Psychological Injury and Law (2024) 17:424–438
https://doi.org/10.1007/s12207-024-09519-2

ORIGINAL RESEARCH

How Psychologists Communicate about Symptom and Performance Validity Testing in Their Reports: Room for Improvement

Brechje Dandachi-FitzGerald[1,2] · Maarten J. V. Peters[1] · Thomas Merten[3] · Jeroen J. Roor[4,5] · Rudolf W. H. M. Ponds[4,6] · Harald Merckelbach[1]

Received: 2 May 2024 / Accepted: 27 August 2024 / Published online: 12 September 2024

Abstract

This archival study sought to determine whether psychological reports adequately communicate the results of Symptom Validity Tests (SVTs) and Performance Validity Tests (PVTs). We collected reports from a pool of 469 clinical psychological and neuropsychological assessments conducted across five Dutch hospitals. To be included, the administered SVT (i.e., Structured Inventory of Malingered Symptomatology; SIMS) and/or PVT (Amsterdam Short-Term Memory test, ASTM) needed to be either passed or failed. Additionally, we selected reports from psychologists who, prior to conducting the assessments, anticipated either problematic or unproblematic validity. A total of 146 reports (57 clinical psychological and 89 neuropsychological reports), authored by 36 psychologists from five different hospitals, were analyzed. Invalid range scores on SIMS and/or ASTM occurred in 48% of the sample. Two researchers independently reviewed and coded reports, resolving mismatches through consensus and crosschecking with original test data. The majority of clinical psychological reports (89.5%) did not reference the SIMS or accurately describe the SIMS results, despite its use. In contrast, most neuropsychological reports mentioned the SIMS and ASTM, and adequately described their results (77.5%). Approximately half of the reports with invalid range scores on these instruments included interpretative statements, often suggesting overreporting and/or underperformance. In about one-third of cases, a fail on the validity test was attributed to factors such as anxiety, fatigue, depression, or pain. Other cognitive tests and psychological questionnaires were frequently interpreted without considering these invalid scores. Treatment recommendations seldom took SVT/PVT fails into account. The findings indicate that a non-negligible number of reports do not accurately report and discuss SVT/PVT results, underscoring the need for enhancing the quality and precision of psychological reports concerning validity testing.

Keywords Symptom Validity · Performance Validity · Psychological Reports · Feedback · Structured Inventory of Malingered Symptomatology · Amsterdam Short-Term Memory Test

✉ Brechje Dandachi-FitzGerald
b.fitzgerald@maastrichtuniversity.nl

1 Department of Clinical Psychological Science, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

2 Faculty of Psychology, Open University, Heerlen, The Netherlands

3 Department of Neurology, Vivantes Klinikum Im Friedrichshain, Berlin, Germany

4 Department of Psychiatry and Psychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, The Netherlands

5 Department of Medical Psychology, VieCuri Medical Center, Venlo, The Netherlands

6 Department of Medical Psychology, Amsterdam University Medical Center, Amsterdam, The Netherlands

Introduction

In certain settings, such as those related to forensic evaluations, individuals may exhibit a tendency to exaggerate or even invent symptoms and/or impairments (Bass & Wade, 2019). Against this backdrop, there has been a continuous interest in the development and utilization of tools to assess the validity of the complaints with which patients present. These tools are commonly known as symptom validity tests (SVTs) and performance validity tests (PVTs), where SVTs assess the tendency to overreport symptoms, while PVTs measure the tendency to engage in cognitive underperformance.

The professional interest in SVTs and PVTs has been driven by three factors. First, empirical research has shown

that many SVTs and PVTs may serve as sensitive indicators of symptom overreporting and cognitive underperformance, respectively**.** Moreover, these tests often demonstrate sufficient specificity. In other words, only a small fraction of individuals with actual cognitive impairment and/or genuine psychopathological conditions exhibit abnormal results, that is, they *fail*[1] on SVTs and/or PVTs (e.g., Dandachi-FitzGerald et al., 2020a; Schroeder & Martin, 2022a). Second, the limitations of relying solely on unstructured clinical interviews and observations to evaluate the validity of self-reported symptoms and impairments have become increasingly apparent. Clinical impressions might be prone to bias and error and therefore often do not meet the standards for evidence-based decision-making (Dandachi-FitzGerald & Martin, 2022; Dandachi-FitzGerald et al., 2017; Ng et al., 2021). Third, although SVTs and PVTs were originally primarily developed for the forensic context, it has become clear that symptom overreporting and cognitive underperformance can transpire in various settings (Dandachi-FitzGerald, et al., 2020b; Merckelbach et al., 2019). For example, in a recent meta-analysis comprising 47 studies (Roor et al., 2024), the base rate of failure on a PVT in clinical neuropsychological assessments was estimated to be approximately 16% (95% CI [13, 19]).

