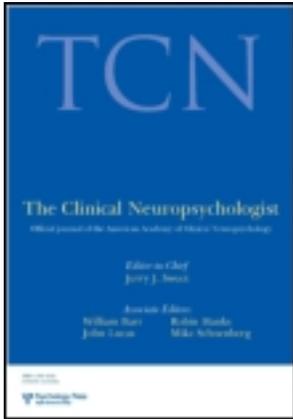


This article was downloaded by: [University of Maastricht]

On: 17 August 2011, At: 12:41

Publisher: Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



The Clinical Neuropsychologist

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/ntcn20>

Cognitive Underperformance and Symptom Over-Reporting in a Mixed Psychiatric Sample

Brechje Dandachi-FitzGerald^a, Rudolf W. H. M. Ponds^a, Maarten J. V. Peters^b & Harald Merckelbach^b

^a Psy-Q Mental Health Care Institute, Maastricht & School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

^b Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands

Available online: 15 Jul 2011

To cite this article: Brechje Dandachi-FitzGerald, Rudolf W. H. M. Ponds, Maarten J. V. Peters & Harald Merckelbach (2011): Cognitive Underperformance and Symptom Over-Reporting in a Mixed Psychiatric Sample, *The Clinical Neuropsychologist*, 25:5, 812-828

To link to this article: <http://dx.doi.org/10.1080/13854046.2011.583280>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Cognitive Underperformance and Symptom Over-Reporting in a Mixed Psychiatric Sample

Brechje Dandachi-FitzGerald¹, Rudolf W. H. M. Ponds¹, Maarten J. V. Peters², and Harald Merckelbach²

¹Psy-Q Mental Health Care Institute, Maastricht & School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

²Forensic Psychology Section, Maastricht University, Maastricht, The Netherlands

The current study examined the prevalence of cognitive underperformance and symptom over-reporting in a mixed sample of psychiatric patients ($N=183$). We employed the Amsterdam Short-Term Memory Test (ASTM) to measure cognitive underperformance and the Structured Inventory of Malingered Symptomatology (SIMS) to measure the tendency to over-report symptoms. We also administered neuropsychological tests (e.g., Concept Shifting Task; Rey's Verbal Learning Test) and the Symptom Checklist-90 (SCL-90) to the patients. A total of 34% of them failed the ASTM, the SIMS or both tests. ASTM and SIMS scores were significantly, albeit modestly, correlated with each other ($r=-.22$). As to the links between underperformance, over-reporting, neuropsychological tasks, and the SCL-90, the association between over-reporting on the SIMS and SCL-90 scores was the most robust one. The subsample that only failed on the ASTM performed significantly worse on a compound index of memory performance. Our findings indicate that underperformance and over-reporting are loosely coupled dimensions and that particularly over-reporting is intimately linked to heightened SCL-90 scores.

Keywords: Underperformance; Over-reporting; Symptom validity tests; Amsterdam Short-Term Memory Test; Structured Inventory of Malingered Symptomatology.

INTRODUCTION

In forensic assessments underperformance on neuropsychological tests and over-reporting of psychiatric symptoms may lead to a more favorable outcome for the examinee (e.g., higher financial compensation or diminished criminal responsibility). With this in mind, experts have developed several methods for assessing effort level and response style in the past two decades (Rogers, 2008). These symptom validity tools are intended to assist clinicians in differentiating between genuine patients and persons deliberately fabricating or exaggerating symptoms in order to obtain an external incentive (i.e., malingering; see *Diagnostic and Statistical Manual of Mental Disorders*; DSM-IV-TR; American Psychiatric Association, 2000). Some of these tools measure underperformance on neuropsychological tests, whereas others measure the tendency to over-report symptoms (see, for examples, Berry, Baer, Rinaldo, & Wetter, 2002; Green, 2007a; Nitch & Glassmire, 2007). In the clinical literature underperformance and symptom over-reporting are often

Address correspondence to: Brechje Dandachi-FitzGerald, School for Mental Health and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.
E-mail: b.fitzgerald@maastrichtuniversity.nl

Accepted for publication: April 8, 2011. First published online: June 27, 2011

conceptualized as behavioral proxies of malingering, with terms like underperformance, poor effort, symptom exaggeration, response bias, and malingering being used interchangeably. However, as Iverson (2006, p. 81) pointed out, “effort tests do not measure malingering, *per se*; they measure behavior associated with malingering”. In the strict sense of the word, malingering refers to the intentional creation of symptoms (American Psychiatric Association, 2000). Because it is difficult, if not impossible, to measure intention, we prefer to use the behavioral terms “underperformance” and “symptom over-reporting” (Berry & Nelson, 2010; Boone, 2007).

Prevalence estimates of cognitive underperformance and symptom over-reporting in forensic settings vary widely, depending on, for example, the symptom validity tests used, the research sample, and the referral question, with most estimates ranging between 20% and 60% (e.g., Hout, Schmand, Weking, & Deelman, 2006; Mittenberg, Patton, Canyock, & Condit, 2002; Schmand et al., 1998). It should be noted, however, that these prevalence rates are rough estimates at best, because they often rely on single diagnostic tools administered to artificially homogenous samples. Nevertheless, the bottom line of much research on symptom validity in forensic settings is that underperformance explains more variance in cognitive (e.g., memory) test performance than, for example, severity of the brain injury or depressive complaints (Green, 2007b; Rohling, Green, Allen, & Iverson, 2002; Stevens, Friedel, Mehren, & Merten, 2008).

