


Performance and Symptom Validity Assessment in Patients with Apathy and Cognitive Impairment

Brechje Dandachi-FitzGerald^{1,*} , Annelien A. Duits², Albert F.G. Leentjens², Frans R.J. Verhey² and Rudolf W.H.M. Ponds²

¹Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD, Maastricht, the Netherlands

²Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, P.O. Box 5800, 6202 AZ, Maastricht, the Netherlands

(RECEIVED February 28, 2019; FINAL REVISION August 13, 2019; ACCEPTED August 21, 2019; FIRST PUBLISHED ONLINE October 29, 2019)

Abstract

Objective: Performance and symptom validity tests (PVTs and SVTs) measure the credibility of the assessment results. Cognitive impairment and apathy potentially interfere with validity test performance and may thus lead to an incorrect (i.e., false-positive) classification of the patient's scores as non-credible. The study aimed at examining the false-positive rate of three validity tests in patients with cognitive impairment and apathy. **Methods:** A cross-sectional, comparative study was performed in 56 patients with dementia, 41 patients with mild cognitive impairment, and 41 patients with Parkinson's disease. Two PVTs – the Test of Memory Malingering (TOMM) and the Dot Counting Test (DCT) – and one SVT – the Structured Inventory of Malingered Symptomatology (SIMS) – were administered. Apathy was measured with the Apathy Evaluation Scale, and severity of cognitive impairment with the Mini Mental State Examination. **Results:** The failure rate was 13.7% for the TOMM, 23.8% for the DCT, and 12.5% for the SIMS. Of the patients with data on all three tests ($n = 105$), 13.5% failed one test, 2.9% failed two tests, and none failed all three. Failing the PVTs was associated with cognitive impairment, but not with apathy. Failing the SVT was related to apathy, but not to cognitive impairment. **Conclusions:** In patients with cognitive impairment or apathy, failing one validity test is not uncommon. Validity tests are differentially sensitive to cognitive impairment and apathy. However, the rule that at least two validity tests should be failed to identify non-credibility seemed to ensure a high percentage of correct classification of credibility.

Keywords: Symptom validity, Mild cognitive impairment, Dementia, Parkinson's disease, Test of Memory Malingering (TOMM), Structured Inventory of Malingered Symptomatology (SIMS), Dot Counting Test (DCT)

INTRODUCTION

Before conclusions are reached about a patient's cognitive abilities and psychological symptoms, it is vital that the assessment validity of the neuropsychological evaluation is established. Assessment validity is defined here as “the accuracy or truthfulness of an examinee's behavioral presentation, self-reported symptoms, or performance on neuropsychological measures” (Bush et al., 2005: p. 420; Larrabee, 2015). Phrased otherwise, assessment validity pertains to the credibility of the clinical presentation. To determine the assessment validity, specific tools have been developed: (1) symptom

validity tests (SVTs) that measure whether a person's complaints reflect his or her true experience of symptoms and (2) performance validity tests (PVTs) that measure whether a person's test performance is reflective of the actual cognitive ability (Larrabee, 2015; Merten et al., 2013). In the validation of these validity tests the cut-scores to classify non-credible presentation are traditionally set at a specificity of $\geq .90$ to safeguard against a false-positive classification of a credible presentation of symptoms and cognitive abilities as non-credible (Boone, 2013, Chapter 2). Also, a “two-test failure rule” (i.e., any pairwise failure on validity tests) has been recommended as a criterion to identify non-credible presentations, particularly in samples with a low base rates of non-credible symptom reports and cognitive test performance (Larrabee, 2008; Lippa, 2018; Victor, Boone, Serpa, Buchler, & Ziegler, 2009). The post-test probability that a validity test failure is due to a

*Correspondence and reprint requests to: Brechje Dandachi-FitzGerald, Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, the Netherlands. Phone: 0031-43-3881948. E-mail: b.fitzgerald@maastrichtuniversity.nl

non-credible presentation depends on the base rate (Rosenfeld et al., 2000). For example, failing one validity test with a specificity of .90 and sensitivity of .80, in a sample with a base rate of .10, the post-test probability that a failure is due to non-credible responding is only .47. Therefore, if the base rate is not taken into account, one runs the risk of incorrect interpretation of the validity test failure in low base rate samples.

The usefulness of PVTs and SVTs has been demonstrated in numerous studies, in particular in patients in which external incentives are present (Bianchini, Curtis, & Greve, 2006; Boone, 2013: Chapter 2), but also in psychological assessments of clinically referred patients (Dandachi-FitzGerald, van Twillert, van de Sande, van Os, & Ponds, 2016; Locke, Smigielski, Powell, & Stevens, 2008). Because of these findings, assessment validity has now become an integral part of a neuropsychological evaluation (Hirst et al., 2017).

Motivational deficiency due to cerebral pathology is one of several potential limiting, but insufficiently researched, factors. Especially, concerns have been raised about the potential influence of apathy on PVTs. An apathetic patient may not be able to invest sufficient effort into testing, and consequently be wrongfully classified by the PVT as non-credible (Bigler, 2015).