In 2005, the National Academy of Neuropsychology positioned that "when a psychological evaluation is deemed medically necessary, validity assessment is a medical necessity" (Bush et al., 2005, p. 419). Since then, the informational value of validity assessment tools has been stressed in guidelines (e.g., Moore et al., 2021), policy and consensus statements of professional psychology organizations (e.g., Chafetz et al., 2015; Sweet et al., 2021), but their importance has also been recognized outside psychology, the report of the Institute of Medicine (2015) being a case in point. Consistent with this trend, survey studies suggest that clinicians are increasingly integrating validity tests into their regular clinical assessments, including evaluations conducted outside forensic settings. In an international sample of neuropsychologists ($N = 654$) surveyed by Hirst et al. (2017), approximately 71% of the participants were in favor of including a validity test in every neuropsychological evaluation. In a Northern American sample ($N = 316$), Martin et al. (2015) found that almost 92% of the participants stated to often or always include a PVT in their clinical assessment. This compares favorably to earlier surveys in the United States (Sharland & Gfeller, 2007), the United Kingdom (McCarter et al., 2009), and six Western European countries (Dandachi-FitzGerald et al., 2013). In those studies, percentages of participants who reported to often or always include an SVT and/or PVT in their clinical neuropsychological assessment ranged from 16 to 56%.

The issue of *how* the results of SVTs and PVTs are presented in psychological reports is important. Arguably, the incremental value of validity assessment tools critically depends on accurate interpretation and clear communication of results to the patient (e.g., Carone et al., 2010; Martin & Schroeder, 2022), their support system (if applicable), and the referring party. Meanwhile, the position papers and consensus statements cited above are largely silent about how to address the outcome of SVTs and PVTs in psychological reports. Insights into this matter can be gleaned from survey studies where neuropsychologists were asked to choose statements they would employ to communicate their findings regarding SVT and/or PVT fails (e.g., Dandachi-FitzGerald et al., 2013; Hirst et al., 2017; Martin et al., 2015). Surveyed clinicians mostly reported a preference for rather neutral statements to the effect that test data are invalid, that no firm conclusions can be drawn, or that the test results are inconsistent with the severity of injury. Respondents rarely opted for malingering as a descriptive term, although in the survey of Martin et al. (2015), a sizeable minority (i.e., 11% of the respondents) said that they used that term in their reports "often" or "always" when they believed the test results to be indicative of symptom overreporting.

Taking a different approach in a more recent survey, Martin and Schroeder (2021) provided neuropsychologists ($N = 209$) with three different case vignettes of patients who evidently failed a PVT during neuropsychological assessment (i.e., exhibited evidence of cognitive underperformance). The neuropsychologists were invited to freely describe how they would communicate this test outcome in their psychological report. Across the three case vignettes, almost all participants (95%—100%) expressed a preference for explicitly mentioning the PVT fail. Yet, respondents disagreed regarding how to best describe the PVT fail, with the most common statements being that "results are invalid, inaccurate or unreliable"; "results indicate poor or variable effort"; and "results show a poor or variable engagement", but none of those interpretations were endorsed by more than half of the respondents. Curiously enough, there was also intraindividual variability in the preference for certain descriptions across the three case-vignettes. For instance, in the case vignette of a patient with mild head injury suspected of malingering, 24% of the professionals endorsed statements indicating that test results were inconsistent with the patient's injury, functioning, or presentation. However, only 6% did so in the case vignette of a patient failing multiple PVTs without apparent external incentives or significant psychological/medical/psychosocial factors that could explain the invalid test performance. Overall, these survey findings indicate that a significant number of clinicians

[1] Here and elsewhere, we use the descriptor "fail" to refer to an invalid range score (Guilmette et al., 2020).

struggle to articulate the appropriate way to characterize failures on validity tests in psychological reports.

An important limitation of the survey results reported above is that they were based on self-reports. To gain a more comprehensive understanding of how validity tests are incorporated in psychological reports, we need to move beyond studies that rely on these *self-reported* practices, as there might be discrepancies between reported and *actual* practices (e.g., Baumeister, 2007). An illustrative example is provided by the study of MacAllister et al. (2019) on how the results of PVTs during pediatric neuropsychological assessments are documented. These authors referred to a survey study among North American pediatric neuropsychologists ($N = 282$), 92% of whom stated that they used at least one validity test in each assessment (Brooks et al., 2016). As a follow-up, MacAllister et al. (2019) inspected the pediatric neuropsychological reports written by professionals with a similar background as those surveyed in the Brooks et al. (2016) study. Of the 131 reports inspected, only six (< 5%) mentioned validity testing.

From these results, one may suspect that surveys possibly provide an overly optimistic view on the use of validity tests in clinical practice due to selection bias (i.e., those psychologists who use validity tests might be more inclined to respond to surveys addressing this topic) and/or social desirability bias (i.e., survey respondents might overstate how often they use validity tests). Another possibility is that professionals do use validity tests on a fairly wide scale but ignore their outcomes when they yield unfavorable outcomes as it complicates diagnostic interpretation and reporting. Thus, failures on SVTs and/or PVTs might for example be explained away by ad hoc explanations (Merten & Merckelbach, 2013), such as "cry for help" (Dandachi-FitzGerald et al., 2024; see Young, 2022, for a differing position). The importance of this issue lies in the ample evidence indicating that failures on SVTs and/or PVTs align with skewed scores on standard clinical instruments like depression scales and memory tests (Dandachi-FitzGerald et al., 2011; Merten et al., 2020). Neglecting to address this correlation could potentially compromise the precision of diagnostic assessments and subsequent treatment recommendations.