In the past 10 years there has been an increased interest in symptom validity and how it relates to routine neuropsychological testing scores in non-litigant settings. Clinical studies in this domain have focused on patients with epilepsy (Cragar, Berry, Fakhoury, Cibula, & Schmitt, 2006; Dodrill, 2008), medically unexplained symptoms (Kemp, Coughlan, Rowbottom, Wilkinson, Teggart, & Baker, 2008), acquired brain damage (Locke, Smigielski, Powell, & Stevens, 2008), and students referred for ADHD evaluation (Suhr, Hammers, Dobbins-Buckland, Zimak, & Hughes, 2008; Sullivan, May, & Galbally, 2007). In these studies the percentage of patients exhibiting underperformance was found to be in the order of 20–30%. However, research addressing this issue in psychiatric samples is still limited. To the best of our knowledge only one study (Gorrissen, Sanz, & Schmand, 2005) examined underperformance in psychiatric patients and how it affected neuropsychological test scores. The authors of that study demonstrated that 25% of their non-psychotic (mainly affective disordered) patients and 72% of their schizophrenic patients failed the Word Memory Test (WMT; Green, 2003), which is a widely used index of underperformance. Most patients in this study had been referred for routine neuropsychological assessment. Failure on the WMT explained up to 35% of the variance in the neuropsychological test scores, a proportion that accords very well with that reported by Stevens et al. (2008) for their forensic sample. Thus the prevalence estimates of underperformance that have been reported in the forensic domain may also hold true for the clinical, non-litigant setting. This would imply that in both settings, underperformance is a non-trivial phenomenon and *when* it occurs, it affects standard clinical test scores to a considerable degree.

An important limitation of the research that addressed symptom validity in a clinical, non-litigant setting is that most of the studies relied on instruments measuring underperformance. So far only a few studies have looked at both the

prevalence of cognitive underperformance and symptom over-reporting in clinical samples. Haggerty, Frazier, Busch, and Naugle (2007) examined cognitive underperformance and symptom over-reporting in a sample of neurological and psychiatric patients. The authors reported significant but modest correlations (i.e., $r_s < .20$) between scores on the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, & Thompson, 1996) indexing underperformance and scores on the Negative Impression Management (NIM) subscale of the Personality Assessment Inventory (PAI; Morey, 1991) indexing symptom over-reporting. Similarly, Whiteside, Dunbar-Mayer, and Waters (2009) found modest correlations in an outpatient sample ($r_s < .39$) between the NIM subscale of the PAI and the Test of Memory Malingering (TOMM; Tombaugh, 1996) as a measure of underperformance. Importantly, these two studies did not address whether failing on symptom validity tests is related to test performance on routine tasks and scores on clinical self-report instruments.

In sum, the studies conducted so far concur that indices of underperformance and over-reporting are only modestly correlated (Stevens et al., 2008; Ruocco et al., 2008; Nelson, Sweet, Berry, Bryant, & Granacher, 2007). Yet none of the studies cited above looked at how underperformance and over-reporting are related to scores on routine clinical tests. With these considerations in mind, the aim of this study was twofold. First, we examined the base rate of underperformance and over-reporting of symptoms in a heterogeneous sample of psychiatric patients who had been referred for neuropsychological assessment. Second, we explored to what extent underperformance and symptom over-reporting in this sample predict performance on standard neuropsychological tests and scores on a widely used inventory of self-reported symptoms.

METHOD

Participants

All participants were patients referred for neuropsychological assessment at Psy-Q, a mental health care institute in Maastricht, the Netherlands. Referrals were received from clinicians or general practitioners. Neuropsychological assessments were conducted for treatment purposes and none of the patients was referred as part of a forensic evaluation. Data were collected between January 2007 and November 2009. Only patients with complete data on the symptom validity tests (see below) were included in the analyses. Ten patients were excluded because of clinically obvious cognitive impairment. In the case of one patient, psychotic symptoms interfered with testing. We excluded four patients who had evident cognitive impairment as a result of chronic alcohol abuse. Two of them were diagnosed with Korsakoff syndrome. Furthermore, we excluded five patients diagnosed with contusion cerebri and/or cerebrovascular accidents. These five patients exhibited serious cognitive impairments and psychiatric comorbidity.

The final sample consisted of 183 patients (122 men). Their mean age was 34.7 years ($SD = 12.52$ range: 17–66). Educational background was quantified with an 8-point scale that is commonly used in the Netherlands for this purpose (De Bie, 1987) and that ranges from primary school (1; fewer than 6 years of education) to

university degree (8; 16 years of education or more). The median educational level was 4, which corresponds to medium vocational training. All participants were literate and had (corrected to) normal vision and hearing. Nearly all patients (approximately 9 out of 10 patients) were outpatients at the time of the assessment. A total of 51 patients (28%) were referred for general assessment of cognitive abilities, 79 patients (43%) were referred for a possible diagnosis of Attention Deficit Hyperactivity Disorder (ADHD), 52 patients (28%) for a possible diagnosis of Autism Spectrum Disorder (ASD), and two patients (1%) for a combined diagnosis of possible ADHD and ASD.

Procedure

Patients completed an informed consent form. Next they completed a neuropsychological test battery (see below). Depending on the referral question there was some variation in the tests administered, but the majority of patients received the same core test battery. The tests were administered by a certified psychological assistant or by trained clinical psychology master students. They were supervised by a clinical neuropsychologist. Before the data were entered in the SPSS database, the first author checked the record of each patient for missing values and outliers.

Symptom validity measures

We used the Amsterdam Short-Term Memory test (ASTM; Schmand, Sterke, & Lindeboom, 1999) as a measure of underperformance. The ASTM is presented as a memory test and basically involves a forced choice word recognition procedure. Please refer to the ASTM manual for additional details regarding test procedure and materials (Schmand & Lindeboom, 2005). Following Schmand et al. (1999), we used a cut-off score of 85. In the original validation studies, a cut-off of 85 best distinguished between experimental simulators ($N=57$) and aggregated groups of patients suffering from neurological disorders such as contusion cerebri, multiple sclerosis, and severe epilepsy ($N=57$), with a sensitivity and specificity of 84% and 90%, respectively. Higher sensitivity and specificity rates (both >90%) were obtained in experimental simulation studies with healthy simulators and controls (e.g., Bolan, Foster, & Bolan, 2002). Validation studies further showed that the test is not suitable for patients with clinically evident cognitive impairment as in dementia and Korsakoff syndrome. A recent study of Rienstra, Spaan, and Schmand (2010) demonstrated that children who were 9 years and older all passed the ASTM. In the current study we used a fixed sequence of tests with the ASTM being positioned at the beginning of the test battery.