Cognitive impairments are another potential limiting factor. PVTs, for example, still require a minimum of cognitive abilities to be able to perform normal. Likewise, the SVTs require a minimum level of reading and verbal comprehension to grasp the items. Despite that the role of cognitive impairment has been studied more than the role of motivational deficiencies, the scientific body of knowledge on this topic is still limited.

The current study aims to extend previous studies on the critical limits of validity test performance. To the best of our knowledge no study yet has examined the accuracy of both types of tests – PVTs and SVTs – in a relatively homogenous and large sample of patients with cognitive impairment and apathy. We hypothesize that in this sample the false-positive rate of the individual validity test would be unacceptably high (i.e., >10%), and related to both cognitive impairment and apathy. We further hypothesize that the PVTs require relatively more motivational effort and cognitive abilities to perform than the SVTs, and are therefore more susceptible to apathy and cognitive impairment. In addition, we examine the accuracy of the “two-test failure rule” to identify non-credible presentations of symptoms and cognitive abilities. We hypothesize that the false-positive rate of this classification rule will be within acceptable limits (i.e., <10%), and thereby constitute a superior approach to determine the assessment validity within patient samples with raised levels of cognitive impairment and apathy.

METHOD

Sample

The study followed a cross-sectional, between groups, design. All patients were clinically evaluated at the

Maastricht University Medical Centre. Patients with Parkinson's disease (PD) were referred by the neurologist or psychiatrist for neuropsychological assessment. The referral reason was either to generally evaluate their cognitive functioning or as part of a preoperative screening to determine their eligibility for deep brain stimulation (DBS). No patients with already-implanted DBS systems were included. Patients with mild cognitive impairment (MCI) or dementia were seen at the memory clinic of the hospital for diagnostic evaluation. The final clinical diagnosis was made in a multi-disciplinary team and based on multiple sources of information such as third-party information (e.g., a spouse or child was interviewed on the presence of symptoms and impairment in daily functioning), neuropsychological assessment, psychiatric and neurological evaluation, and brain imaging (e.g., MRI scan). The patients evaluated at the memory clinic were participating in a larger research project following the course of cognitive decline (Aalten et al., 2014).

Inclusion criteria were: (1) a clinical diagnosis of MCI or dementia based on the National Institute of Neurological and Communicative Disorders and Stroke – Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) criteria (Albert et al., 2011; Dubois et al., 2007) or a diagnosis of PD according to the Queens Square Brain Bank criteria (Lees, Hardy, & Revesz, 2009); (2) mental competency to give informed consent; (3) native Dutch speaker; (4) a minimum of 8 years of formal schooling, and no history of mental retardation. Exclusion criteria were: (1) comorbid major depressive disorder as defined by the Diagnostic and Statistical Manual of Mental Disorders -IV (*DSM-IV*) criteria (American Psychiatric Association, 2000); (2) other neurological diseases (e.g., epilepsy and multiple sclerosis); (3) a history of acquired brain injury (e.g., cerebrovascular accident and cerebral contusion); (4) involvement in juridical procedures (e.g., litigation).

Measures

Performance validity tests

Test of Memory Malingering (TOMM). The TOMM is a 50-item forced choice picture recognition task (Tombaugh, 2006). We used the cut-score of 45 on the second recognition trial. In the validation studies, this score was associated with a specificity of 1.00, and a sensitivity of .90 in a sample of healthy controls either instructed to malingering brain injury or to perform honestly (Tombaugh, 2006). Further, the validation studies showed a specificity of .92 in a sample of patients with dementia ($N=37$), and of .97 in a clinical sample of patients with cognitive impairment, aphasia, and traumatic brain injury ($N=108$) (Tombaugh, 2006).

Dot Counting Test (DCT). The DCT requires the participant to count grouped and ungrouped dots as quickly as possible. An effort index (i.e., E-score) is calculated from the response time and number of errors. We used the standard cut-score of 17, which in validation studies was associated with a

specificity of .90 and a sensitivity of .79 (Boone et al., 2002). We also applied the recommended cut-score for mild dementia (i.e., E-score ≥ 22). In a sample of 16 patients with mild dementia this cut-score was associated with a specificity of .94 and a sensitivity of .62.

Symptom validity test

Structured Inventory of Malingered Symptomatology (SIMS). The SIMS is a 75-item self-report questionnaire addressing bizarre and/or rare symptoms that are rated on a dichotomous (yes-no) scale. A study of the Dutch research version of the SIMS with a group of 298 participants revealed a specificity of .98 and sensitivity of .93 with a cut-score of 16 (Merckelbach & Smith, 2003). A recent meta-analysis recommended to raise the cut-score to 19 in clinical samples (van Impelen, Merckelbach, Jelicic, & Merten, 2014). Therefore, we will provide information of both the standard cut-score of >16 and the recommend cut-score of >19 .