With these considerations in mind, the current study was undertaken to determine how SVT and PVT results are communicated in psychological reports. Specifically, our focus centered on three key qualities: (1) explicit reference to SVTs and/or PVTs in reports; (2) accuracy in scoring and interpreting the results of these tests; and (3) the extent to which conclusions pertaining to psychopathology, personality characteristics, cognitive functioning, and treatment recommendations appropriately consider the outcomes of the validity assessment. A unique aspect of our study was that we had access to the raw scores of the validity tests administered. Thus, we could systematically compare original test data and whether/how they were presented in the clinical reports. In doing so, the current study overcomes the limitations of relying on clinicians' self-reported practices or on reports without the original test scores.

Method

Participants

This archival study constituted a segment of a broader research initiative focusing on validity assessment among clinically referred hospital outpatients (e.g., Dandachi-FitzGerald et al., 2016, 2017). Ethical approval for this study was obtained by the standing ethical committee of the Medical Ethical Committee of Maastricht University Medical Centre [METC 12–4-022.6/pl]. Our approach involved soliciting 153 psychological reports from psychologists across five hospitals in the southern part of the Netherlands. These reports pertained to patients referred for clinical psychological and neuropsychological assessment between July 2012 and May 2013. Ultimately, we successfully obtained 146 reports (95.4%), encompassing 57 clinical psychological and 89 neuropsychological assessments conducted by 36 different psychologists. All reports were anonymized to protect patient information and de-identified with regard to authorship to ensure the confidentiality of the psychologists who authored them. Consequently, specific data on the number of reports contributed by each individual psychologist are not available. Although the five hospitals did not contribute equally, multiple psychologists from each hospital were involved in providing the reports.

For clinical psychological assessments, the psychologists used the Structured Inventory of Malingered Symptomatology was used (SIMS; Smith & Burger, 1997; details below). In neuropsychological assessments, both the SIMS and the Amsterdam Short-Term Memory Test (ASTM; Schmand & Lindeboom, 2005; details below) were employed. The SIMS and ASTM were selected due to their prominence in Dutch validity assessment practices. Both the Dutch adaption of the SIMS (Merckelbach & Smith, 2003) and the ASTM, which was developed in the Netherlands, were ranked as the most commonly used tools among Dutch neuropsychologists in a European survey (Dandachi-FitzGerald & Ponds, 2013; Dandachi-FitzGerald et al., 2013). Although psychologists also relied on other types of validity tests (e.g., MMPI-2 validity scales), the frequencies with which they did so were so low that meaningful analyses was not possible.

In the context of the 57 clinical psychological assessments, there was a fail on the SIMS in 21 cases, representing a failure rate of 37%. Among the 89 neuropsychological assessment reports, both the SIMS and the ASTM were failed in 49 cases, corresponding to a failure rate of

55%. Note that these statistics reflect how the reports were selected from the original pool of 469 reports (see Fig. 1 in supplemental file 1). Briefly, patients with clinically obvious cognitive impairments (e.g., due to dementia, brain tumors, or severe traumatic brain injury) were excluded. In their case, the administration of SVTs/PVTs make less sense – given that they may generate false positives—and it would be understandable if experts would disregard SVTs/PVTs altogether (e.g., Lippa, 2018). As per the ASTM manual (Schmand & Lindeboom, 2005, p. 4) "clinically obvious symptoms" are those evident during informal contact or history taking (e.g., repeatedly providing the same information or failing to recall a previous topic of conversation), without the need for formal cognitive testing to reveal them. Furthermore, we sampled from reports describing patients who either passed or failed the SIMS in the case of clinical psychological assessments and either passed or failed both the SIMS and the ASTM in the case of neuropsychological assessments. Additionally, we selected reports from psychologists who, prior to conducting the assessments, expected either problematic or unproblematic symptom validity (as per Dandachi-FitzGerald et al., 2017) (see Table 1 in supplemental file 1). Although such a priori impressions have been found to be inaccurate—since psychologists cannot reliably determine in advance who will pass or fail validity testing (Dandachi-FitzGerald et al., 2017)—we included it in our selection strategy, because, at the time of study design, we believed it might help us explore whether a priori expectations influenced the interpretation of validity test results. However, the results did not allow for such a detailed analysis of the data.

At the point of writing their psychological reports, the psychologists participating in the study were unaware that their documents would be later screened. However, it is important to note that they had provided informed consent for us to potentially analyze their reports. Note further that the assessments encompassed a variety of additional instruments alongside the SIMS and/or the ASTM. In clinical psychological assessments, scales addressing symptoms, coping strategies, and personality characteristics were employed, occasionally supplemented by an intelligence measure. Neuropsychological assessments featured an array of cognitive tests, sometimes accompanied by a symptom checklist and/or a coping style questionnaire.