We employed a Dutch research version of the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997; see also Merckelbach & Smith, 2003) as a measure of symptom over-reporting. The SIMS is a self-report scale consisting of 75 yes–no items. As with the ASTM, we refer to the manual for more detailed information about this scale (Widows & Smith, 2005). Following the recommendations of Rogers, Hinds, and Sewell (1996), in the current study we used a cut-off score of 16. A study of the Dutch research version of the SIMS

(Merckelbach & Smith, 2003) with a group of 298 participants, revealed a specificity of 98% and sensitivity of 93% with the cut-off set at 16. Follow-up studies indicated that the SIMS attains high sensitivity rates ($\geq 80\%$; e.g., Clegg, Fremouw, & Mogge, 2009), even in coached simulators (Jelicic, Hessels, & Merckelbach, 2006). In the current study the SIMS was positioned at the end of the test battery.

Clinical measures

In 74 patients intelligence was measured with a validated Dutch language version of the Wechsler Adult Intelligence Scale III (WAIS III; Wechsler, 1997), while in 106 patients the short form of the Groninger Intelligence Test-2 was employed (Luteijn & Barelds, 2004). In three patients, no intelligence test was administered.

As an index of processing speed we used the a version of Stroop Color Word Test most commonly used in the Netherlands (SCWT; Stroop, 1935; Van der Elst, van Boxtel, van Breukelen, & Jolles, 2006a). It consists of three stimulus cards: color word naming (I); color naming (II); and naming of color words printed in a different color (inference task III). The time needed to complete each card is scored.

The Concept Shifting Task (CST) was used to measure processing speed and cognitive flexibility (Van der Elst, van Boxtel, van Breukelen, & Jolles, 2006b). It consists of three subtasks (A, B, and C). On each test sheet 16 small circles are grouped into one larger circle. In the smaller circles the test items appear in a fixed random order. In subtask A patients are asked to cross out numbers (1–16) in the right order as quickly and accurately as possible. In part B, the circles contain letters (A–P) that have to be crossed out in alphabetical order. In the third part the card displays both numbers and letters, and patients are requested to alternate between numbers and letters. The time needed to complete each card is scored.

Rey's Verbal Learning Test (VLT) was used to evaluate learning and retrieval capacity (Van der Elst, van Boxtel, van Breukelen, & Jolles, 2005). In the VLT 15 words are presented in a fixed order on a computer screen, one after another, in five trials. After each trial the patient is asked to reproduce the words (immediate recall). Then 20 minutes after the last trial the patient is asked again to reproduce the words (delayed recall). Following the delayed recall a list of 30 words is presented in a fixed order on a computer screen, one after another. Patients have to indicate whether or not the presented word was on the learning list (recognition). Dependent variables were the total number of words recalled on the immediate and delayed recall and the number of correctly recognized items on the recognition trial.

A subsample of 112 patients completed the Dutch version of the Symptom Checklist-90 (SCL-90; Arrindell & Ettema, 2003). The SCL-90 is a 90-item self-report measure that covers a broad range of psychological symptoms (e.g., anxiety, depression). Patients rate on 5-point scales (0 = not at all; 4 = extremely) how much they have experienced these symptoms during the last week. For the purpose of the present study a total raw score was calculated by summing across item.

Data reduction and analysis

Raw test scores were screened for outliers. We decided not to exclude patients with extreme outliers. By excluding them we would potentially obscure the relationship between symptom validity measures and deviant responding on our clinical measures, since those patients who exhibit poor symptom validity might be inclined to produce outliers. With this in mind we employed another widely used method of handling outliers, namely replacement of outliers by the sample mean plus two standard deviations (e.g., Field, 2005). Less than 8% of the patients had an extreme outlier in their test protocol.¹ Raw test scores for related tasks were clustered to yield compound performance indices for three domains: memory, executive control, and speed of information processing (e.g., Stevens et al., 2008; Van Boxtel, Langerak, Houx & Jolles, 1996). This was done so as to reduce the number of dependent variables while improving the robustness of the underlying cognitive construct. Next, test scores were transformed to Z scores. Following this, the mean of the Z-transformed scores that were included in a compound performance index was calculated for each patient. Thus a memory score was derived from the Z-transformed immediate recall (IR), delayed recall (DR), and recognition (RC) scores of the VLT. An executive control score was derived from the Z-transformed scores of the SCWT III and CST C. Speed of information processing was calculated from Z-transformed scores of SCWT I and II and CST A and B. The sign of the speed scores was inverted such that for all compound scores positive values denote above average performance and negative values denote below average performance.² IQ (population-based, age-corrected normative score) and SCL-90 (raw score) test results were treated as separate variables.

In the first step of the analysis we determined the base rate of failing the ASTM and SIMS using the recommended cut-off scores (ASTM < 85 and SIMS > 16). Second, Pearson product-moment correlations were calculated between ASTM and SIMS, and also between these two symptom validity tests and age and educational level. As a third step, to explore a potential interaction effect between underperformance and over-reporting, we performed an omnibus repeated-measures $2 \times 2 \times 5$ Analysis of Variance (ANOVA), with underperformance and over-reporting as between-participant variables and clinical tests as repeated measures. Fourth, we formed four groups of patients based on their scores on the ASTM and SIMS: those who passed both tests; those who only failed the ASTM (i.e., underperformance); those who only failed the SIMS (i.e., over-reporting); and those who failed both tests (i.e., underperformance and over-reporting). We performed one-way ANOVAs to examine whether the four groups differed with regard to IQ, compound scores of memory, executive control and speed of information processing, and SCL-90 scores. When there was a significant

¹A total of 12 patients produced an extreme outlier on only one of the measures (SCWT I: 3 outliers; SCWT II: 1 outlier; CST B: 2 outliers; Rey's VLT recognition: 6 outliers). Only one patient obtained three extreme outliers (SCWT I, CST A, and B).