Clinical measures

Apathy Evaluation Scale (AES). The AES is a scale that consists of 18 items phrased as questions that are to be answered on a four-point Likert scale (Marin, Biedrzycki, & Firinciogullari, 1991). The AES is a reliable and valid measure to screen for and to assess the severity of apathy in PD and dementia (Clarke et al., 2007; Leentjens et al., 2008). We used the clinician rated version of the AES, and a cut-score of 38/39 to identify clinical cases of apathy (Pluck & Brown, 2002).

Mini Mental State Examination (MMSE). As a global index for cognition, we used the MMSE (Folstein, Folstein, & McHugh, 1975), with the standard cut-score of <24 to identify clinical cases of cognitive impairment.

Procedure

The study was approved by the Medical Ethical Committee of Maastricht University Medical Centre (MEC 10-3-81). All patients received an information letter and informed consent form. Patients with PD were informed about the study by the nurse practitioner or the referring psychiatrist or neurologist. Patients with MCI or dementia were participating in a larger research project for which a generic information brochure for participants was constructed in order to prevent confusion and information overload (Aalten et al., 2014). For all patients there was a minimum of 1 week reflection time before entering the study. When patients provided informed consent, the validity tests were included in the neuropsychological test battery. The AES and the MMSE were already part of the test battery.

Data Analysis

Data were checked for errors, missing data, outliers, and score distributions. Missing data were excluded pairwise per

analysis. Outliers were not removed from the dataset. First, descriptive statistics were calculated for the total sample and for each diagnostic category. Differences between the three diagnostic groups on the dependent variables were examined with one-way ANOVA and *post-hoc* Bonferroni corrected pairwise comparisons (age), Chi square analysis (gender, education), and with Kruskal-Wallis tests with *post-hoc* Dunn-bonferroni pairwise comparisons (AES and MMSE). Second, the percentage of participants failing the TOMM, DCT, and SIMS was calculated. Before determination of the accuracy of the “two test failure rule,” the nonparametric bivariate correlation between the validity tests was calculated to make sure that the individual tests do not highly correlate in order to avoid inflation of type I error (Larrabee, 2008). We calculated the number of validity tests failed. All calculations were conducted twice, with the standard cut-score on the SIMS (>16) and DCT (≥ 17), and with the adjusted cut-scores on the SIMS (>19) and DCT (≥ 22). Third, with Fisher’s exact tests we compared the frequency of TOMM, DCT, and SIMS failures in the group of patients with (i.e., AES ≥ 39) and without apathy (i.e., AES < 39), and in the group of patients with (i.e., MMSE < 24) and without a significant degree of cognitive impairment (i.e., MMSE ≥ 24). Analyses were performed with SPSS version 23.

RESULTS

Demographics and Psychometrics

In total, 145 patients were assessed of whom 7 were excluded because of too many missing data. In the final sample ($N = 138$) there were 8 (5.8%) missing records for the AES, 14 (10.1%) for the TOMM, 16 (11.6%) for the DCT, and 18 (13.0%) for the SIMS. Missing item scores on the SIMS were imputed with the mean item score of that person when there were no more than 15 item scores in total and no more than 3 item scores per subscale missing. There were three extreme outliers on the TOMM, and no extreme outliers on the SIMS extrapolated total score, and on the DCT. Most variables, that is, the TOMM, DCT, SIMS, MMSE, and the AES were not normally distributed.

As can be seen in Table 1, the sample consisted of 138 participants (56.5% male) with a mean age of 71.9 years ($SD = 9.1$, range: 48–89). In total, 56 patients were diagnosed with dementia, 41 patients with MCI, and 41 patients with PD. Of the 41 PD patients, 9 (22%) were referred for a neuropsychological evaluation as part of the preoperative screening to determine the eligibility for DBS. Gender [$\chi^2(2) = .62$, $p = .74$], and education [$\chi^2(4) = 5.19$, $p = .27$] did not differ significantly between the three diagnostic groups (i.e., dementia, MCI, and PD). There was a group difference in age [$F(2, 135) = 35.10$, $p < .00$]. The group of patients with PD was younger than the two other diagnostic groups. The mean rank MMSE score [$H(2) = 59.10$, $p < .01$] and the mean rank AES score [$H(2) = 24.51$, $p < .01$] differed between the three diagnostic groups. As expected, the mean rank MMSE

Table 1. Demographics and psychometric data

	Total sample (<i>N</i> = 138)	1. Dementia (<i>n</i> = 56)	2. MCI (<i>n</i> = 41)	3. PD (<i>n</i> = 41)	<i>p</i>	Significant group comparisons
Age (mean, <i>SD</i>) ^a	71.9 (9.1)	74.8 (7.1)	78.0 (7.2)	63.7 (8.1)	.00	3 < 1, 2
Gender (% male) ^b	55.8	51.8	58.5	58.5	.74	
Education (%) ^{b*}					.27	
Low	13.8	19.6	4.9	14.6		
Medium	61.6	58.9	70.7	56.1		
High	24.6	21.4	24.4	29.3		
CDR (mean, <i>SD</i>)		.6 (.4)	.5		n/a	
Hoehn & Yahr (mean, <i>SD</i>)				2.0 (.6)	n/a	
AES						
Total score (median, IQR) ^c	32 (15)	39 (16)	33 (15.5)	26 (7.5)	.00	1 > 2, 3
Cut-of > 38 (%) ^b	33.8	54.9	28.9	12.2	.00	
MMSE						
Total score (median, IQR) ^c	26 (4.0)	24 (3.8)	26 (3.5)	29 (2.0)	.00	1 < 2 < 3
Cut-of < 24 (%) ^b	24.3	46.4	17.1	.0	.00	
TOMM 2 (median, IQR) ^c	49.0 (2.0)	48.0 (5.0)	49.5 (3.0)	50.0 (.8)	.00	1 < 2 < 3
DCTE-score (median, IQR) ^c	11.8 (7.2)	14.3 (7.1)	11.8 (7.8)	9.3 (4.0)	.00	1 > 3
SIMS (median, IQR) ^c	9.1 (8.0)	10.7 (7.8)	8.3 (7.0)	9.0 (8.8)	.13	