Measures

Structured Inventory of Malingered Symptomatology (SIMS) The SIMS is a 75-item self-report questionnaire listing less plausible or even bizarre symptoms that are rated on a dichotomous (yes–no) scale. The cutoff score of > 16 as recommended by Rogers et al. (1996) was used. For this cutoff score, an older paper on the Dutch adaptation of the SIMS ($N = 298$) found a sensitivity of 0.93 and a specificity of 0.98 (Merckelbach & Smith, 2003). However, later studies, such as the meta-analysis by van Impelen et al. (2014) and the systematic review by Shura et al. (2022), raised concerns about the specificity of the SIMS and recommended a higher cutoff score (e.g., > 19). To clarify whether the adoption of an overly liberal SIMS cutoff might have biased our results, we examined the percentage of SIMS scores falling between > 16 and ≤ 19 in the total sample of clinical assessments in this research project ($N = 469$; Dandachi-FitzGerald et al., 2016). We found that only 28 assessments (6%) had SIMS scores in this range. Of these 28 cases, 19 (68%) failed a second validity test, either the MMPI validity scales or the ASTM. Thus, there is little basis for arguing that, in our study, failures on the SIMS were frequently close to the liberal cutoff and therefore false positives that should be disregarded.

Amsterdam Short-Term Memory Test (ASTM) The ASTM is based on a multiple-choice word recognition procedure. The test consists of 30 trials in which a list of five semantically related words is first presented and then, after a simple distraction task, a second list of five words is given, including three from the first list and two distractors (Schmand & Lindeboom, 2005). We used the standard cutoff score of < 85. In the original validation studies, this cutoff score was associated with a specificity of 0.98 and a sensitivity of 0.77 (Schmand & Lindeboom, 2005). That study contrasted experimental malingerers ($n = 84$) and patients suffering from neurological disorders, such as moderate to severe traumatic brain injury, advanced Parkinson's disease, stroke, multiple sclerosis, and severe epilepsy ($n = 206$).

Exit questionnaire After completion of the data collection of the Dandachi-FitzGerald et al. (2016) study, clinicians received an exit questionnaire via e-mail. The exit questionnaire focused on the following clinician characteristics: name, age, function, years of work experience, estimated number of diagnostic assessments in the past year, type of psychological assessment (i.e., clinical psychological, neuropsychological or both), and experience with forensic evaluations and, if so, the number of forensic evaluations in the past year. Clinicians were also queried about their use of validity tests before entering the study, and whether their participation in this study had sensitized them to distorted symptom presentations during their interviews with the patient. Note that this background information has been previously reported for the subgroup of clinicians performing neuropsychological assessments (see Dandachi-FitzGerald et al., 2017). The last two items of the questionnaire related to communication of the validity assessment. First, clinicians were asked whether they *always*, *sometimes*, or *never* commented on the validity assessment in their report.

Second, clinicians were given five statements to describe test results in the invalid range and were asked to indicate the frequency of use for each statement on a five-point Likert scale (i.e., 1 = *never*, 2 = *rarely*, 3 = *sometimes*, 4 = *often*, 5 = *always*) (see supplemental file 3).

Rating list for psychological reports Before we inspected the selected reports, we developed a list of criteria to evaluate the quality of reporting on SVTs and PVTs. Based on previous surveys (e.g., Dandachi-FitzGerald et al., 2013) and guidelines for psychological reports (e.g., Dutch Association of Psychologists, 2017; Martin et al., 2022), the criteria were as follows:

1) Are SVTs and PVTs explicitly mentioned in the method section of the psychological report?
2) Does the behavioral observations section of the report comment upon the patient's motivation and their test effort or task engagement?
3) Are SVT and PVT outcomes described in the results section of the report? Are the validity test results correctly interpreted as a pass or a fail?
4) Are conclusions drawn from pass or fail? In case of a fail: are other test results interpreted? Does the report give an explanation for failing validity testing?
5) Are treatment recommendations given? In case of a fail: Is this outcome taken into account in the treatment recommendation?

The rating list with coding scheme, and interrater agreement can be found in the appendix.

Procedure and Data-Analysis

The psychological reports were independently reviewed and coded by the first two authors (BD and MP) using the rating list. Subsequently, the two databases were combined to evaluate interrater agreement on the main variables of interest. The Kappa coefficients were satisfactory to excellent for all variables (see appendix). To strengthen accuracy, disagreements were identified and resolved through consensus discussion between BD and MP. This involved a thorough review of the report as well as crosschecking of actual SIMS and ASTM test scores. Frequency analysis was conducted on the rating list criteria, and Chi-square analysis was used to compare the reports of clinical psychological and neuropsychological assessments on relevant aspects. Analyses were carried out using SPSS version 25.

Results

Characteristics of Participating Clinicians

In total, 34 (94%) of the clinicians returned the exit questionnaire. Mean age of the clinicians was 37 years (range: 27–62 years). On average, they had 11 years of working experience (range: 1–35). On average, they estimated to have conducted 65 assessments in the past year (range: 0–230). Further background details can be found in supplemental file 2.

