



UvA-DARE (Digital Academic Repository)

Indicators of suboptimal performance embedded in the Wechsler Memory Scale

Fourth Edition (WMS-IV)

Bouman, Z.; Hendriks, M.P.H.; Schmand, B.A.; Kessels, R.P.C.; Aldenkamp, A.P.

DOI

[10.1080/13803395.2015.1123226](https://doi.org/10.1080/13803395.2015.1123226)

Publication date

2016

Document Version

Final published version

Published in

Journal of Clinical and Experimental Neuropsychology

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Bouman, Z., Hendriks, M. P. H., Schmand, B. A., Kessels, R. P. C., & Aldenkamp, A. P. (2016). Indicators of suboptimal performance embedded in the Wechsler Memory Scale: Fourth Edition (WMS-IV). *Journal of Clinical and Experimental Neuropsychology*, 38(4), 455-466. <https://doi.org/10.1080/13803395.2015.1123226>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Indicators of suboptimal performance embedded in the Wechsler Memory Scale–Fourth Edition (WMS–IV)

Zita Bouman, Marc P.H. Hendriks, Ben A. Schmand, Roy P. C. Kessels & Albert P. Aldenkamp

To cite this article: Zita Bouman, Marc P.H. Hendriks, Ben A. Schmand, Roy P. C. Kessels & Albert P. Aldenkamp (2016) Indicators of suboptimal performance embedded in the Wechsler Memory Scale–Fourth Edition (WMS–IV), *Journal of Clinical and Experimental Neuropsychology*, 38:4, 455–466, DOI: [10.1080/13803395.2015.1123226](https://doi.org/10.1080/13803395.2015.1123226)

To link to this article: <http://dx.doi.org/10.1080/13803395.2015.1123226>



© 2016 The Author(s). Published by Taylor & Francis



Published online: 16 Feb 2016.



[Submit your article to this journal](#)



Article views: 264



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

Indicators of suboptimal performance embedded in the Wechsler Memory Scale–Fourth Edition (WMS–IV)

Zita Bouman^{a,b}, Marc P.H. Hendriks^{a,b}, Ben A. Schmand^{c,d}, Roy P. C. Kessels^{b,e} and Albert P. Aldenkamp^{a,f,g,h}

^aAcademic Centre for Epileptology, Kempenhaeghe, Heeze, the Netherlands; ^bDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands; ^cDepartment of Psychology, University of Amsterdam, Amsterdam, the Netherlands; ^dDepartment of Medical Psychology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; ^eDepartment of Medical Psychology, Radboud University Medical Center, Nijmegen, the Netherlands; ^fDepartment of Neurology, University Hospital Gent, Gent, Belgium; ^gSignal Processing System Group, Faculty of Electrical Engineering, Technical University Eindhoven, Eindhoven, the Netherlands; ^hDepartment of Neurology and School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, the Netherlands

ABSTRACT

Introduction. Recognition and visual working memory tasks from the Wechsler Memory Scale–Fourth Edition (WMS–IV) have previously been documented as useful indicators for suboptimal performance. The present study examined the clinical utility of the Dutch version of the WMS–IV (WMS–IV–NL) for the identification of suboptimal performance using an analogue study design.

Method. The patient group consisted of 59 mixed-etiology patients; the experimental malingerers were 50 healthy individuals who were asked to simulate cognitive impairment as a result of a traumatic brain injury; the last group consisted of 50 healthy controls who were instructed to put forth full effort.

Results. Experimental malingerers performed significantly lower on all WMS–IV–NL tasks than did the patients and healthy controls. A binary logistic regression analysis was performed on the experimental malingerers and the patients. The first model contained the visual working memory subtests (Spatial Addition and Symbol Span) and the recognition tasks of the following subtests: Logical Memory, Verbal Paired Associates, Designs, Visual Reproduction. The results showed an overall classification rate of 78.4%, and only Spatial Addition explained a significant amount of variation ($p < .001$). Subsequent logistic regression analysis and receiver operating characteristic (ROC) analysis supported the discriminatory power of the subtest Spatial Addition. A scaled score cutoff of <4 produced 93% specificity and 52% sensitivity for detection of suboptimal performance.

Conclusion. The WMS–IV–NL Spatial Addition subtest may provide clinically useful information for the detection of suboptimal performance.

ARTICLE HISTORY

Received 20 July 2015

Accepted 18 November 2015

KEYWORDS

Assessment; malingering; suboptimal performance; effort indicators; logistic regression.

Assessment of memory functioning plays a key role in neuropsychological evaluation of patients with a variety of neurological and psychiatric disorders. There are several well-developed and standardized memory tests and batteries available, such as the Wechsler Memory Scale (WMS; Lezak, Howieson, Bigler, & Tranel, 2012). However, one of the difficulties that arise when validating neuropsychological tests is the assumption that the test performance of the examinee is a true reflection of his or her actual level of ability (Brennan &

Gouvier, 2006; Larrabee, 2012; Merckelbach, Smeets, & Jelicic, 2009; Slick, Sherman, & Iverson, 1999). Therefore, it is recommended to assess performance validity routinely in neuropsychological evaluations (American Academy of Clinical Neuropsychology, 2007; Bush et al., 2005; Heilbronner et al., 2009).

One possible cause for invalid test performance is malingering, which is defined as “the intentional production of false or grossly exaggerated physical or psychological problems.

CONTACT Zita Bouman  z.bouman@donders.ru.nl  Donders Institute for Brain Cognition and Behaviour, Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, the Netherlands.

© 2016 The Author(s). Published by Taylor & Francis

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motivation for malingering is usually external (e.g., avoiding military duty or work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs)” (*Diagnostic and statistical manual of mental disorders—Fifth Edition, DSM-V*; American Psychiatric Association, 2013). There are several performance validity tests (PVTs) that are designed with the purpose of assessing whether an individual’s test performance on data obtained by neuropsychological tests is valid (Dandachi-FitzGerald, Ponds, & Merten, 2013; Larrabee, 2012). Examples of PVTs are the Test of Memory Malingering (TOMM; Tombaugh, 1996) and the Amsterdam Short Term Memory Test (ASTM: Schmand, Lindeboom, & Merten, 2005; Schagen, Schmand, de Sterke, & Lindeboom, 1997).

In addition to PVTs, several studies have proposed methodologies to derive indicators of suboptimal performance within common neuropsychological tests, so called “embedded” validity indicators (Larrabee, 2012; Slick et al., 1999). Well-established embedded indicators for suboptimal performance are poor performance on recognition tasks in relation to relatively adequate performance on delayed recall tasks (Bernard, 1990; Haines & Norris, 2001; Langeluddecke & Lucas, 2003) and relatively poor performance on tasks involving immediate span of attention, as they may be perceived as memory tasks by malingerers while tapping simple attentional functions (Axelrod, Fichtenberg, Millis, & Wertheimer, 2006; Heinly, Greve, Bianchini, Love, & Brennan, 2005; Iverson & Tulsky, 2003; Langeluddecke & Lucas, 2003).