²Thus, the following formulas were used:

Memory = $(ZVLT-IR + ZVLT-DR + ZVLT-RC)/3$;

Executive Control = $-(ZCST-C + ZSCWT-III)/2$;

Speed of information processing = $-(ZCST-A + ZCST-B + ZSCWT-I + ZSCWT-II)/4$.

Table 1 DSM-IV-TR diagnoses (percentages between parentheses) in the sample ($N=183$) and number of patients failing the ASTM and SIMS

	Primary diagnosis		Secondary diagnosis		Fail ASTM		Fail SIMS	
ADHD	56	(31%)	4	(2%)	8	(14%)	8	(14%)
ASD	25	(14%)	1	(1%)	4	(16%)	4	(16%)
Mood and anxiety disorders	34	(19%)	40	(22%)	8	(24%)	7	(21%)
Psychotic disorder	8	(4%)	1	(1%)	2	(25%)	3	(38%)
Cognitive disorder NOS	7	(4%)	8	(4%)	4	(57%)	4	(57%)
Personality Disorders	16	(9%)	3	(2%)	5	(31%)	6	(38%)
Abuse/Dependence of substances	11	(6%)	14	(8%)	2	(18%)	2	(18%)
Other psychiatric disorders	13	(7%)	18	(10%)	4	(31%)	3	(23%)
Postponed diagnosis	5	(3%)	4	(2%)	2	(40%)	1	(20%)
No diagnosis	8	(4%)	90	(49%)	0	(0%)	0	(0%)

ASTM=Amsterdam Short Term Memory test; SIMS=Structured Inventory of Malingered Symptomatology; ADHD=Attention Deficit Hyperactivity Disorder; ASD=Autism Spectrum Disorder; Cognitive Disorder NOS=Cognitive Disorder Not Otherwise Specified.

group difference, we performed post-hoc Bonferroni tests. Cases with missing data were excluded. Analyses were performed with SPSS version 15.0 for Windows. Alpha level was set at $p < .05$ (two-tailed).

RESULTS

Descriptive statistics

Table 1 depicts DSM-IV diagnoses in our sample and the number of patients failing the ASTM and SIMS within each diagnostic category. As can be seen, the most common primary diagnoses were ADHD, mood and anxiety disorders, and ASD. A total of 89 patients (49%) had a comorbid psychiatric diagnosis, the most common of which were mood and anxiety disorders and substance (in the majority of cases: cannabis) abuse or dependence.

Prevalence of underperformance and over-reporting

Table 2 shows the proportions of patients who passed both tests, patients who failed the ASTM (i.e., underperformance), patients who failed the SIMS (i.e., over-reporting), and patients who failed both tests. As can be seen, 62 patients (33.9%) of the sample failed one or both tests. The proportions of patients who failed one or both symptom validity test(s) did not differ: $\chi^2(2) = 2.46$, $p = .29$. Table 1 shows these proportions for the separate diagnostic categories. Because developmental disorders in adulthood (notably ADHD and ASD) were over-represented in our sample, we performed separate analyses for these categories and the other diagnostic groups. ADHD or ASD patients passed the SIMS, ASTM, or both measures somewhat more often than patients in other diagnostic categories, a difference that attained borderline significance, $\chi^2(3) = 7.50$, $p = .06$, $\phi = 0.21$.

Table 2 Number of patients (percentages in parentheses) who passed or failed the ASTM and the SIMS

		ASTM			
		Pass		Fail	
SIMS	Pass	121	(66%)	24	(13%)
	Fail	23	(13%)	15	(8%)

N = 183. ASTM = Amsterdam Short Term Memory test; SIMS = Structured Inventory of Malingered Symptomatology.

There were significant, but small correlations between ASTM and age ($r = -.18$, $p < .05$; $r^2 = .03$) and between ASTM and educational level ($r = .23$, $p < .05$; $r^2 = .05$), with lower scores on the ASTM being associated with higher age and lower education level. Also, there was a negative correlation between SIMS and educational level ($r = -.25$, $p < .05$; $r^2 = .06$), with higher scores on the SIMS being associated with lower educational level. The correlation between SIMS and age fell short of significance ($r = -.02$). The first-order Pearson product-moment correlation between ASTM and SIMS was $-.22$ ($p < .05$),³ with poorer performance on the ASTM being associated with higher scores on the SIMS. When controlling for age and educational level, ASTM remained significantly correlated with SIMS (partial $r = -.17$, $p < .05$), although the amount of explained variance was small ($r^2 = .03$).

Relation with clinical instruments

A repeated-measures ANOVA with underperformance and over-reporting as between-participants factors and clinical instruments as repeated variables, indicated that there was no significant interaction between underperformance and over-reporting, $F(1, 94) = .80$, $p = .77$. The main effect of underperformance remained non-significant, indicating that this factor did not have an overall impact on clinical test outcome, $F(1, 94) = .70$, $p = .41$. In contrast, the main effect of over-reporting was significant, $F(1, 94) = 45.52$, $p < .05$; $\eta_p^2 = .33$. Thus over-reporting as measured with the SIMS explained a substantial proportion of variance in clinical test results.