Regarding their communication of SVT/PVT fails, thirteen clinicians (38%) said that they *always* commented on the validity of the psychological assessment, twenty (59%) stated that they *sometimes* did so in their reports; only one clinician (2.9%) indicated to never comment on the validity of the assessment. There was little consensus on how to describe fails on validity tests. In general, most clinicians (73%) indicated that they *often* or *always* framed fails in terms of underperformance or overreporting. Similarly, most clinicians (84%) said they *never* or *rarely* employed terms such as malingering or feigning. Other statements to communicate fails on validity tests were endorsed to be used *often* or *always* by a minority of the respondents, namely "test results are invalid" (45.2%), "the test results are inconsistent with the severity of the injury or condition" (32.3%), and "no conclusions can be drawn from test results" (28.2%). More comprehensive details can be found in supplemental file 3.

Psychological Reports: Method section and Behavioral Observations

Out of the 146 reports, 118 (81%) listed the test battery in their method section. The test battery was significantly more often specified in neuropsychological reports (84 out of 89) than in clinical psychological reports (34 out of 57; $\chi^2(1) = 27.04$, $p < 0.01$, $\eta = 0.430$).

Table 1 shows that when validity tests were mentioned, test acronyms were most frequently used. However, in clinical psychological reports, references to validity tests were frequently omitted in the method section, even in those reports that had specified the test battery. In fact, not mentioning a validity test significantly more often characterized clinical psychological (89.5%) than neuropsychological (21.3%) reports; $\chi^2(1) = 64.61$, $p < 0.01$, $\eta = 0.665$.

Clinical psychological and neuropsychological reports differed markedly with regard to behavioral observations relevant for patients' motivation and commitment. As can be seen in Table 2, all neuropsychological reports contained observations, whereas most clinical psychological

Table 1 Mention of validity tests in the Method section of psychological reports (% of samples)

Categories	Full sample ($N=146$)		Clinical psychological reports ($n=57$)		Neuropsychological reports ($n=89$)	
	n	%	*n*	%	*n*	%
Acronyms (SIMS, ASTM)	65	44.5	6	10.5	59	66.3
Validity test not reported	60	41.5	51	89.5	9	10.1
Incomplete (one validity test mentioned, other left out)	10	6.8	0	0	10	11.2
One validity test in full, one acronym	8	5.5	0	0	8	9.0
Validity tests as a generic term	3	2.1	0	0	3	3.4
Full names	0	0	0	0	0	0

SIMS Structured Inventory of Malingered Symptomatology, *ASTM* Amsterdam Short-Term Memory test

Table 2 General impression of motivation and effort mentioned in psychological reports

Categories	Full sample ($N=146$)		Clinical psychological reports ($n=57$)		Neuropsychological reports ($n=89$)	
	n	%	*n*	%	*n*	%
Cooperative attitude and good effort	72	49.3	8	14.0	64	71.9
No observations	44	30.1	44	77.2	0	0
Observations without comment on cooperative attitude and effort	13	8.9	4	7.0	9	10.1
Cooperation and effort fluctuate during the evaluation	13	8.9	1	1.8	12	13.5
Factors that impact the cooperation and effort (e.g., nervousness, tension)	3	2.1	0	0	3	3.4
Possibly feigned symptom presentation	1	0.7	0	0	1	1.1

reports did not. In those reports that contained behavioral observations, clinicians mostly reported that the patient was cooperative and made a sufficient effort during the assessment. In about one out of every ten reports, clinicians noted observable fluctuations in motivation and test effort, with some reports specifically mentioning factors that impacted them. In one report, the clinician raised the possibility of a feigned symptom presentation.

Psychological Reports: Result Section

A total of 123 out of 146 reports (84%) described the outcomes of individual tests and questionnaires. In the remaining reports, test results were only summarized. The proportion of clinical psychological reports without description of individual test results (21 out of 57) was significantly higher than that of neuropsychological reports (2 out of 89; $\chi^2(1)=31.33, p<0.01, \eta=0.463$). As shown in Table 3, the description of SVT/PVT results varied considerably. Only about half of the psychological reports (51%) correctly described the scores of the validity test(s), either in full or summarized form. In the other reports (49%), validity test(s) were either not mentioned at all, or reported incompletely, or incorrectly. In neuropsychological reports, SVT/PVT performance was more often correctly described (77.5%) as compared to clinical psychological reports (10.5%; $\chi^2(1)=62.44, p<0.001, \eta=0.654$).

The discussion of a validity test failure with the patient was detailed in just two (2.9%) of the 70 reports where such a failure occurred. In particular, one clinical psychological report outlined the discussion of poor symptom validity, with the patient admitting to responding rather impulsively and uncritically. Additionally, a separate neuropsychological report indicated that the assessment was concluded prematurely due to underperformance on the ASTM.

Psychological Reports: Conclusion Section

Interpretation and explanation of SVT/PVT fails Of the 70 reports that pertained to patients who had failed on validity tests, 33 (47%) included one or more interpretative statements of this result. Clinical psychological and neuropsychological reports differed in this respect: only two out of 21 clinical psychological reports (11%) contained interpretative statements of a fail against 31 of 49 neuropsychological reports (63%); ($\chi 2\ (1)=17.04, p<0.01, \eta=0.493$).