In particular, memory tests have been examined to determine their efficacy in identifying suboptimal performance (cf. Lu, Rogers, & Boone, 2007; Suhr & Barrash, 2007), mainly because tests designed to assess memory and concentration are particularly susceptible to exaggeration or fabrication of cognitive impairment. This is hardly surprising given that it is well known that memory and concentration disorders are common symptoms following head injury (Mittenberg, Azrin, Millsaps, & Heilbronner, 1993; Williams, 1998). Several studies have examined indicators and patterns of suboptimal performance using the Wechsler Memory Scale—Third Edition (WMS—III; Wechsler, 1997), with varying levels of success. Some of these studies have used the entire instrument and

demonstrated that malingering traumatic brain injury (TBI) patients returned lower WMS—III mean scores than nonmalingerers TBI patients (Langeluddecke & Lucas, 2003; Ord, Greve, & Bianchini, 2008). Other studies have examined the use of specific subtests (Faces; Glassmire et al., 2003), rarely missed items (Rarely Missed Index; Bortnik et al., 2010; Killgore & DellaPietra, 2000; Lange, Sullivan, & Anderson, 2005; L. J. Miller, Ryan, Carruthers, & Cluff, 2004; Swihart, Harris, & Hatcher, 2008), and difference-scores for index and subtests (Lange, Iverson, Sullivan, & Anderson, 2006; Langeluddecke & Lucas, 2003) to discriminate between malingering and nonmalingerers patients.

For the latest editions of the Wechsler intelligence and memory batteries, the Wechsler Adult Intelligence Scale—Fourth Edition (WAIS—IV; Wechsler, 2008) and the Wechsler Memory Scale—Fourth Edition (WMS—IV; Wechsler, 2009), the additional Advanced Clinical Solutions (ACS) package provides several embedded measures for the detection of malingering including the Reliable Digit Span from the WAIS—IV (Greiffenstein, Baker, & Gola, 1994), the four recognition tasks (Logical Memory Recognition, LM-Rec; Verbal Paired Associates Recognition, VPA-Rec; Visual Reproduction Recognition, VR-Rec; and Designs Recognition, DE-Rec) from the WMS—IV, and the newly developed Word Choice Test (which has a similar format to that of the Warrington Memory Test; Holdnack & Drozdick, 2009).

So far, only two studies found promising results for the WMS—IV ACS package as an effective tool for detection of suboptimal performance (Holdnack & Drozdick, 2009; J. B. Miller et al., 2011). Furthermore, a recent study by Young, Caron, Baughman, and Sawyer (2012) identified the Symbol Span subtest as an indicator of suboptimal performance. This is not surprising as the Symbol Span is a visual analogue of the Digit Span task, which has proven to be able to detect malingering according to a number of validation studies (Axelrod et al., 2006; Babikian, Boone, Lu, & Arnold, 2006; Heinly et al., 2005; Iverson & Tulsky, 2003).

The WMS is one of the most widely used memory batteries to assess memory function (Rabin, Barr, & Burton, 2005). Several studies have reported effective embedded validity indicators using previous versions of the WMS, but so far only few studies used the WMS—IV. The aim of our study is to examine whether several tasks of

the WMS-IV can be used as embedded validity indicators using the Dutch version of this battery (WMS-IV-NL; Hendriks, Bouman, Kessels, & Aldenkamp, 2014). We selected a number of tasks that we expected to distinguish between malingering participants and nonmalingering neurological patients. First, we selected the visual working memory tasks Spatial Addition (SA) and Symbol Span (SSP), as working memory tests were previously found to be sensitive in other WMS studies (Lange et al., 2006; Young et al., 2012). Secondly, we selected the recognition tasks LM-Rec, VPA-Rec, DE-Rec, and VR-Rec, because these subtests were already shown to be sensitive in previous research using the WMS-IV (Holdnack & Drozdick, 2009; J. B. Miller et al., 2011).

Method

Participants

A three-group design was used to compare WMS-IV-NL performance of healthy volunteers who were instructed to simulate cognitive impairment due to TBI (i.e., “experimental malingerers”), mixed-etiology patients, and healthy controls. The first sample of experimental malingerers consisted of 50 healthy participants who were instructed to pretend to be cognitively impaired as a result of a TBI. This group of participants was recruited by the researchers through their network. Exclusion criteria for this sample were: inability to speak/understand the Dutch language; significant hearing or visual impairment; psychiatric or neurologic disorder; substance abuse affecting cognitive functioning; use of medicines affecting cognitive functioning; and not following the malingering instruction, as established by a questionnaire and a PVT: the ASTM (see also Procedure section).

Second, a total of 59 mixed-etiology patients were recruited from several rehabilitation centres in the Netherlands: Bavo-Europoort Center for Neuropsychiatry/Acquired Brain Injury, Rotterdam ($n = 21$); Bravis Hospital Roosendaal ($n = 20$); Rehabilitation Centre Groot Klimmendaal Arnhem ($n = 14$); and Sophia Rehabilitation Centre, The Hague ($n = 4$). Of these patients, 27 were diagnosed with TBI; 23 with a stroke (cerebrovascular accident; CVA); 4 with postanoxic encephalopathy; 2 with a tumor; 2 with multiple sclerosis; and 1 with

meningococcal meningitis. Patients were excluded if they met the following exclusion criteria: inability to speak/understand the Dutch language; significant hearing or visual impairment; evidence for suboptimal performance (based on performance validity testing or expert opinion).

The third sample of participants consisted of 50 healthy controls selected from the Dutch version of the WMS-IV (WMS-IV-NL) standardization study (see Hendriks et al., 2014, for a detailed description of the participant selection) and were matched for age, sex, and education level with the other groups. Moreover, healthy controls were excluded if they met the following exclusion criteria: inability to speak/understand the Dutch language; significant hearing or visual impairment; psychiatric or neurologic disorder; substance abuse affecting cognitive functioning; and use of medicines affecting cognitive functioning. Participant characteristics are summarized in Table 1.

Measures

The primary measure in this study was the WMS-IV-NL, which was administered and scored according to the test manual (Hendriks et al., 2014). The authorized Dutch version of the WMS-IV is equivalent to the original American version. The nonverbal visual stimuli are identical in both language versions, and the instruction, auditory stimuli, and scoring criteria were translated and adapted to the Dutch language. A previous study revealed that the WMS-IV and WMS-IV-NL have a similar factor structure (Bouman, Hendriks, Kerkmeier, Kessels, & Aldenkamp, 2015).