As follow-up analyses we carried out one-way ANOVAs with group (1 = pass both, 2 = fail only ASTM, 3 = fail only SIMS, 4 = fail both) as between-participants variable and age, education, IQ, compound indices (i.e., memory, executive control and speed of information processing), and SCL-90 scores as dependent variables. Table 3 summarizes the results of these ANOVAs. Group differences were significant at $p < .05$ for all dependent variables. Post-hoc Bonferroni corrected pairwise comparisons revealed that the group that only underperformed (as measured with the ASTM) was on average older and performed worse on the

³Because ASTM scores followed a skewed distribution while SIMS scores did not, we also computed the Spearman rho correlation between both measures: $\rho = -.18$. This value comes close to the Pearson r that we found.

Table 3 Mean scores (*SD*) on demographic and clinical outcome variables of the four groups

Measure	Number of patients	Group 1		Group 2		Group 3		Group 4		F	<i>p</i> -value	Eta squared	Significant group comparisons (Bonferroni corrected)
		Passed both	Failed both	Failed only ASTM	Failed only SIMS	Failed both	Failed both						
Age	183	33.21 (12.19)	42.83 (12.82)	34.52 (12.78)	34.20 (10.21)	4.16	.01	.07	2 > 1				
Education	183	4.44 (1.94)	3.88 (1.99)	3.70 (2.14)	3.13 (1.13)	2.90	.04	.05	ns				
IQ	180	95.60 (16.16)	92.08 (19.81)	84.83 (14.28)	78.86 (14.48)	6.36	.00	.10	3,4 < 1				
Compound score Memory	180	.19 (.80)	-.60 (1.00)	-.22 (.81)	-.25 (.85)	6.75	.00	.10	2 < 1				
Compound score Speed of information processing	161	.14 (.69)	-.21 (.90)	-.31 (.92)	-.46 (1.14)	4.00	.01	.07	ns				
Compound score Executive Control	112	1.46 (.83)	-.38 (1.11)	-.32 (.84)	-.23 (.79)	3.48	.02	.06	ns				
SCL-90 raw total score	112	159.72 (37.73)	186.13 (61.16)	262.85 (71.99)	271.89 (61.54)	28.37	.00	.44	3,4 < 1 3,4 < 2				

memory compound score than the group that passed both symptom validity measures. The group that only over-reported symptoms (as measured with the SIMS) and the group that underperformed and over-reported symptoms had significantly lower average IQ scores than the group that passed both symptom validity measures. Furthermore, these two groups reported significantly more psychological symptoms than the group that passed both tests and the group that only underperformed. In terms of effect size (eta squared), the link between over-reporting on the SIMS and heightened SCL-90 levels was the most substantial one.

As a next step we performed analysis of covariance to control for the potentially confounding effects of age and educational background in the relation between group membership and clinical test outcomes. When controlling for the effects of age and educational level, the group differences remained significant for IQ, the compound memory index, and SCL-90. However, group differences in speed of information processing, $F(3, 155)=2.04$, $p=.11$, and executive control, $F(3,155)=1.33$, $p=.27$, became non-significant.

As covariates both age and educational level were significantly related to IQ, $F(1, 174)=5.16$, $p < .05$, and $F(1, 174)=97.40$, $p < .05$, respectively. After adjusting for these demographic variables, group differences remained significant for IQ. $F(3, 174)=4.34$, $p < .05$. Both covariates were also significantly related to the compound memory score—age: $F(1, 174)=8.91$, $p < .05$; educational level: $F(1, 174)=10.63$, $p < .05$. However, group differences for the memory compound score remained significant after controlling for age and educational level, $F(3, 174)=3.31$, $p < .05$. Neither age, $F(1, 106)=.02$, $p=.88$, nor educational level, $F(1, 106)=.78$, $p=.38$, was significantly related to the SCL-90. Groups continued to differ significantly in their report of psychological symptoms after controlling for both covariates, $F(3, 106)=26.10$, $p < .05$.

DISCUSSION

The purpose of this study was to examine the base rate of underperformance on cognitive tests and symptom over-reporting in a mixed psychiatric, non-litigant sample. We also wanted to know how these phenomena are related to routine clinical measures. Almost 34% of the patients in our sample failed the ASTM, the SIMS or both tests. This suggests that even outside the forensic domain, underperformance on cognitive tests and/or symptom over-reporting are not unusual among psychiatric patients. Our finding that 21% of the patients in a standard clinical setting failed the ASTM (i.e., underperformed) concurs with the findings of previous studies in various clinical settings that 20–30% of the patients underperform on symptom validity tests during neuropsychological assessment (Gorissen et al., 2005; Kemp et al., 2008; Locke et al., 2008). Furthermore, our finding that almost 21% of the patients failed the SIMS (i.e., over-reported symptoms) is in accordance with other clinical studies. For example, using a taxometric analysis of the MMPI-2 infrequency scales, Strong, Greene, and Schinka (2000) found a base rate of 27% for symptom over-reporting in a large sample of psychiatric inpatients. Likewise, using the SIMS, Beilen, Griffioen, Gross, and

Leenders (2009) demonstrated that 23% of their patients with medically unexplained neurological complaints over-reported symptoms.

Although we found that SIMS and ASTM were significantly correlated, even when the influence of age and education was partialled out, this correlation was by all standards, small. Apparently both instruments measure related but different dimensions. The few studies that looked at both underperformance and over-reporting in clinical (Haggerty et al., 2007; Whiteside et al., 2009) and legal settings (Nelson et al., 2007; Ruocco et al., 2008; Stevens et al., 2008) came to a similar conclusion, although it should be added that different measures were used in these previous studies. Taken together, these findings provide strong support for Iverson's (2006) point that symptom over-reporting and underperformance on cognitive tests may occur independently of each other. The implication of this for clinical practice is that both dimensions should be addressed during neuropsychological testing. Our findings also illustrate that the two dimensions have domain-specific correlates when it comes to clinical test results. That is to say, failing the SIMS had its strongest link with heightened levels of symptom reporting on the SCL-90 (33% explained variance). On the other hand, failing the ASTM seemed to have more limited effects and was primarily, although modestly, related to poor memory performance.