Table 3 Description of SVT/PVT results in the Results Section of psychological reports

Categories	Full sample ($N=146$)		Clinical psychological reports ($n=57$)		Neuropsychological reports ($n=89$)	
	n	%	*N*	%	*n*	%
Correct	75	51.4	6	10.5	69	77.5
- Only pass or fail	34	23.3	6	10.5	28	31.5
- Full description (including obtained test scores)	24	16.4	0	0	24	27.0
- Summarized as validity assessment passed or failed	17	11.6	0	0	17	19.1
Incorrect	71	48.6	51	89.5	20	22.5
- Not mentioned	50	34.2	49	86.0	1	1.1
- Incomplete (i.e., not all administered validity tests are mentioned)[a]	15	10.3	2	3.5	13	14.6
- Incorrectly reported[b]	6	4.1	0	0	6	6.7

[a]In eleven reports the SIMS was not mentioned. In three reports, the SIMS and ASTM were not mentioned, but the results of another validity test were reported; namely the Test of Memory Malingering ($n=1$) or the MMPI-2 ($n=2$). [b]In six reports, only raised scores on the subscales of the SIMS were mentioned without giving the total SIMS score; in three of those reports, the total score was non-deviant; in the other three reports, the total score was deviant (> 16)

The most common interpretation was that the test results suggested or pointed towards underperformance or over-reporting, followed by the notion that the tests results were invalid (see Table 4). The suggestion of malingering or feigning was never made. In seven of the 70 reports in which there was a fail on validity tests (10%), it was stated that the assessment was nevertheless judged to be valid. In one report, this interpretation was substantiated (see below), but in the other six reports it appeared that the clinicians' own judgment just overruled the invalid range scores on the SVT/PVT. That is, the clinician considered the assessment valid due to the patient's seemingly cooperative behavior and good effort observed during testing. To illustrate this, one report stated: "A suboptimal performance was observed on a performance validity test. However, behavioral observations showed a cooperative work attitude, good effort, and performance focus. Therefore, the overall assessment was considered valid."

Table 4 Interpretation and explanation of SVT/PVT fail in psychological reports

Categories	Reports with SVT/PVT fail ($N=70$)		Clinical psychological reports with SIMS fail ($n=21$)		Neuropsychological reports with with SIMS & ASTM fail ($n=49$)	
Interpretations	*n*	%	*n*	%	*n*	%
The test suggests or points towards overreporting or underperformance	26[a]	38.5	1	4.8	26[a]	53.0
The test results are invalid	11	15.7	1	4.8	10	20.4
No conclusions can be drawn	4	5.7	0	0	4	8.2
The test results are inconsistent with the severity of the injury or condition	2	2.9	0	0	2	4.1
The test results suggest or indicate malingering / feigning	0	0	0	0	0	0
The assessment is valid	7	10.0	1	4.8	6	12.2
Explanations						
No explanation	39	55.7	19	90.4	20	40.8
Unsubstantiated explanation (e.g., due to anxiety, mood problems, fatigue)	22	31.4	1	4.8	21	42.9
Substantiated explanation	2	2.9	0	0	2	4.1
Other	7	10.0	1	4.8	6	12.2

SVT Symptom Validity Test, *PVT* Performance Validity Test, *SIMS* Structured Inventory of Malingered Symptomatology, *ASTM* Amsterdam Short-Term Memory Test. [a]One report mentioned that test results suggested underperformance but did not mention overreporting

In many reports (55.7%), no explanation was given for SVT/PVT fail (see Table 4). However, in about one third of the reports, a fail on validity tests was apparently explained by attributing it to factors such as (performance) anxiety, fatigue, depression, or pain. To illustrate this, a few sample statements are presented here (translated into English):

1) "There were indications of underperformance and symptom overreporting, and no reliable conclusions could be drawn. Apparently, the patient was hindered to perform optimally by symptoms such as concentration difficulties, fatigue, and mood problems."
2) "The patient exhibited a strong illness perception and was likely to suffer from a depressive or anxiety disorder. Reduced selective attention and, resulting from it, reduced memory performance may have coincided with this."
3) "There were indications of insufficient performance validity. The patient appeared particularly nervous when explicitly instructed to memorize items."

Only two reports offered a more articulated explanation. One of them was a report about a 59-year-old patient with a history of encephalitis resulting in moderate to severe deficits with physical and cognitive impairment. The patient was assessed in the context of a follow-up assessment three years after rehabilitation. There was a consistent test profile, with prominent deficits in working memory. The report notes that, despite the patient's fail on the ASTM (with a score of 80 correct out of 90), the performance validity was deemed acceptable, considering the influence of impaired working memory on this PVT. The second report related to the assessment of a 51-year-old patient with persistent cognitive complaints after a motor vehicle accident. The patient scored low on the ASTM (72 correct) and high on the SIMS (27 endorsed items). The report explicitly mentioned that current legal proceedings, along with pain, fatigue, and personality characteristics, could explain both the cognitive underperformance and distorted symptom presentation.