The WMS-IV-NL contains one optional subtest, the Brief Cognitive Status Exam (BCSE), and six primary subtests: Logical Memory (LM), Verbal Paired Associates (VPA), Designs (DE), Visual Reproduction (VR), Spatial Addition (SA) and Symbol Span (SSP). Of these, four subtests (LM, VPA, DE, and VR) have immediate and delayed recall conditions. The primary subtests were converted into age-adjusted scaled scores ($M = 10$, $SD = 3$), which were used in all analyses. These subtest scaled scores can be used to calculate five index scores: Auditory Memory Index (AMI), Visual Memory Index (VMI), Immediate Memory Index (IMI), Delayed Memory Index (DMI) and Visual Working Memory Index (VWMI). Several subtests

Table 1. Participant characteristics.

	Experimental malingerers	Mixed-etiology patients	Controls	Statistic	<i>p</i>
<i>N</i>	50	59	50		
Age (years; mean, <i>SD</i>)	40.6 (14.7)	46.1 (16.3)	40.3 (14.6)	$F(2, 156) = 2.57$.083
Sex distribution (M/F)	28/22	40/19	28/22	$\chi^2(2, n = 159) = 2.16$.340
Education level (low/average/high)	15/21/14	26/18/12	16/21/13	$\chi^2(2, n = 159) = 3.05$.218
DART IQ (mean, <i>SD</i>)	102.3 (13.7)	94.8 (14.4)	101.8 (15.4)	$F(2, 149) = 4.33$.015

Note. Education level was assessed by classifying formal schooling in the Netherlands according to the grouping of the Central Office of Statistics for the Netherlands (CBS, 2011), which is based on the International Standard Classification of Education (ISCED: United Nations Educational, Scientific and Cultural Organisation Institute for Statistics UNESCO-UIS, 2011). The National Adult Reading Test (NART) IQ was not available from all patients. DART = Dutch version of NART; M = male; F = female. The sample size: experimental malingerers, $n = 50$; mixed-etiology patients, $n = 54$; and healthy controls, $n = 48$.

also include optional tasks, including recognition tasks (for the subtests LM, VPA, DE and VR), separate scores for DE content and spatial scores, a word recall task for VPA (in which the examinee is asked to sum up as many of the words from the pairs as he or she can recall), and a copy task for VR (in which the examinee is asked to draw the figures while looking at them). Because the score distribution of the recognition tasks and the VR copy task are highly skewed, there are no scaled scores available in the WMS-IV. Thus, in the following analyses raw scores were used for these tasks.

In addition, the Dutch version of the National Adult Reading Test (NART: Nelson, 1982; DART: Schmand, Lindeboom, & Van Harskamp, 1992) was administered to all participants to obtain an estimation of premorbid verbal intelligence. Moreover, the experimental malingerers underwent short structured interviews at the beginning and the end of the examination and completed the ASTM (Schagen et al., 1997; Schmand et al., 2005). The ASTM is a forced-choice verbal memory test that is designed to assess (in)valid performance. Individual performance on the ASTM was used to perform a manipulation check (i.e., to check whether a experimental malingerer performed below the previously established cutoff score of ≤ 84). With a cutoff score of ≤ 84 the ASTM has a sensitivity of 91% and a specificity of 89% (Schmand et al., 2005). Also, two questionnaires were used to determine how they interpreted the complaints accompanying TBI (for the detailed questionnaires see the Appendix).

Procedure

This study was approved by the Institutional Review Board of the Faculty of Social Sciences of Radboud

University in Nijmegen, and patient data were collected as part of the routine clinical assessment of each participating centre. Written informed consent was obtained from all participants.

The examiner provided the participants in the experimental malingerer group with the following scenario and instructions containing symptom coaching two days before testing. This scenario was based on previous studies (Brennan & Gouvier, 2006; Brennan et al., 2009; Suhr & Gunstad, 2000; Tan, Slick, Strauss, & Hultsch, 2002; Weinborn, Woods, Nulsen, & Leighton, 2012) and the recommendations outlined by Suhr and Gunstad (2000).

Instructions: Six months ago you were involved in a car accident, and you don't suffer any consequences from it at the moment. Imagine that your lawyer tells you that you could get a large sum of money from an insurance company, but only if it is determined that you suffer from brain damage. In a few days, you will undergo neuropsychological tests to assess whether you have brain damage. You have decided to simulate the symptoms of brain damage. Commonly experienced problems in brain damage are: fatigue, memory problems and problems with attention, depression, slowed response, irritability and anxiety. Try to imagine how a person with brain damage would perform on these tests you're about to take. Do keep in mind that you have to make it seem believable; some of the tests you will take can be specifically designed to detect people faking. When the results of the assessment show that you have been faking, you will not get the money. If you think it is necessary you may look for information about brain damage to prepare yourself. You cannot ask the test assessor any questions about your role though.

This scenario was successfully used in prior research as an example of extrinsic motivation to malingering (Brennan & Gouvier, 2006; Jellic, Merckelbach, Candel, & Geraerts, 2007). Furthermore, the described TBI symptoms were

likely to be found online or to be provided by a client's lawyer in a real litigation case. If a participant was unable or unwilling to follow the instructions, he or she was excluded from the study.

Prior to testing, all experimental malingerers underwent a structured interview about their complaints to simulate a true neuropsychological assessment. Following the completion of the neuropsychological tests according to the standardized procedures—ASTM, WMS-IV-NL—the experimental malingerers completed a questionnaire requiring them to report whether or not they followed instructions to feign cognitive impairment (for the detailed questionnaires see the [Appendix](#)). Finally, the experimental malingerers were asked to put forth their full effort on the DART.

For the patients, the WMS-IV-NL and DART were administered as part of a comprehensive neuropsychological evaluation; for the healthy controls, the WMS-IV-NL and DART were administered as part of the Dutch standardization study (Hendriks et al., 2014). All these participants were asked to put forth their full effort on all (neuro)psychological tests.

Analyses

First, we compared the three groups (experimental malingerers, mixed-etiology patients, and healthy controls) using a one-way multivariate analyses of variance (MANOVA) with group (experimental malingerers, patients, healthy controls) as between-subjects factor and 15 WMS-IV-NL subtest scores as dependent variables. Furthermore, as the WMS-IV-NL BCSE and subtest recognition scores were not normally distributed, Kruskal-Wallis analyses were carried out. Significant differences were analyzed with Bonferroni-corrected post hoc analyses.