Symptom validity measures, almost by definition, do not intend to measure cognitive abilities or variables strongly related to these abilities. Therefore we were surprised to find that lower scores on the ASTM were associated with lower educational levels and higher age. Although the original validation studies on the ASTM (Schagen, Schmand, de Sterke, & Lindeboom, 1997; Schmand et al., 1999) did not find a relationship with age or educational level, Stulemeijer, Andriessen, Brauer, Vos, and Van der Werf (2007) also reported that patients failing the ASTM had significantly lower educational levels. Their study relied on a homogeneous sample of patients with mild traumatic brain injury. Unlike our study, the authors did not observe a significant age difference between patients passing and failing the ASTM. Clinical studies relying on other instruments than the ASTM to measure underperformance yielded mixed results, with some studies finding small yet significant correlations with educational level and/or age (Haggerty et al., 2007; Kemp et al., 2008), whereas others found no significant associations (Gorrissen et al., 2005; Locke et al., 2008). This inconsistency may have to do with the homogeneity of the samples used. It may well be the case that only in samples that are mixed with regard to psychiatric background, educational level, and age, significant correlations between symptom validity measures and these demographic variables emerge. Small correlations ($\leq .20$) with demographical variables may not be unusual for instruments like the ASTM: even though they are simple, they require some minimum level of cognitive functioning (e.g., reading skills, working memory). The more important question is whether age and education correlate with the proportion of false positives on these instruments. This issue warrants future studies in which instruments like the ASTM are administered to large samples that are highly diverse with regard to age and education.

We found that patients failing only the ASTM performed worse on the compound memory score than the group passing both tests. After adjusting for age and education, this difference remained significant, although failing the ASTM

explained only a small proportion of the variance in memory scores (i.e., 5.4% after partialling out age and education). Thus, in contrast to previous studies, we were unable to show that a substantial overlap exists between ASTM failure and poor memory performance. This discrepancy might have to do with differential sensitivity of symptom validity instruments to cognitive impairment. For example, Merten, Bossink, and Schmand (2007) studied the lower-bound reliability limits of several instruments and concluded that the TOMM is more robust against genuine cognitive impairment (e.g., a result of dementia or other neurological conditions) than the ASTM. These authors also found that the ASTM was most strongly related to neuropsychological tests measuring working memory, a cognitive function that is often impaired in various neuropsychiatric conditions (Alloway & Gathercole, 2006). On the basis of the Merten et al. (2007) study, and the results of the current study, it appears that the ASTM is not the most suitable test for detecting underperformance in neuropsychiatric samples. The more general point here is, of course, that just because a symptom validity instrument may have proven its validity in, for example, a medico-legal setting or with a specific disorder (e.g., mild traumatic brain injury), this does not imply that it can also be administered with the same empirically established cut-off scores to other settings and disorders.

The most robust relationship in the current study was that between the SIMS and the SCL-90. Patients failing the SIMS reported more psychological symptoms than the group that only failed the ASTM and the group passing both tests. At first glance this suggests that the SIMS is correctly identifying those patients who are exaggerating their symptoms. There was, however, also a modest relationship between the SIMS and intelligence, such that patients failing the SIMS obtained a lower IQ score. Furthermore, higher scores on the SIMS were associated with lower educational levels. A plausible interpretation of this constellation is that less-intelligent and less-educated patients may have a preference for more blatant forms of symptom exaggeration, while more-intelligent patients might exhibit a more subtle style of symptom over-reporting (Solomon, Boone, Miora, & Skidmore, 2010). Another possibility is that persons with higher IQ are more likely to see through the rationale of the SIMS. However, there is little empirical support for this interpretation. For example, Jelacic et al. (2006) demonstrated that students who were instructed to simulate and received coaching still failed on the SIMS. A third possibility is that the links between SIMS scores, IQ, and education reflect cognitive abilities. Thus, less-intelligent persons may answer more items positively because they do not fully grasp the questions. Although the manual of the SIMS recognizes that "In some cases, a respondent may exhibit genuine cognitive incapacity, such as mental retardation, and be unable to complete the SIMS" (Widows & Smith, 2005, p. 5), it does not specify the minimum intelligence level for a reliable administration of the SIMS. This issue is important because Graue et al. (2007) did indeed find a reduced specificity of the SIMS in persons with mild mental retardation ($IQ \leq 70$). While the average IQ level in our sample was higher than 70, our findings highlight the need to further examine the relationship between SIMS and intelligence levels in order to establish appropriate cut-off scores to maintain adequate specificity.

A limitation of the current study is that we did not explore external incentives in our sample. Although all patients were seeking treatment, we cannot rule out the

possibility that some of them had external incentives. In a sample of Dutch psychiatric outpatients ($N=166$) Van Egmond, Kummeling, and Balkom (2005) found that 42% of their patients fostered expectations of gaining specific benefits from being in therapy other than getting better (e.g., they wanted help with obtaining a disability status, leave of absence from work, or a new accommodation). Interestingly, in most cases the clinician was unaware of these expectations. Furthermore, patients with these expectations showed less improvement during treatment than patients without these expectations. Whether such expectations may drive underperformance and over-reporting is an important issue that deserves systematic study. Meanwhile, we agree with Berry and Nelson (2010, p. 297) that it seems likely that “most behaviors are driven by both intrinsic and extrinsic factors” and that is notoriously difficult to discriminate between these factors.

Our finding that failing the ASTM or the SIMS is related to poor memory and heightened symptom levels, respectively, is silent about the causal direction that underlies these associations. It might be the case that intentional feigning is the driving force behind these associations, but another possibility is, of course, that genuine pathology interferes with performance and symptom reporting. Whatever the interpretation, these associations do show that clinicians cannot take clinical test outcomes at face value when patients fail symptom validity instruments.

