingering in people presenting for treatment of posttraumatic stress disorder: Methods, obstacles, and recommendations. J Anx Disorders 2007;21:22-41

Tolin DF, Maltby N, Weathers FW, Litz BT, Knight J, Keane TM. The use of the MMPI-2 Infrequency–Psychopathology scale in the assessment of posttraumatic stress disorder in male veterans. J Psychopathol Behav Assess 2004;26:23-29

Tombaugh TN. Trail Making Test A and B: Normative data stratified by age and education. Arch Clin Neuropsychol 2004;19:203-214

Tydecks S, Merten T, Gubbay J. The Word Memory Test and the One-in-Five-Test in an analogue study with Russian speaking participants. Internat J Forens Neuropsychol 2006:1(3):29-37

Wallis BJ, Bogduk N. Faking a profile: Can naive subjects simulate whiplash responses? Pain 1996;66:223-227

Wechsler D. Wechsler Adult Intelligence Scale–Third Edition. San Antonio, TX: The Psychological Corporation 1997

Weiss DS, Marmar CR. The Impact of Event Scale–Revised. In Wilson JP, Keane TM (eds), Assessing psychological trauma and PTSD (pp. 399-411). New York: Guilford 1996

Wetter MW, Baer RA, Berry DTR, Reynolds SK. The effect of symptom information on faking in the MMPI-2. Assess 1994;1:199-207

Wetter MW, Corrigan SK. Providing information to clients about psychological tests: A survey of attorneys' and law students' attitudes. Profess Psychol Res Pract 1995;26:474-477

Widows MR, Smith GP. SIMS – Structured Inventory of Malingered Symptomatology. Professional Manual. Lutz, FL: Psychological Assessment Resources 2005

Williamson A. Using self-report measures in neurobehavioural toxicology: Can they be trusted? NeuroToxicol 2007;28:227–234