Notes: MCI = mild cognitive impairment; PD = Parkinson's disease; CDR = clinical dementia rating; AES = Apathy Evaluation; MMSE = Mini Mental State Exam; TOMM 2 = Test of Memory Malingering second recognition trial; DCT E-score = Dot Counting Test Effort-score; SIMS = Structured Inventory of Malingered Symptomatology; n/a = not applicable; ^aone-way ANOVA; ^b χ^2 ; ^cKruskall-Wallis test; *Low = at most primary education; medium = junior vocational training, and high = senior vocational or academic training (Van der Elst et al., 2005).

score was lowest in the patients with dementia, followed by the group of patients with MCI, and then the group of patients with PD (all $ps < .01$). Patients with dementia had a higher mean rank AES score than the group of patients with MCI ($p < .05$), and the group of patients with PD ($p < .01$), who did not differ from each other ($p = .12$). Regarding the SVTs, the mean rank TOMM score was the lowest in the group of patients with dementia, followed by the group of patients with MCI, and then the group of patients with PD (all $ps < .05$). The mean rank DCT E-score was higher in the group of patients with dementia than the group of patients with PD ($p < .01$). The mean rank SIMS score did not significantly differ between the three groups.

False-Positive Rate of the Individual Validity Tests

Table 2 shows that 17 of the 124 participants who completed the TOMM scored below the cut-score of 45 on the second recognition trial. In total, 120 participants completed the SIMS, of whom 12 and 15 patients obtained a deviant score using a cut-score of >19 or >16, respectively. As for the DCT, 29 of the 122 participants who completed the DCT scored above the standard cut-score of 17, whereas only 7 participants failed when using the adjusted cut-score of 22.

False-Positive Rate of the "Two-Test Failure Rule"

Table 3 shows that the TOMM, DCT, and SIMS were at most weakly correlated. Of the 105 patients who completed all three validity tests, most (83.8%) passed, and none failed all three tests (Table 2). A substantial number of participants

Table 2. Failure rate (%) of the validity tests

	<i>N</i>	Standard cut-scores ^a	Adjusted cut-scores ^b
TOMM 2	124	13.7	13.7
DCT E-score	122	23.8	5.7
SIMS	120	12.5	10.0
Number of tests failed	105		
0		67.6	83.8
1		27.6	13.3
2		4.8	2.9
3		.0	.0

Notes: TOMM 2 = Test of Memory Malingering recognition trial 2; DCT E-score = Dot Counting Test Effort-score; SIMS = Structured Inventory of Malingered Symptomatology; ^aTOMM recognition trial 2 < 45, SIMS total score > 16, DCT E-score ≥ 17 ; ^bTOMM recognition trial 2 < 45, SIMS total score > 19, DCT E-score ≥ 22 .

failed on one validity test: 27.6% with the standard cut-scores, and 13.3% with the adjusted cut-scores on the SIMS and DCT. By contrast, only 3 (adjusted cut-scores) to 5 (standard cut-scores) patients out of 105 patients failed on two validity tests. Thus, the "two-failure rule" correctly classified 95–97% as credible.

Relationship between Validity Test Failure, Apathy, and Cognitive Impairment

Fisher's Exact Test (FET) revealed no significant difference between the group of patients with and without apathy in

Table 3. Spearman's rho correlation between the validity tests and clinical measures

	TOMM 2	DCT E-score	SIMS	MMSE	AES
TOMM 2	–				
DCT E-score	-.20*	–			
SIMS	-.27**	.20*	–		
MMSE	.47**	-.37**	-.23*	–	
AES	-.31**	.09	.27**	-.32**	–

Notes: SVTs = symptom validity tests; TOMM 2 = Test of Memory Malingered recognition trial 2; DCT E-score = Dot Counting Test Effort score; SIMS = Structured Inventory of Malingered Symptomatology; MMSE = Mini Mental State Examination; AES = Apathy Evaluation Scale; * $p < .05$; ** $p < .01$.

failure rate on the TOMM (19% vs. 9%; $p = .14$), and on the DCT (8% vs. 4%; $p = .38$) (see Table 4). On the SIMS, however, the failure rate of patients with apathy was much higher than those without apathy (21% vs. 4%; $p = .01$) (see Table 4). Regarding cognitive impairment, the opposite pattern occurred. Patients with cognitive impairment on the MMSE failed more frequently than those without cognitive impairment on the TOMM (31% vs. 9%; $p = .01$), and on the DCT (22% vs. 1%; $p = .00$). Patients with and without cognitive impairment did not differ in the frequency of failure on the SIMS (17% vs. 9%; $p = .26$). Using the standard cut-scores for the DCT (≥ 17) and SIMS (> 16) did not change the pattern of significant and non-significant statistical test results.