In conclusion, our results strongly support the notion that underperformance and over-reporting can be viewed as “separate but related aspects of the broader construct of symptom exaggeration” (Haggerty et al., 2007, p. 926). They imply that a thorough assessment of symptom validity needs to take both dimensions into account, also during routine neuropsychological evaluation in a psychiatric setting. Particularly, the SIMS seems to be of value in evaluating the validity of symptom reporting. Our study also demonstrates that only establishing the base rates for failing symptom validity tests in various groups is not sufficient to understand the consequences and origins of underperformance and symptom over-reporting.

ACKNOWLEDGMENTS

We thank Anouk Peeters for her contribution to the collection of data.

REFERENCES

- Alloway, T. P., & Gathercole, S. E. (2006). *Working memory and neurodevelopmental disorders*. London: Psychology Press.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: APA.
- Arrindell, W. A., & Ettema, J. H. M. (2003). *Symptom Checklist (SCL-90)*. Handleiding bij een multidimensionele psychopathologie-indicator [SCL-90: Manual for a multidimensional psychopathology indicator]. Lisse, The Netherlands: Swets & Zeitlinger.
- Beilen van, M., Griffioen, B. T., Gross, A., & Leenders, K. L. (2009). Psychological assessment of malingering in psychogenic neurological disorders and non-psychogenic neurological disorders: relationship to psychopathology levels. *European Journal of Neurology*, 16, 1118–1123.

- Berry, D. T. R., Baer, R. A., Rinaldo, J. C., & Wetter, M. W. (2002). Assessment of malingering. In J. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (2nd ed., pp. 269–302). New York: Oxford University Press.
- Berry, D. T. R., & Nelson, N. W. (2010). DSM-5 and malingering: A modest proposal. *Psychological Injury and Law*, 3, 295–303.
- Bolan, B., Foster, J. K., & Bolan, S. (2002). A comparison of three tests to detect feigned amnesia. *The effects of feedback and the measurement of response latency*. *Journal of Experimental and Clinical Neuropsychology*, 24, 154–167.
- Boone, K. B. (2007). A reconsideration of the Slick et al. (1999) criteria for malingered neurocognitive dysfunction. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 29–49). New York: Guilford Press.
- Clegg, C., Fremouw, W., & Mogge, M. (2009). Utility of the Structured Inventory of Malingered Symptomatology (SIMS) and the Assessment of Depression Inventory (ADI) in screening for malingering among outpatients seeking to claim disability. *Journal of Forensic Psychiatry and Psychology*, 20, 239–254.
- Cragar, D. E., Berry, D. T. R., Fakhoury, T. A., Cibula, J. E., & Schmitt, F. A. (2006). Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *The Clinical Neuropsychologist*, 20, 552–566.
- De Bie, S. E. (1987). *Standaardvragen 1987: Voorstellen voor uniformering van Vraagstellingen naar achtergrondkenmerken en interviews [Standard questions 1987: Proposal for uniformisation of questions regarding background variables]* (2nd ed.). Leiden, The Netherlands: Leiden University Press.
- Dodrill, C. B. (2008). Do patients with psychogenic nonepileptic seizures produce trustworthy findings on neuropsychological tests? *Epilepsia*, 49, 691–695.
- Field, A. (2005). *Discovering statistics using SPSS, second edition*. London: Sage Publications.
- Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, 78, 199–208.
- Graue, L. O., Berry, D. T. R., Clark, J. A., Sollman, M. J., Cardi, M., Hopkins, J., et al. (2007). Identification of feigned mental retardation using the new generation of malingering detection instruments: Preliminary findings. *The Clinical Neuropsychologist*, 21, 929–942.
- Green, P. (2003). *The Word Memory Test*. Edmonton/Seattle: Green's Publishing Inc.
- Green, P. (2007a). Spoiled for choice, making comparisons between forced-choice effort tests. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 50–77). New York: Guilford Press.
- Green, P. (2007b). The pervasive influence of effort on neuropsychological tests. *Physical Medicine and Rehabilitation Clinics of North America*, 18, 43–68.
- Haggerty, K. A., Frazier, Th. W., Busch, R. M., & Naugle, R. I. (2007). Relationships among Victoria Symptom Validity Test indices and Personality Assessment Inventory validity scales in a large clinical sample. *The Clinical Neuropsychologist*, 21, 917–928.
- Hout, M. S. E., Schmand, B., Weking, E. M., & Deelman, B. G. (2006). Cognitive functioning in patients with suspected chronic toxic encephalopathy: Evidence for neuropsychological disturbances after controlling for insufficient effort. *Journal of Neurology, Neurosurgery and Psychiatry*, 77, 296–303.
- Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology*, 13, 77–90.
- Jelicic, M., Hessels, A., & Merckelbach, H. (2006). Detection of feigned psychosis with the Structured Inventory of Malingered Symptomatology (SIMS): A study of coached and uncoached simulators. *Journal of Psychopathology and Behavioral Assessment*, 28, 19–22.