Group means of overall performance reveal little information about the test's ability to detect suboptimal performance, and, therefore, we also performed logistic regression analyses. As the working memory and recognition subtests are expected to indicate malingering based on previous research and theoretical background, we used these six scores in a logistic regression analysis (SA, SSP, LM II Rec, VPA II Rec, DE II Rec, and VR II Rec). Only experimental malingerers and patients were included, as the differentiation between these two groups was of interest here. If a selection of WMS-

IV-NL score(s) were found to contribute substantially to the model's ability to predict outcome, a subsequent logistic regression analysis that contains only these important predictor(s) was fitted. The Hosmer-Lemeshow goodness-of-fit statistic (Hosmer & Lemeshow, 2000) was used to determine whether the models provided a good fit for the data. A significant Hosmer-Lemeshow value means that the calibration is insufficient, but large values ($p > .05$) indicate that the models are well calibrated and fit the data. Furthermore, receiver operating characteristics (ROC) analyses were performed on the selection of significant predictor(s). ROC analysis generates an area under the curve (AUC) value, which indicates the discriminative power of the predictor.

Results

All experimental malingerers reported on the questionnaire that they were successful in following our malingering instructions. In line with this, all participants scored below the previously established cutoff score of 84 on the ASTM (range = 34 to 83), which indicates that all participants followed the instructions and adequately feigned (mild) brain damage according to the present scenario. As a result, none of the experimental malingerers had to be removed from the sample.

Group comparisons

The three groups were equivalent for age, sex, and education level (all $p > .08$), but significant differences were found for verbal intelligence level (DART IQ), $F(2, 149) = 4.33$, $p < .05$, $\eta_p^2 = .06$. Bonferroni-corrected post hoc analyses revealed that the patients and healthy controls did not differ significantly ($p = .05$), whereas the patients revealed lower verbal intelligence level than the experimental malingerers ($p = .03$). Correlation analyses revealed that there were low correlations between DART IQ and the WMS-IV-NL subtest scores (Pearson product-moment correlation coefficients ranging from $-.02$ to $.34$), therefore, no covariates were included in the analyses.

The MANOVA with group (experimental malingerers, patients, and healthy controls) as between-subjects factor and 15 WMS-IV-NL subtest scores as dependent variables revealed an overall main effect for group, $F(30, 266) = 5.67$,

$p < .001$, $\eta_p^2 = .39$. Moreover, the Kruskal–Wallis analyses revealed significant main effects of group for the WMS–IV–NL BCSE and subtest recognition scores (all $p < .001$). Bonferroni-corrected post hoc tests revealed that patients performed worse than healthy controls on all WMS–IV–NL scores, except for the VR II Copy task. Moreover, the experimental malingers performed worse than healthy controls on all WMS–IV–NL scores, except for the process-score DE I Content; and they performed worse than the patients on LM I, LM II, VR I, SA, SSP, BCSE, and three of the four recognition tasks (LM-Rec, VPA-Rec, and VR-Rec). The average WMS–IV–NL subtest, BCSE, recognition, and process scores for the experimental malingers, patients, and healthy controls are presented in Table 2.

Classification accuracy statistics

A logistic regression model was fitted to determine which of the WMS–IV–NL tasks best

discriminated between patients and experimental malingers. Given our a priori hypothesis, the WMS–IV–NL visual working memory subtests (SA and SSP) and recognition tasks (LM-Rec, VPA-Rec, DE-Rec, and VR-Rec) were entered as independent variables into the initial model. A test of the model with these six variables against a constant-only model was statistically significant, $\chi^2(6) = 45.78$, $p < .001$, indicating that this combination of variables was able to distinguish between patients and experimental malingers. Moreover, the value of the Hosmer–Lemeshow goodness-of-fit statistic was 4.11, and the corresponding p -value was .85, which indicated that this model was well calibrated. The model as a whole explained between 36.2% (Cox and Snell R square) and 48.2% (Nagelkerke R squared) of the variance and correctly classified 78.4% of cases. As reported by the Wald criterion, only the SA subtest explained a significant amount of variation ($p < .001$), recording an odds ratios of 0.60.

Table 2. Mean scores and standard deviations of the WMS–IV–NL indexes and subtests for experimental malingers, mixed-etiology patients, and healthy controls.

WMS–IV–NL subtest scores	Experimental malingers Mean (SD)	Mixed-etiology patients Mean (SD)	Controls Mean (SD)	F	p	Mal vs. controls	Patients vs. controls
						η_p^2	η_p^2
LM I	5.9 (2.7) ^a	8.0 (2.8) ^a	10.3 (3.4) ^a	27.14	<.001	.34	.13
LM II	5.2 (2.4) ^a	7.4 (2.8) ^a	10.1 (3.1) ^a	39.18	<.001	.44	.17
VPA I	7.1 (1.9) ^a	7.3 (3.5) ^b	9.6 (3.4) ^{ab}	11.04	<.001	.18	.11
VPA II	6.3 (2.3) ^a	7.3 (3.6) ^b	9.7 (3.6) ^{ab}	15.25	<.001	.25	.10
VPA II Word Recall	6.3 (2.1) ^a	7.2 (3.6) ^b	9.9 (3.3) ^{ab}	18.90	<.001	.30	.14
DE I	7.6 (2.7) ^a	8.0 (3.3) ^b	10.0 (3.6) ^{ab}	8.18	<.001	.13	.08
DE I Content	8.3 (2.7)	8.2 (3.4) ^a	9.8 (3.4) ^a	4.24	.016	.06	.06
DE I Spatial	7.9 (2.7) ^a	8.5 (3.0) ^b	10.3 (3.4) ^{ab}	8.45	<.001	.14	.07
DE II	8.1 (2.9) ^a	8.1 (2.9) ^b	10.7 (3.2) ^{ab}	13.58	<.001	.16	.16
DE II Content	8.4 (2.9) ^a	8.0 (2.7) ^b	9.9 (2.4) ^{ab}	6.94	.001	.07	.12
DE II Spatial	8.9 (2.6) ^a	8.5 (2.3) ^b	10.6 (3.4) ^{ab}	8.15	<.001	.07	.12
VR I	5.8 (2.9) ^a	8.0 (3.7) ^a	10.5 (3.0) ^a	27.62	<.001	.40	.13
VR II	6.1 (2.7) ^a	7.3 (3.1) ^b	10.3 (3.5) ^{ab}	23.45	<.001	.31	.17
SA	4.0 (2.3) ^a	8.0 (3.2) ^a	10.5 (2.8) ^a	67.29	<.001	.61	.14
SSP	5.2 (3.1) ^a	7.6 (3.3) ^a	10.6 (2.9) ^a	37.60	<.001	.45	.19
<i>WMS–IV–NL BCSE score</i>				χ^2	p	r	r
BCSE total score (max = 58)	35.1 (12.8) ^a	48.9 (9.3) ^a	53.9 (6.3) ^a	53.85	<.001	.67	.26
<i>WMS–IV–NL recognition scores</i>				χ^2	p	r	r
LM-Rec (max = 30)	20.5 (3.6) ^a	22.9 (3.7) ^a	25.5 (2.5) ^a	45.37	<.001	.66	.38
VPA-Rec (max = 40)	32.0 (5.2) ^a	35.3 (5.8) ^a	38.8 (1.9) ^a	56.98	<.001	.76	.38
DE-Rec (max = 24)	13.5 (3.1) ^a	14.6 (3.0) ^b	16.6 (3.1) ^{ab}	21.00	<.001	.75	.43
VR-Rec (max = 7)	3.5 (1.9) ^a	5.2 (1.7) ^a	6.5 (0.8) ^a	61.11	<.001	.45	.29
<i>WMS–IV–NL process score</i>				χ^2	p	r	r
VR Copy (max = 43)	40.3 (4.6) ^a	41.9 (2.1)	42.4 (1.1) ^a	10.24	.006	.31	.09