DISCUSSION

Our study shows that the failure rate on the individual validity tests ranged between 12.6% and 23.8%. Raising the cut-scores for the SIMS and the DCT yielded lower false-positive failure rates. In particular, the DCT retained a satisfactory accuracy with adjustment of the cut-score, namely .94 correct classification of valid assessment results. The SIMS attained with the recommended cut-score of > 19 an accuracy of .90.

We found a differential sensitivity of the PVTs and SVTs for cognitive impairment and apathy. Failure on the PVTs, the

TOMM and the DCT, was related to cognitive impairment: the accuracy remained above .90 in patients who scored equal or higher than 24 on the MMSE, whereas it dropped to .69 for the TOMM, and to .78 for the DCT in patients with a MMSE score lower than 24. This finding is in line with other studies that found a lower MMSE score to be related to PVT failure in genuine patient samples (McGuire, Crawford, & Evans, 2019; Merten, Bossink, & Schmand, 2007). Comparable to our findings is a recent study that examined failure on a PVT, the Word Memory Test, in a sample of 30 patients with PD and essential tremor who were neuropsychologically evaluated as part of the screening for the indication for DBS (Rossetti, Collins, & York, 2018). These patients had a motivation to perform well, since clinically relevant cognitive impairment might jeopardize their candidacy for the surgical intervention of DBS. The accuracy of the PVT in this sample was .90. Similarly, in patient samples with other neurological disorders such as sickle cell disease patients (Dorociak, Schulze, Piper, Molokie, & Janecek, 2018), and Huntington disease (Sieck, Smith, Duff, Paulsen, & Beglinger, 2013), the accuracy of stand-alone PVT was $\geq .90$. So, the evidence points toward a critical limit of moderate to severe cognitive impairment for PVTs. Note that merely the presence of a neurological condition in itself does not automatically mean that the PVTs are not valid anymore. Also, the PVTs have been shown to differ in their sensitivity to cognitive impairment. In our study the DCT had a higher failure rate than the TOMM had, which might be explained by difference in load on working memory of both tests (Merten et al., 2007).

Failure on the PVTs was not related to apathy. The false-positive rate of the TOMM and the DCT did not significantly differ between patients with and without clinical levels of apathy. Thus our findings are not supportive of the notion that apathy might lead to an increased risk of false-positive classification of PVTs.

Intriguingly, failure on the SVT, the SIMS, showed the opposite pattern from the PVTs: failure was related to apathy, but not to cognitive impairment. To the best of our knowledge, this is the first study to examine the limits of an SVT in a sample of patients with cognitive impairment and apathy.

Table 4. Relation validity test failure with apathy and cognitive impairment

	Apathy		p	Cognitive impairment		p
	No (AES ≤ 38)	Yes (AES > 38)		No (MMSE ≥ 24)	Yes (MMSE < 24)	
% failure (number of patients)						
TOMM 2 (< 45)	8.9 (7 out of 79)	18.9 (7 out of 37)	.14	8.6 (8 out of 93)	31.0 (9 out of 29)	.01
DCT E-score (≥ 22)	3.9 (3 out of 78)	8.3 (3 out of 36)	.38	1.1 (1 out of 93)	22.2 (6 out of 27)	.00
SIMS (> 19)	4.0 (3 out of 76)	21.1 (8 out of 38)	.01	8.5 (8 out of 94)	16.7 (4 out of 24)	.27

Notes: AES = Apathy Evaluation Scale; MMSE = Mini Mental State Exam; TOMM 2 = Test of Memory Malingered recognition trial 2; DCT E-score = Dot Counting Test Effort-score; SIMS = Structured Inventory of Malingered Symptomatology.

These first, and therefore preliminary, results are reassuring that cognitive impairment in patients with dementia, MCI, or PD does not directly lead to an inability to understand and answer the questions of the SIMS. This suggests that in neuropsychological assessments the SIMS is an instrument that can be used to measure the validity of self-reported symptoms, even in neurological populations. Although in need of replication, apathy might be a critical limit for the SIMS.