- Kemp, S., Coughlan, A. K., Rowbottom, C., Wilkinson, K., Teggart, V., & Baker, G. (2008). The base rate of effort test failure in patients with medically unexplained symptoms. *Journal of Psychosomatic Research, 65*, 319–325.
- Locke, D. E. C., Smigielski, J. S., Powell, M. R., & Stevens, S. R. (2008). Effort issues in post acute outpatient acquired brain injury rehabilitation seekers. *Neurorehabilitation, 23*, 273–281.
- Luteijn, F., & Barelids, D. P. H. (2004). *GIT-2, Groninger Intelligentie Test 2 [Groningen Intelligence Test 2]*. Amsterdam, The Netherlands: Harcourt Assessment.
- Merkelbach, 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.
- Merten, Th., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology, 29*, 308–318.
- 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.
- Morey, L. C. (1991). *The Personality Assessment Inventory*. Odessa, FL: Psychological Assessment Resources.
- Nelson, N. W., Sweet, J. J., Berry, D. T. R., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*, 440–449.
- Nitch, S. R., & Glassmire, D. M. (2007). Non-forced measures to detect noncredible cognitive performance. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 78–102). New York: Guilford Press.
- Rienstra, A., Spaan, P. E. J., & Schmand, B. (2010). Validation of symptom validity tests using a “child-model” of adult cognitive impairments. *Archives of Clinical Neuropsychology, 25*, 371–382.
- Rogers, R. (Ed.). (2008). *Clinical assessment of malingering and deception*. New York: Guilford Press.
- Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment, 67*, 244–257.
- Rohling, M. L., Green, P., Allen, L. M., & Iverson, G. L. (2002). Depressive symptoms and neurocognitive test scores in patients passing symptom validity tests. *Archives of Clinical Neuropsychology, 17*, 205–222.
- Rosen, G. M., & Philips, W. R. (2004). A cautionary lesson from simulated patients. *The Journal of the American Academy of Psychiatry and the Law, 3*, 132–133.
- Ruocco, A. C., Swirsky-Sacchetti, Th., Chute, D. L., Mandel, S., Platek, S. M., & Zillmer, E. A. (2008). Distinguishing between neuropsychological malingering and exaggerated psychiatric symptoms in a neuropsychological setting. *The Clinical Neuropsychologist, 22*, 547–564.
- Schagen, S., Schmand, B., de Sterke, S., & Lindeboom, J. (1997). Amsterdam Short-Term Memory Test: A new procedure for the detection of feigned memory deficits. *Journal of Experimental and Clinical Psychology, 19*, 43–51.
- Schmand, B., de Sterke, S., & Lindeboom, J. (1999). *Amsterdamse Korte Termijn Geheugen test: Handleiding [Amsterdam Short-Term Memory Test: A Manual]*. Lisse, The Netherlands: Swets & Zeitlinger.

- 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 and Psychiatry*, *64*, 339–343.
- Schmand, B., & Lindeboom, J. (2005). *Amsterdam Short-Term Memory Test. Manual*. Leiden, The Netherlands: PITS.
- Slick, D., Hopp, G., Strauss, E., & Thompson, G. B. (1997). *VSVT: Victoria Symptom Validity Test, version 1.0, professional manual*. Odessa, FL: Psychological Assessment Resources.
- Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academic Psychiatry and the Law*, *25*, 183–189.
- Solomon, R. E., Boone, K. B., Miora, D., & Skidmore, S. (2010). Use of the Picture Completion Subtest as an embedded measure of effort; Abstracts of Division 40 Annual Meeting. *The Clinical Neuropsychologist*, *24*, 936.
- Stevens, A., Friedel, E., Mehren, G., & Merten, Th. (2008). Malingering and uncooperativeness in psychiatric and psychological assessment: Prevalence and effects in a German sample of claimants. *Psychiatry Research*, *157*, 191–200.
- Strong, D. R., Greene, R. L., & Schinka, J. A. (2000). A taxometric analysis of MMPI-2 infrequency scales (F and Fp) in clinical settings. *Psychological Assessment*, *12*, 166–173.
- Stroop, J. R. C. (1935). Studies of interference in serial verbal reaction. *Journal of Experimental Psychology*, *18*, 643–662.
- Stulemeijer, M., Andriessen, T. M. J. C., Brauer, J. M. P., Vos, P. E., & Van der Werf, S. (2007). Cognitive performance after mild traumatic brain injury: The impact of poor effort on test results and its relation to distress, personality and litigation. *Brain Injury*, *21*, 309–318.
- Suhr, J., Hammers, D., Dobbins-Buckland, K., Zimak, E., & Hughes, C. (2008). The relationship of malingering test failure to self-reported symptoms and neuropsychological findings in adults referred for ADHD evaluation. *Archives of Clinical Neuropsychology*, *23*, 521–530.
- 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.
- Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. North Tonawanda, NY: Multi Health Systems.
- Van Boxtel, M. P. J., Langerak, K., Houx, P. J., & Jolles, J. (1996). Self-reported physical activity, subjective health, and cognitive performance in older adults. *Experimental Aging Research*, *22*, 363–379.
- Van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P., & Jolles, J. (2005). Rey's verbal learning test: Normative data for 1855 healthy participants aged 24–81 years and the influence of age, sex, education, and mode of presentation. *Journal of the International Neuropsychological Society*, *11*, 290–302.
- Van der Elst, W., Van Boxtel, M. P. J., van Breukelen, G. J. P., & Jolles, J. (2006a). The Stroop color-word test: Influence of age, sex, and education; and normative data for a large sample across the adult age range. *Assessment*, *13*, 62–79.
- Van der Elst, W., van Boxtel, M. P. J., van Breukelen, G. J. P., & Jolles, J. (2006b). The Concept Shifting Test: Adult normative data. *Psychological Assessment*, *18*, 424–432.
- Wechsler, D. (1997). *WAIS-III Nederlandstalige bewerking, technische handleiding [WAIS-III: Dutch version; technical manual]*. Lisse, The Netherlands: Swets & Zeitlinger.
- Van Egmond, J., Kummeling, I., & Balkom, T. (2005). Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry*, *20*, 416–421.

- Whiteside, D. M., Dunbar-Mayer, P., & Waters, D. P. (2009). Relationship between TOMM performance and PAI validity scales in a mixed clinical sample. *The Clinical Neuropsychologist*, 23, 523–533.
- Widows, M. R., & Smith, G. P. (2005). *Professional manual: Structured inventory of malingered symptomatology*. Lutz, FL: Psychological Assessment Resources.