Note. WMS–IV–NL = Wechsler Memory Scale–Fourth Edition, Dutch version; Mal = experimental malingers; LM = Logical Memory; VPA = Verbal Paired Associates; DE = Designs; VR = Visual Reproduction; SA = Spatial Addition; SSP = Symbol Span; Rec = Recognition; BCSE = Brief Cognitive Status Exam. The WMS–IV–NL subtest scores are presented in age-adjusted scaled scores ($M = 10$, $SD = 3$). Groups with the same notation (a, b, c) differ significantly (Bonferroni-corrected $p < .05$).

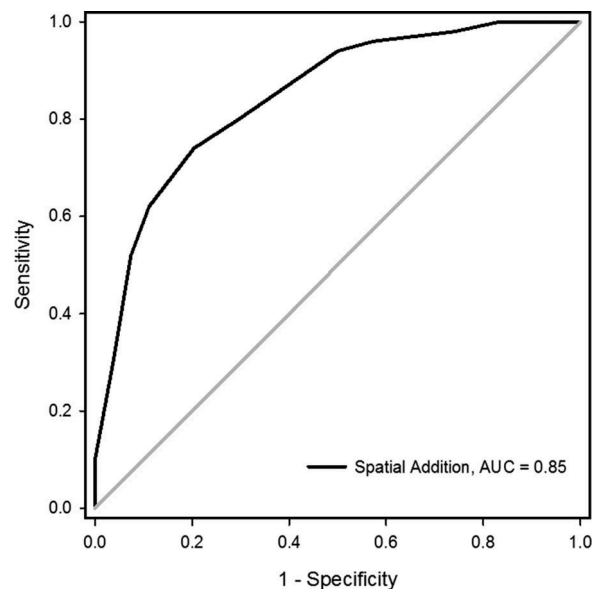
Table 3. Logistic regressions for predictive value of subtests differentiating experimental malingerers from mixed-etiology patients for the full model and single variable models.

Model	Predictor	B	SE	Wald	df	p	Odds ratio	95% CI for odds ratio	
Full model	<i>WMS-IV-NL subtest scores</i>								
	SA	-0.514	0.134	14.770	1	<.001	0.598	[0.460, 0.777]	
	SSP	0.041	0.108	0.142	1	.706	1.041	[0.843, 1.287]	
	<i>WMS-IV-NL recognition scores</i>								
	LM-Rec	-0.010	0.098	0.011	1	.915	0.990	[0.817, 1.199]	
	VPA-Rec	-0.056	0.072	0.590	1	.442	0.946	[0.821, 1.090]	
	DE-Rec	0.042	0.097	0.188	1	.665	1.043	[0.862, 1.262]	
	VR-Rec	-0.060	0.198	0.091	1	.763	0.942	[0.639, 1.388]	
	Constant	4.525	2.288	3.912	1	.048	92.284		
Single variable	SA	-0.545	0.108	25.623	1	<.001	0.580	[0.470, 0.716]	
	Constant	3.092	0.656	22.227	1	<.001	22.014		

Note. WMS-IV-NL = Wechsler Memory Scale-Fourth Edition, Dutch version; LM = Logical Memory; VPA = Verbal Paired Associates; DE = Designs; VR = Visual Reproduction; SA = Spatial Addition; SSP = Symbol Span; Rec = Recognition; CI = confidence interval.

Next, a univariate logistic regression model that contained the stand-alone SA subtest was fitted. This model was performed to determine whether the SA subtest alone revealed a similar model classification. A test of the model with this variable against a constant-only model was statistically significant, $\chi^2(1) = 44.30$, $p < .001$, indicating that the SA subtest was able to distinguish between patients and experimental malingerers. Moreover, the value of the Hosmer-Lemeshow goodness-of-fit statistic was 1.58, and the corresponding p -value was .99, which indicated that this model was well calibrated. The SA explained between 34.7% (Cox and Snell R square) and 46.3% (Nagelkerke R squared) of the variance and correctly classified 76.9% of cases. The odds ratio was 0.58, which indicates that for every additional subtest scaled score point on the subtest SA, respondents were 0.58 times less likely to malingering. The regression coefficients for both models are presented in Table 3.

Predictive performance of the subtest SA was further examined using a ROC analysis, which revealed that SA produced a good separation between the groups as indicated by the AUC of 0.85 ($SD = 0.04$, $p < .001$, 95% CI [0.77, 0.92]; Hosmer & Lemeshow, 2000). Figure 1 shows the ROC curve for the SA subtest for detecting sub-optimal performance. As the current study evaluates cutoff scores for measuring performance validity, high specificity rates are required to minimize false-positive errors—that is, misdiagnosing an individual with real cognitive deficits (Larrabee & Berry, 2007). A specificity of 90% is recommended (Axelrod et al., 2006; Babikian et al., 2006), but this reduces the sensitivity to 52%. In



Receiver operating characteristic (ROC) curve for Spatial Addition subtest scaled score for distinguishing experimental malingerers from mixed-etiology patients. AUC = area under the curve.

some contexts, other preassigned values for sensitivity and specificity may be preferred, and, therefore, a range of probability cutoff scores for SA and their associated diagnostic efficiency found in this sample is presented in Table 4.

Discussion

The present study aimed to examine whether several tasks of the WMS-IV-NL could be used as embedded indicators for the differentiation between malingerers and patients with mild to severe acquired brain injuries. Overall, the Spatial Addition subtest may provide clinically useful

Table 4. Sensitivity and specificity for different Spatial Addition subtest scaled score cutoff scores.

Cutoff value	Sensitivity (%)	Specificity (%)
<2	10	100
<3	30	96
<4	52	93
<5	62	89
<6	74	80
<7	80	70
<8	94	50
<9	96	43
<10	98	26
<11	100	17

Note. Extreme values for the curve were omitted to increase readability.

information for the detection of suboptimal performance.