The second hypothesis is supported by the data. The decision rule that at least two validity tests should be failed for the determination of non-credibility ensured a high correct rating of the assessment validity (.95). Note that this holds true with using the standard cut-scores on the SIMS and DCT. Moreover, the “two-test failure rule” with the standard cut-scores on the validity tests yielded fewer false positives than the individual test with the adjusted cut-score. Thus, our findings suggest that the “two-test failure rule” is superior to adjustment of cut-scores when it comes to maintaining adequate accuracy in patient samples with genuine psychopathology. This is an important finding, because customizing cut-scores for different patient groups poses specific problems. In general, a higher specificity of an individual validity test comes with the price of a lower sensitivity. For example, the sensitivity of the DCT to correctly classify a person who is feigning cognitive impairment is substantially lower with a cut-score of 22 (i.e., .62) than with the standard cut-score of 17 (i.e., .79) (Boone et al., 2002). Further, it is often not evident beforehand whether cognitive impairment or psychopathology is present. The reason why psychological assessment is requested is actually to objectify the presence of cognitive impairment and psychopathology. To avoid hindsight bias, the clinician will have to select the appropriate cut-score before psychological testing. It is questionable whether this approach is feasible in clinical practice. Relatedly, selection of the cut-score for the validity test in an individual psychological assessment leaves the door wide open to discussion about the appropriateness of the chosen cut-score, and consequently of the classification of the assessment validity. This complicates the diagnostic assessment and clinical decision making.

Our findings on the accuracy of “two-test failure rule” to classify non-credible test performance compare favorably with those of Davis’s (Davis, 2018) with a very large sample of older patients diagnosed with normal cognition, cognitive impairment, MCI, and dementia. In that study 13.2% of the patients with MCI and 52.8% of the patients with dementia failed two validity tests, and their performance was, therefore, wrongfully classified by the “two-test failure rule” as non-credible. A plausible explanation for these divergent findings relates to the type of validity measures used. In the study of Davis five embedded indicators of performance validity were used, with a higher average intercorrelation between measures (i.e., .41) than the freestanding validity tests in our study (i.e., .22). The “two-test failure rule” is based on the chaining of likelihood ratios (Larrabee, 2008) and can be applied only when the different validity measures are independent. Also,

the embedded indicators are generally more susceptible to cognitive impairment than the freestanding validity tests (Merten et al., 2007; Loring et al., 2016).

Our study is not without limitations. Importantly, there is no gold standard for determination of assessment validity available. Therefore, strictly speaking we cannot exclude the possibility that a failure on the validity test was actually a true positive score. This is not only a limitation of our study but of the field of assessment validity research in general. However, we excluded patients who held external incentives, and the clinical diagnosis was based on multiple sources of information besides neuropsychological assessment (e.g., third-party information on daily functioning and brain imaging). Therefore, we are confident that the assessment validity of the neuropsychological evaluation can be assumed in the vast majority of the patients. Then, our findings only hold for the constellation of the validity tests we used and cannot be automatically extrapolated to other – combinations of – validity tests. Therefore, replication as well as research of different validity tests is called for.

Of note, the two-test failure rule based on chaining of likelihood ratios seems appropriate in samples with low base rates of non-credible responding (Larrabee, 2008; Lippa, 2018). In specific samples, for example, in patients with no or only mild neurological impairment seen in a forensic context, a heightened base rate of non-credible clinical presentations might be expected, and here even a single validity test failure should raise concern about the validity of the obtained test data (Proto et al., 2014; Rosenfeld et al., 2000).

In conclusion, the results show that in a sample of older patients diagnosed with MCI, dementia, or PD, failing on one validity test is not uncommon, and that the validity tests are differentially sensitive to cognitive impairment and apathy. However, also in this sample, it remains rare to score abnormal on two independent validity tests. Therefore, using the “two test failure rule” is probably better to identify non-credibility. More generally, the implication is that in clinical assessments, validity tests can be used without running an unacceptably high risk of incorrectly classifying a genuine presentation of symptoms and cognitive test scores as non-credible. The study findings can serve as contextual background information for a psychological assessment, in which the failure on two independent validity tests in an examinee who does not have moderate to severe cognitive impairment or clinical relevant apathy due a neurological condition can most likely be classified as an invalid assessment.

ACKNOWLEDGEMENT

We thank Fleur Smeets-Prompers for her contribution to the collection of data.

FINANCIAL SUPPORT

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

CONFLICT OF INTEREST

The authors have nothing to disclose.

PROVENANCE AND PEER REVIEW

Not commissioned; externally peer reviewed.

CONTRIBUTORSHIP STATEMENT

BDF contributed to the study design, data collection, and data analysis, and wrote the first draft of the article; RP commented on earlier drafts and contributed to the design, data collection, and analysis; AD, AL, and FV were involved in the data collection and commented on earlier drafts.