Interpretation of other Cognitive Tests and Psychological Questionnaires

In most cases, cognitive tests and psychological questionnaires were interpreted without giving due consideration to failures on the validity test(s) (see Table 5). None of the clinical psychological reports explicitly took the SIMS failure into account when interpreting personality questionnaires or symptom checklists. Moreover, almost half of the neuropsychological reports did not explicitly consider the failure on two validity tests when interpreting standard tests. In a notable minority of neuropsychological reports, the failures on the SVT/PVT were rationalized or apparently explained away. For example, one report stated: "The fluctuating memory performance can be attributed to the individual's fixation on memory complaints, as well as their high level of uncertainty and distress". In another report, the psychologist wrote: "Suboptimal performance on some tests (reduced mental speed) can be understood in the context of the mood disorder problems."

Relatedly, in seven out of the 49 neuropsychological reports in which the SIMS and ASTM were failed (14.3%), there was a disparity such that cognitive test results were deemed invalid due to PVT fail, but psychological questionnaires were still treated as though they were valid despite SVT fail. As an example, the neuropsychological report of a case with both the ASTM (80 correct) and SIMS (25 points) scores in the invalid range, stated: "The obtained test results

Table 5 Interpretation of other cognitive tests and psychological questionnaires in case of SVT/PVT in the invalid range

Categories	Reports with SVT/PVT fail ($n=70$)		Clinical psychological reports with SIMS fail ($n=21$)		Neuropsychological reports with SIMS & ASTM fail ($n=49$)	
	n	%	*n*	%	*n*	%
Yes, without consideration of SVT/PVT fail	44	62.9	21	100	23	46.9
Yes, while explaining away SVT/PVT fail (e.g., due to pain, fatigue, depression) or with only a general warning to "interpret with caution"	8	12.9	0	0	8	16.3
Fail PVT considered, fail SVT not	7	10.0	0	0	7	14.3
No interpretation	5	7.1	0	0	5	10.2
Yes, with taking SVT/PVT fail into account	2	2.9	0	0	2	4.1
Other	4	5.7	0	0	4	8.2

S*VT* Symptom Validity Test, *PVT* Performance Validity Test, *SIMS* Structured Inventory of Malingered Symptomatology, *ASTM* Amsterdam Short-Term Memory test

do not validly reflect actual cognitive functioning", but also that: "The score on a depression questionnaire is raised and indicative of a severe depressive disorder." In only a minority of the neuropsychological reports (7 out of 49; 14.3%), the tests were either interpreted with consideration of the fail on the validity tests, or they were not interpreted at all because of a fail on the SVT/PVT. As an illustration, the report about a 67-year-old patient, experiencing difficulties in adjusting to retirement from work, and presenting with somatic, cognitive, and depressive symptoms, concluded that the patient showed signs of underperformance on cognitive tests. While his scores on the standard cognitive tests were generally within the normal range, there was an exception in the case of memory tasks. The observed discrepancy in memory performance was attributed to suboptimal effort in completing these specific tests, and no objectively verified cognitive deficits in memory were identified.

Psychological Reports: Treatment Recommendations

As shown in Table 6, most psychological reports included some sort of treatment recommendation, such as psychoeducation regarding mood and cognitive functioning, pragmatic supportive treatment, psychotherapy, or referral to a psychiatrist to check the indication for an antidepressant medication, without addressing the failing on validity tests. Only in a small minority of psychological reports (3 out of 70 reports; i.e., 3.4%), SVT/PVT fail was considered in the treatment recommendation to varying degrees. One report mentioned neuropsychological counseling in order to discuss the findings of the neuropsychological assessment (i.e., indications of cognitive underperformance and symptom overreporting). Another report stated that the patient did admit to have produced invalid test data due to underperformance but was willing to explore possible underlying mechanisms. The psychologist, therefore, recommended a referral to a psychologist or psychotherapist outside the hospital setting. The third report mentioned that the patient and his spouse were relieved by the results of the neuropsychological assessment (i.e., no objectified cognitive deficits). The patient, experiencing cognitive difficulties since a 2005 myocardial infarction, acknowledged the possibility of hyperfocus on cognitive errors, leading to an overemphasis on those specific complaints. A follow-up appointment was scheduled, and the patient was willing to practice focusing less on mistakes and discontinuing excessive self-monitoring.

Discussion

This study examined 146 reports of hospital outpatients undergoing routine psychological assessment, focusing on the outcomes of validity tests. The results of this study offer a number of insights that are valuable for understanding how psychologists communicate the results of SVT and PVT failures. Our results also provide useful guidance on how to promote improvements in this area.