Our findings concerning the between-group comparisons indicated that both the experimental malingerers and the mixed-etiology patients performed significantly lower than healthy controls on all WMS-IV-NL scores, which is in line with previous studies (Carlozzi, Grech, & Tulsky, 2013; Langeluddecke & Lucas, 2003; Ord et al., 2008). Furthermore, in comparison with the patients, experimental malingerers scored significantly worse on the optional cognitive screener (BCSE), two auditory verbal memory subtests (LM I and LM II), one visual memory subtest (VR I), both visual working memory subtests (SA and SSP), and three of the four recognition tasks (LM-Rec, VPA-Rec, and VR-Rec). These results are in agreement with the notion that malingerers have a tendency to overestimate the magnitude of cognitive deficits arising from brain injury and, as a result, show even poorer performances than patients on previous editions of the WMS (Langeluddecke & Lucas, 2003; Rogers, 2007; Schwartz, Gramling, Kerr, & Morin, 1998).

Since differences in group means of overall performance reveal little information about the test's ability to detect suboptimal performance, the classification accuracy statistics are noteworthy. In our first logistic regression analysis, the visual working memory subtests (SA and SSP) and recognition tasks (LM-Rec, VPA-Rec, DE-Rec, and VR-Rec) were found to discriminate 78.4% of cases. Of the variables entered in the model, only the SA subtest differentiated significantly between patients and experimental malingerers. These results are not fully in agreement with some studies that have showed the use of multiple WMS-IV scores for the detection of suboptimal performance (J. B. Miller et al., 2011; Holdnack & Drozdick, 2009;

Young et al., 2012). J. B. Miller and colleagues (2011) found that four of the five WMS-IV ACS scores (i.e., Word Choice Test, Digit Span, VPA-Rec, and VR-Rec) performed well in discriminating between moderate to severe TBI patients and coached experimental malingerers. This study by J. B. Miller et al. (2011) also included the newly developed Word Choice Test as part of the optional Advanced Clinical Solutions package for the WMS-IV, which is not available in the Netherlands, which may partly explain the discrepancy in findings. However, this cannot fully explain the differences between our and their study results, as the recognition tasks are equivalent. Moreover, although the study performed by J. B. Miller et al. (2011) included healthy adults coached to feign cognitive impairment, which is similar to our design, these were only compared to patients with traumatic brain injury. This design differs from our study that recruited mixed-etiology neurological patients. Another study (Young et al., 2012) found that the SSP subtest differentiated well between adequate and inadequate effort in a mixed clinical group of veterans, which we did not find. However, no other WMS-IV subtests were administered in Young et al. (2012). Finally, it should be stressed that the authorized Dutch version of the WMS-IV is equivalent to the originally published U.S. version (Hendriks et al., 2014; Wechsler, 2009), with a similar factor structure (Bouman et al., 2015). Therefore, it is likely that our results can be extended to other-language versions of the WMS-IV.

Our second logistic regression analysis and the ROC analysis on the stand-alone SA subtest reveal that this subtest alone has good overall discriminative validity in the detection of malingering with an AUC value of 0.85. This result is comparable to the AUCs reported for the WMS-IV Word Choice Test and WMS-IV SSP subtest (i.e., AUC values of 0.84 and 0.75, respectively: J. B. Miller et al., 2011; Young et al., 2012), but lower than the AUC value of 0.95 that was found for the WMS-IV ACS package (including the WMS-IV recognition tasks, the Word Choice Test and reliable digit span: J. B. Miller et al., 2011). Furthermore, the SA subtest has a sensitivity of 52% at a specificity of 93%. Thus, when the performance on SA results in a score of 4 or less, there is a substantial risk of approximately 50% false negatives (i.e., missing feigned cognitive impairment) but, more importantly, there is only a risk of approximately 10%

false positives (i.e., misclassifying an individual with real cognitive deficits). These results are comparable to the previously reported average sensitivity of .53 and specificity of .91 for 5 embedded indicators on standard neuropsychological and psychological tests (Larrabee, 2003). Moreover, the sensitivity is somewhat higher than the sensitivity of 26% that was found at a specificity of 93% for the SSP subtest reported by Young et al. (2012).

Notably, the experimental malingerers were coached about what symptoms to expect, as well as being warned about performance validity tests. These processes can affect malingering performance and may have dropped the identification accuracy of the embedded indicators in the WMS-IV examined in this study (Jelicic et al., 2007; Schenk & Sullivan, 2010). However, it is likely that the experimental malingerers adequately feigned (mild) brain damage, as they all scored below the cutoff score of 84 on the ASTM (range = 34 to 83). Moreover, with a cutoff score of ≤ 83 the ASTM has a specificity of 95%, so less than 5% of the neurologically impaired patients in the validation study performed that low (Schmand et al., 2005).

Several limitations of this study need to be addressed. First, in comparison to the study by J. B. Miller et al. (2011), we included a sample of analogue malingerers. Although analogue study designs have been recommended (Bush et al., 2005; Heilbronner et al., 2009), this design has sometimes been criticized for external validity concerns as it remains unclear whether the experimental malingering performance of these healthy controls is comparable to real-world malingering (cf. Haines & Norris, 1995; Larrabee, 2007; Rogers, 2007; Suhr & Gunstad, 2000; Vickery et al., 2004). Further research is warranted to replicate these findings in clinical studies with suspected, real-world malingerers.

One could also argue that it is a limitation that the clinical sample is heterogeneous—that is, consisting of patients having different neurological disorders. However, we purposely included a heterogeneous sample, as we wanted to enhance the external validity of our findings relevant for use in a mixed-etiology patient group. In future, it would be interesting to examine the applicability of the WMS-IV, and in particular the SA subtest, in the identification of malingering in specific neurological (or psychiatric) disorders, comparing, for instance, mildly, moderately, and severely cognitively impaired patients, as well as different subgroups (e.g., different types of stroke or different

subtypes of MS), as well as other settings. Furthermore, only patients who did not show evidence for suboptimal performance (based on performance validity testing or expert opinion) were referred to our study based on the inclusion criteria. As a result, our patient sample did not complete the same PVT, as performance validity testing was done as part of the diagnostic work-up of the individual clinics using different, yet widely used PVTs.

In conclusion, findings from the current study show that the WMS-IV-NL visual working memory subtest Spatial Addition might be a valid embedded indicator for the detection of suboptimal performance. However, it should be stressed that the test's sensitivity is lower than its specificity, making it important to not base the detection of suboptimal effort on a single test; rather the Spatial Addition subtest might have added value in clinical practice when used in combination with other measures for the detection of suboptimal performance.