DATA SHARING STATEMENT

The data underlying this study have been uploaded to the Maastricht University Dataverse and are accessible using the following link: <https://hdl.handle.net/10411/8LKJF9>

REFERENCES

- Aalten, P., Ramakers, I.H.G.B., Biessels, G.J., Deyn, P., Koek, H.L., OldeRikkert, M.G.M., Oleksik, A.M., Richard, E., Smits, L.L., van Swieten, J.C., Teune, L.K., van der Lugt, A., Barkhof, F., Teunissen, C.E., Rozendaal, N., Verhey, F.R., & Flier, W. (2014). The Dutch Parelinoer Institute – Neurodegenerative diseases; methods, design and baseline results. *BMC Neurology*, *14*(1), 254. doi: [10.1186/s12883-014-0254-4](https://doi.org/10.1186/s12883-014-0254-4)
- Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., Gamst, A., Holtzman, D.M., Jagust, W.J., Petersen, R.C., Snyder, P.J., Carrillo, M.C., Thies, B., & Phelps, C.H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279. doi: [10.1016/j.jalz.2011.03.008](https://doi.org/10.1016/j.jalz.2011.03.008)
- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV* (4th ed., text rev.). Washington, DC: Author.
- Bianchini, K.J., Curtis, K.L., & Greve, K.W. (2006). Compensation and malingering in Traumatic Brain Injury: A dose-response relationship? *The Clinical Neuropsychologist*, *20*(4), 831–847. doi: [10.1080/13854040600875203](https://doi.org/10.1080/13854040600875203)
- Bigler, E.D. (2015). Neuroimaging as a biomarker in symptom validity and performance validity testing. *Brain Imaging and Behavior*, *9*(3), 421–444. doi: [10.1007/s11682-015-9409-1](https://doi.org/10.1007/s11682-015-9409-1)
- Boone, K.B. (2013). Assessment of neurocognitive symptom validity, In K.B. Boone (Ed.), *Clinical practice of forensic neuropsychology* (pp. 22–72). New York: The Guilford Press.
- Boone, K.B., Lu, P., Back, C., King, C., Lee, A., Philpott, L., Shamieh, E., & Warner-Chacon, K. (2002). Sensitivity and specificity of the Rey Dot Counting Test in patients with suspect effort and various clinical samples. *Archives of Clinical Neuropsychology*, *17*(7), 625–642. doi: [10.1016/S0887-6177\(01\)00166-4](https://doi.org/10.1016/S0887-6177(01)00166-4)
- Bush, S.S., Ruff, R.M., Tröster, A.I., Barth, J.T., Koffler, S.P., Pliskin, N.H., Reynolds, C.R., & Silver, C.H. (2005). Symptom validity assessment: Practice issues and medical necessity NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. doi: [10.1016/j.acn.2005.02.002](https://doi.org/10.1016/j.acn.2005.02.002)
- Clarke, D.E., Reekum, R.V., Simard, M., Streiner, D.L., Freedman, M., & Conn, D. (2007). Apathy in dementia: An examination of the psychometric properties of the Apathy evaluation scale. *Journal of Neuropsychiatry*, *19*(1), 57–64. doi: [10.1176/appi.neuropsych.19.1.57](https://doi.org/10.1176/appi.neuropsych.19.1.57)
- Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y., & Ponds, R.W.H.M. (2016). Poor symptom and performance validity in regularly referred hospital outpatients: Link with standard clinical measures, and role of incentives. *Psychiatry Research*, *239*, 47–53. doi: [10.1016/j.psychres.2016.02.061](https://doi.org/10.1016/j.psychres.2016.02.061)
- Davis, J.J. (2018). Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. *Journal of Clinical and Experimental Neuropsychology*, *40*(10), 1013–1021. doi: [10.1080/13803395.2018.1472221](https://doi.org/10.1080/13803395.2018.1472221)
- Dorociak, K.E., Schulze, E.T., Piper, L.E., Molokie, R.E., & Janecek, J.K. (2018). Performance validity testing in a clinical sample of adults with sickle cell disease. *The Clinical Neuropsychologist*, *32*(1), 81–97. doi: [10.1080/13854046.2017.1339830](https://doi.org/10.1080/13854046.2017.1339830)
- Dubois, B., Feldman, H.H., Jacova, C., DeKosky, S.T., Barberger-Gateau, P., Cummings, J., Delacourte, A., Galasko, D., Gauthier, S., Jicha, G., Meguro, K., O'Brien, J., Pasquier, F., Robert, P., Rossor, M., Salloway, S., Stern, Y., Visser, P.J., & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria. *The Lancet Neurology*, *6*(8), 734–746. doi: [10.1016/S1474-4422\(07\)70178-3](https://doi.org/10.1016/S1474-4422(07)70178-3)
- Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198.
- Hirst, R.B., Han, C.S., Teague, A.M., Rosen, A.S., Gretler, J., & Quittner, Z. (2017). Adherence to validity testing recommendations in neuropsychological assessment: A survey of INS and NAN members. *Archives of Clinical Neuropsychology*, *32*(4), 456–471. doi: [10.1093/arclin/acx009](https://doi.org/10.1093/arclin/acx009)
- Larrabee, G.J. (2008). Aggregation across multiple indicators improves the detection of malingering: Relationship to likelihood ratios. *The Clinical Neuropsychologist*, *22*(4), 666–679. doi: [10.1080/13854040701494987](https://doi.org/10.1080/13854040701494987)
- Larrabee, G.J. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, *70*, 779–788. doi: [10.1037/a0039835](https://doi.org/10.1037/a0039835)
- Leentjens, A.F.G., Dujardin, K., Marsh, L., Martinez-Martin, P., Richard, I.H., Starkstein, S.E., Weintraub, D., Sampaio, C., Poewe, W., Rascol, O., Stebbins, G.T., & Goetz, C.G. (2008). Apathy and anhedonia rating scales in Parkinson's disease: Critique and recommendations. *Movement Disorders*, *23*(14), 2004–2014. doi: [10.1002/mds.22229](https://doi.org/10.1002/mds.22229)
- Lees, A.J., Hardy, J., & Revesz, T. (2009). Parkinson's disease. *The Lancet*, *373*(9680), 2055–2066. doi: [10.1016/S0140-6736\(09\)60492-X](https://doi.org/10.1016/S0140-6736(09)60492-X)
- Lippa, S.M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(4), 391–421. doi: [10.1080/13854046.2017.1406146](https://doi.org/10.1080/13854046.2017.1406146)
- Locke, D.E., Smigielski, J.S., Powell, M.R., & Stevens, S.R. (2008). Effort issues in post-acute outpatient acquired brain injury rehabilitation seekers. *NeuroRehabilitation*, *23*(3), 273–281.