Acknowledgements

We thank Pearson Assessment BV, Amsterdam, The Netherlands, for authorizing and funding the development of the WMS-IV-NL. The authors would like to express their gratitude towards Coby van Drie, Judith Grit, Henriëtte van der Zee, and Luciano Fasotti for their assistance in the data collection of mixed-etiology patients; towards Ajla Mujcic and Karlijne Grootjans for their assistance in the data collection of experimental malingerers; and towards Dirk Bertens for his helpful comments on our paper.

Disclosure statement

No potential conflicts of interest was reported by the author(s).

Funding

This study was funded by Pearson Assessment BV and Academic Centre for Epileptology, Kempenhaeghe, Heeze, the Netherlands. The funder had no role in study design, analyses, or the decision to publish the results.

References

- American Academy of Clinical Neuropsychology. (2007). American Academy of Clinical Neuropsychology (AACN) practice guidelines for neuropsychological assessment and consultation. *The*

- Clinical Neuropsychologist*, 21, 209–231. doi:10.1080/13825580601025932
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.
- Axelrod, B. N., Fichtenberg, N. L., Millis, S. R., & Wertheimer, J. C. (2006). Detecting incomplete effort with Digit Span from the Wechsler Adult Intelligence Scale—Third Edition. *The Clinical Neuropsychologist*, 20, 513–523. doi:10.1080/13854040590967117 (2013).
- Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, 20, 145–159. doi:10.1080/13854040590947362
- Bernard, L. C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology*, 12, 715–728. doi:10.1080/01688639008401014
- Bortnik, K. E., Boone, K. B., Marion, S. D., Amano, S., Ziegler, E., Cottingham, M. E., . . . Zeller, M. A. (2010). Examination of various WMS–III logical memory scores in the assessment of response bias. *The Clinical Neuropsychologist*, 24, 344–357. doi:10.1080/13854040903307268
- Bouman, Z., Hendriks, M. P., Kerkmeer, M. C., Kessels, R. P., & Aldenkamp, A. P. (2015). Confirmatory factor analysis of the Dutch version of the Wechsler Memory Scale—Fourth Edition (WMS–IV–NL). *Archives of Clinical Neuropsychology*, 30, 228–235. doi:10.1093/arclin/acv013
- Brennan, A. M., & Gouvier, W. D. (2006). Are we honestly studying malingering? A profile and comparison of simulated and suspected malingerers. *Applied Neuropsychology*, 13, 1–11. doi:10.1207/s15324826an1301_1
- Brennan, A. M., Meyer, S., David, E., Pella, R., Hill, B. D., & Gouvier, W. D. (2009). The vulnerability to coaching across measures of effort. *The Clinical Neuropsychologist*, 23, 314–328. doi:10.1080/13854040802054151
- Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., . . . Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, 20, 419–426. doi:10.1016/j.acn.2005.02.002
- Carlozzi, N. E., Grech, J., & Tulskey, D. S. (2013). Memory functioning in individuals with traumatic brain injury: An examination of the Wechsler Memory Scale—Fourth Edition (WMS–IV). *Journal of Clinical and Experimental Neuropsychology*, 35, 906–914. doi:10.1080/13803395.2013.833178
- Central Office of Statistics for the Netherlands [Centraal Bureau voor Statistiek]. (2011). *Datalevering Enquête Beroepsbevolking*. Retrieved from <http://www.cbs.nl>
- Dandachi-FitzGerald, B., Ponds, R. W., & Merten, T. (2013). Symptom validity and neuropsychological assessment: A survey of practices and beliefs of neuropsychologists in six European countries. *Archives of Clinical Neuropsychology*, 28, 771–783. doi:10.1093/arclin/act073
- Glassmire, D. M., Bierley, R. A., Wisniewski, A. M., Greene, R. L., Kennedy, J. E., & Date, E. (2003). Using the WMS–III Faces subtest to detect malingered memory impairment. *Journal of Clinical and Experimental Neuropsychology*, 25, 465–481. doi:10.1076/jcen.25.4.465.13875
- Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, 6, 218–224. doi:10.1037/1040-3590.6.3.218
- Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychological Review*, 5, 125–148. doi:10.1007/BF02208438
- Haines, M. E., & Norris, M. P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits. *The Clinical Neuropsychologist*, 15, 171–182. doi:10.1076/clin.15.2.171.1891
- Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Corrected Participants1. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, 23, 1093–1129. doi:10.1080/13854040903155063
- Heinly, M. T., Greve, K. W., Bianchini, K. J., Love, J. M., & Brennan, A. (2005). WAIS digit span-based indicators of malingered neurocognitive dysfunction: Classification accuracy in traumatic brain injury. *Assessment*, 12, 429–444. doi:10.1177/1073191105281099
- Hendriks, M. P. H., Bouman, Z., Kessels, R. P. C., & Aldenkamp, A. P. (2014). *Wechsler Memory Scale—Fourth Edition, Dutch Edition* (WMS–IV–NL). Amsterdam: Pearson Assessment.
- Holdnack, J. A., & Drozdick, L. W. (2009). *Advanced clinical solutions for use with WAIS-IV and WMS-IV*. San Antonio, TX: Pearson Education.
- Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: Wiley Interscience.
- Iverson, G. L., & Tulskey, D. S. (2003). Detecting malingering on the WAIS–III. Unusual Digit Span performance patterns in the normal population and in clinical groups. *Archives of Clinical Neuropsychology*, 18, 1–9. doi:10.1016/S0887-6177(01)00176-7
- Jelicic, M., Merckelbach, H., Candel, I., & Geraerts, E. (2007). Detection of feigned cognitive dysfunction using special malingering tests: A simulation study in naïve and coached malingerers. *The International Journal of Neuroscience*, 117, 1185–1192. doi:10.1080/00207450600934697
- Killgore, W. D., & DellaPietra, L. (2000). Using the WMS–III to detect malingering: Empirical validation

- of the rarely missed index (RMI). *Journal of Clinical and Experimental Neuropsychology*, 22, 761–771. doi:10.1076/jcen.22.6.761.960
- Lange, R. T., Iverson, G. L., Sullivan, K., & Anderson, D. (2006). Suppressed working memory on the WMS–III as a marker for poor effort. *Journal of Clinical and Experimental Neuropsychology*, 28, 294–305. doi:10.1080/13803390490918156
- Lange, R. T., Sullivan, K., & Anderson, D. (2005). Ecological validity of the WMS–III rarely missed index in personal injury litigation. *Journal of Clinical and Experimental Neuropsychology*, 27, 412–424. doi:10.1080/13803390490520319
- Langeluddecke, P. M., & Lucas, S. K. (2003). Quantitative measures of memory malingering on the Wechsler Memory Scale–Third Edition in mild head injury litigants. *Archives of Clinical Neuropsychology*, 18, 181–197. doi:10.1016/S0887-6177(01)00195-0
- Larrabee, G. J. (2003). Detection of malingering using atypical performance patterns on standard neuropsychological tests. *The Clinical Neuropsychologist*, 17, 410–425. doi:10.1076/clin.17.3.410.18089
- Larrabee, G. J. (2007). Introduction: Malingering, research designs, and base rates. In G. J. Larrabee (Ed.), *Assessment of malingered neurocognitive deficits* (pp. 3–13). New York, NY: Oxford University Press.
- Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 18, 625–630. doi:10.1017/S1355617712000240
- Larrabee, G. J., & Berry, D. T. R. (2007). Diagnostic classification statistics and diagnostic validity of malingering assessment. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 14–26). New York, NY: Oxford University Press.
- Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Lu, P. H., Rogers, S. A., & Boone, K. B. (2007). Use of standard memory tests to detect suspect effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 128–151). New York, NY: Guilford Press.
- Merckelbach, H., Smeets, T., & Jelicic, M. (2009). Experimental simulation: Type of malingering scenario makes a difference. *Journal of Forensic Psychiatry and Psychology*, 20, 378–386. doi:10.1080/14789940802456686
- Miller, J. B., Millis, S. R., Rapport, L. J., Bashem, J. R., Hanks, R. A., & Axelrod, B. N. (2011). Detection of insufficient effort using the advanced clinical solutions for the Wechsler Memory Scale, Fourth Edition. *The Clinical Neuropsychologist*, 25, 160–172. doi:10.1080/13854046.2010.533197
- Miller, L. J., Ryan, J. J., Carruthers, C. A., & Cluff, R. B. (2004). Brief screening indexes for malingering: A confirmation of Vocabulary minus Digit Span from the WAIS–III and the Rarely Missed Index from the WMS–III. *The Clinical Neuropsychologist*, 18, 327–333. doi:10.1080/13854040490501592
- Mittenberg, W., Azrin, R., Millsaps, C., & Heilbronner, R. (1993). Identification of malingered head injury on the Wechsler Memory Scale–Revised. *Psychological Assessment*, 5, 34–40. doi:10.1037/1040-3590.5.1.34
- Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual*. Windsor: NFER-Nelson.
- Ord, J. S., Greve, K. W., & Bianchini, K. J. (2008). Using the Wechsler Memory Scale–III to detect malingering in mild traumatic brain injury. *The Clinical Neuropsychologist*, 22, 689–704. doi:10.1080/13854040701425437
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20, 33–65.
- Rogers, R. (2007). *Clinical assessment of malingering and deception* (3rd ed.). New York, NY: Guilford Press.
- 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 Clinical and Experimental Neuropsychology*, 19, 43–51. doi:10.1080/01688639708403835
- Schenk, K., & Sullivan, K. A. (2010). Do warnings deter rather than produce more sophisticated malingering? *Journal of Clinical and Experimental Neuropsychology*, 32, 752–762.
- Schmand, B., Lindeboom, J., & Merten, T. (2005). *Amsterdam Short-Term Memory Test: Manual*. Leiden: PITS.
- Schmand, B., Lindeboom, J., & Van Harskamp, F. (1992). *Dutch adult reading test*. Lisse: Swets & Zeitlinger.
- Schwartz, S. M., Gramling, S. E., Kerr, K. L., & Morin, C. (1998). Evaluation of intellect and deficit specific information on the ability to fake memory deficits. *International Journal of Law and Psychiatry*, 21, 261–272. doi:10.1016/S0160-2527(98)00004-1
- Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, 13, 545–561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545
- Suhr, J. A., & Barrash, J. (2007). Performance on standard attention, memory, and psychomotor speed tasks as indicators of malingering. In G. J. Larrabee (Ed.), *Assessment of malingered neuropsychological deficits* (pp. 131–170). New York, NY: Oxford University Press.
- Suhr, J. A., & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology*, 15, 415–424. doi:10.1016/S0887-6177(99)00033-5
- Swihart, A. A., Harris, K. M., & Hatcher, L. L. (2008). Inability of the Rarely Missed Index to identify simulated malingering under more realistic assessment conditions. *Journal of Clinical and Experimental Neuropsychology*, 30, 120–126. doi:10.1080/13803390701249044

- Tan, J. E., Slick, D. J., Strauss, E., & Hultsch, D. F. (2002). How'd they do it? Malingering strategies on symptom validity tests. *The Clinical Neuropsychologist*, *16*, 495–505. doi:10.1076/clin.16.4.495.13909
- Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. North Tonawanda, NY: Multi-Health Systems.
- United Nations Educational, Scientific and Cultural Organisation Institute for Statistics (UNESCO-UIS). (2011). *International Standard Classification of Education (ISCED)*. Montreal: UNESCO-UIS.
- Vickery, C. D., Berry, D. T., Dearth, C. S., Vagnini, V. L., Baser, R. E., Cragar, D. E., & Orey, S. A. (2004). Head injury and the ability to feign neuropsychological deficits. *Archives of Clinical Neuropsychology*, *19*, 37–48. doi:10.1016/S0887-6177(02)00170-1
- Wechsler, D. (1997). *Wechsler Memory Scale-Third Edition (WMS-III)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligent Scale-Fourth Edition*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2009). *Wechsler Memory Scale-Fourth Edition (WMS-IV)*. San Antonio, TX: Pearson Assessment.
- Weinborn, M., Woods, S. P., Nulsen, C., & Leighton, A. (2012). The effects of coaching on the verbal and nonverbal medical symptom validity tests. *The Clinical Neuropsychologist*, *26*, 832–849. doi:10.1080/13854046.2012.686630
- Williams, J. M. (1998). The malingering of memory disorder. In C. R. Reynolds (Ed.), *Detection of malingering during head injury litigation* (pp. 105–132). New York, NY: Plenum Press.
- Young, J. C., Caron, J. E., Baughman, B. C., & Sawyer, R. J. (2012). Detection of suboptimal effort with symbol span: Development of a new embedded index. *Archives of Clinical Neuropsychology*, *27*, 159–164. doi:10.1093/arclin/acr109

Appendix

Questionnaires for the experimental malingerers before and after testing

Semistructured questions before testing (translated)

- (1) Have you experienced any differences in your behaviour or well-being since the accident? What kind of differences have you experienced? When did the (particular complaint) start? Has it worsened over time? How does it interfere with your everyday life?

Questions to be answered after testing (translated)

- (1) How did you try to simulate brain damage?
- (2) How successful do you think you were at simulating brain damage?
Very unsuccessful–very successful
1 2 3 4 5
- (3) Did you search for extra information on brain damage in order to prepare for your role?
Yes No
- (4) If YES: how did you search for extra information?
— I looked up information online.
— I looked up information in books.
— I asked a friend/acquaintance for help.
— Other: _____
- (5) If YES: what information did you use to help you simulate brain damage?