- Loring, D.W., Goldstein, F.C., Chen, C., Drane, D.L., Lah, J.J., Zhao, L., & Larrabee, G.J. (2016). False-positive error rates for reliable digit span and auditory verbal learning test performance validity measures in amnesic mild cognitive impairment and early Alzheimer disease. *Archives of Clinical Neuropsychology*, *31*, 313–331. doi: [10.1093/arclin/acw014](https://doi.org/10.1093/arclin/acw014)
- Marin, R.S., Biedrzycki, R.C., & Firinciogullari, S. (1991). Reliability and validity of the apathy evaluation scale. *Psychiatry Research*, *38*(2), 143–162. doi: [10.1016/0165-1781\(91\)90040-V](https://doi.org/10.1016/0165-1781(91)90040-V)
- McGuire, C., Crawford, S., & Evans, J.J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114–131. doi: [10.1093/arclin/acy012](https://doi.org/10.1093/arclin/acy012)
- Merckelbach, H., & Smith, G.P. (2003). Diagnostic accuracy of the Structured Inventory of Malingered Symptomatology (SIMS) in detecting instructed malingering. *Archives of Clinical Neuropsychology*, *18*(2), 145–152. doi: [10.1016/S0887-6177\(01\)00191-3](https://doi.org/10.1016/S0887-6177(01)00191-3)
- Merten, T., 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*(3), 308–318. doi: [10.1080/13803390600693607](https://doi.org/10.1080/13803390600693607)
- Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, B.A., Santamaría, P., & González-Ordí, H. (2013). Symptom validity assessment in European countries: Development and state of the art. *Clínica y Salud*, *24*, 129–138.
- Pluck, G.C., & Brown, R.G. (2002). Apathy in Parkinson's disease. *Journal of Neurology, Neurosurgery & Psychiatry*, *73*(6), 636. doi: [10.1136/jnnp.73.6.636](https://doi.org/10.1136/jnnp.73.6.636)
- Proto, D.A., Pastorek, N.J., Miller, B.I., Romesser, J.M., Sim, A.H., & Linck, J.F. (2014). The dangers of failing one or more performance validity tests in individuals claiming mild traumatic brain injury-related postconcussive symptoms. *Archives of Clinical Neuropsychology*, *29*(7), 614–624. doi: [10.1093/arclin/acu044](https://doi.org/10.1093/arclin/acu044)
- Rosenfeld, B., Sands, S.A., & Van Gorp, W.G. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology*, *15*(4), 349–359. doi: [10.1093/arclin/15.4.349](https://doi.org/10.1093/arclin/15.4.349)
- Rossetti, M.A., Collins, R.L., & York, M.K. (2018). Performance validity in deep brain stimulation candidates. *Archives of clinical neuropsychology*, *33*(4), 508–514. doi: [10.1093/arclin/acx081](https://doi.org/10.1093/arclin/acx081)
- Sieck, B.C., Smith, M.M., Duff, K., Paulsen, J.S., & Beglinger, L.J. (2013). Symptom validity test performance in the Huntington disease clinic. *Archives of Clinical Neuropsychology*, *28*(2), 135–143. doi: [10.1093/arclin/acs109](https://doi.org/10.1093/arclin/acs109)
- Tombaugh, T.N. (2006). *Test of Memory Malingering: TOMM*. New York: Pearson Assessments.
- van Impelen, A., Merckelbach, H., Jelicic, M., & Merten, T. (2014). The Structured Inventory of Malingered Symptomatology (SIMS): A systematic review and meta-analysis. *The Clinical Neuropsychologist*, *28*(8), 1336–1365. doi: [10.1080/13854046.2014.984763](https://doi.org/10.1080/13854046.2014.984763)
- 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.
- Victor, T., Boone, K., Serpa, J.G., Buehler, J., & Ziegler, E. (2009). Interpreting the meaning of multiple symptom validity test failure. *The Clinical Neuropsychologist*, *23*(2), 297–313. doi: [10.1080/13854040802232682](https://doi.org/10.1080/13854040802232682)