To begin with, reports often did not consistently describe the validity test(s) administered and its/their result. Interestingly, we noted a difference in this regard between clinical psychological and neuropsychological reports. Specifically, close to 90% of psychological reports (against 21% of the neuropsychological reports) did not mention the SIMS, even though this SVT was utilized. Relatedly, psychological reports ignored patients' fail on the SIMS. We can only speculate about the reason for this discrepancy. One contributing factor may be that clinical psychological reports tended to be more unstructured than neuropsychological reports, lacking a clearly defined method section, listing the test battery, and results section. However, this factor may not fully account for the discrepancy, as clinical psychological reports still

Table 6 Treatment recommendations provided in case of SVT/PVT in the invalid range

Categories	Reports with SVT/PVT fail ($n=70$)		Clinical psychological reports with SIMS fail ($n=21$)		Neuropsychological reports with SIMS & ASTM fail ($n=49$)	
	n	%	*n*	%	*n*	%
Yes, without explicit consideration of SVT/PVT fail	55	78.6	21	100	35	71.4
No	8	11.4	0	0	8	16.3
Yes, with explicit consideration of SVT/PVT fail	3	4.3	0	0	3	6.1
Other	3	4.3	0	0	3	6.1

SVT Symptom Validity Test, *PVT* Performance Validity Test, *SIMS* Structured Inventory of Malingered Symptomatology, *ASTM* Amsterdam Short-Term Memory test

incorporated a (summarized) interpretation of findings based on standard clinical scales that had been administered. Possibly, the discrepancy reflects different research traditions, with a more pronounced interest in validity tests in the field of neuropsychology as compared to clinical psychology (e.g., Sweet et al., 2021). Still, about one in five neuropsychological reports omitted references to the SIMS and/or ASTM despite the fact that they had been administered.

One obvious improvement for both types of reports might be to adopt a structured format in which the tests administered are consistently listed along with their results. This approach aligns with guidelines for the use of tests of the Dutch Association of Psychologists (2017) emphasizing that "it is necessary to specify the psychological instruments used in the report. These can be included in the body of the report or in an attachment" (p. 35). This resonates with the recommendation in the American Academy of Clinical Neuropsychology (AACN) Consensus Conference Statement on neuropsychological validity: "In their reports, neuropsychologists list the PVTs and validity assessment procedures that are utilized in evaluations" (p. 1067, Sweet et al., 2021). However, there are limits to how much information psychologists should provide about symptom and performance validity tests, given that too much transparency might undermine test integrity. To safeguard test integrity, Schroeder and Martin (2022b) recommended to refer to validity tests by their acronyms. In line with this advice, the current study found that acronyms were most often used whenever validity tests were explicitly mentioned in the reports.

A second conclusion that can be drawn from our findings is that many clinicians seem to struggle with interpreting fails on symptom and performance validity tests. At the very least, there is a lack of consensus on how approach such interpretation. A common strategy to navigate uncertainties surrounding interpretations of validity test outcomes appears to be avoiding mentioning them altogether. A notable proportion of reports with failed validity testing lacked interpretative statements on this result (89% in psychological reports and 37% in neuropsychological reports). A minority of reports (39%) stated that the test results indicated symptom overreporting and cognitive underperformance. An even smaller proportion of reports (15%) proffered the interpretation that the test results were invalid. Both interpretations accurately describe what it means when a patient fails a validity test, with the latter offering a more comprehensive view by addressing the data integrity of other test results (but see also Guilmette et al., 2020).

The reluctance to provide interpretative statements in cases of symptom validity fails stands in sharp contrasts to the willingness of psychologists to communicate their observations on the patient's supposed cooperation and effort during testing. Indeed, the majority of neuropsychological reports contained subjective impressions as to the cooperative attitude and good effort of the patient. Meanwhile, such impressions have been shown to be notoriously unreliable (Dandachi-FitzGerald & Martin, 2022; Dandachi-FitzGerald et al., 2017; Faust et al., 1988). To enhance clarity, it would be advisable to refrain from overall clinical impressions and only describe distinct deviations when they do occur during testing (e.g., fluctuations in effort, boredom, or signs of uncooperativeness).

Third, around one-third of the reports about patients who did not pass validity testing attributed this fail to factors such as anxiety, fatigue, depression, or pain (see, for example, the three statements in the Results section). Empirical findings suggest that these factors do not genuinely account for validity test fails, except perhaps in the most severe cases. Consequently, they can be regarded as unsubstantiated explanations (Green & Merten, 2013; Dandachi-FitzGerald et al., 2024).

Fourth and related, scores on standard cognitive tests and psychological questionnaires were often interpreted without taking into account that the patient failed on the SIMS and/or ASTM. Likewise, treatment recommendations were regularly given without due consideration of the fact that the patient had failed a validity test. To some extent, the challenge at hand is the issue of data integration, as highlighted by Faust (1989) and Wedding and Faust (1989). Clinicians, like humans in general, face difficulty in amalgamating information from multiple sources into a cohesive and comprehensive interpretation (Faust & Furman, 2022). Nevertheless, what distinguishes validity assessments is their potential to raise red flags. It is clear that failures in validity tests should be accorded informational priority. The credibility of clinical information obtained from routine clinical instruments is crucial for diagnostic or therapeutic interpretations, and encountering a fail introduces a significant challenge in ensuring the reliability of such clinical data (e.g., Dandachi-FitzGerald et al., 2011; Green et al., 2001).

In summary, our findings underscore clinicians' challenges in assessing symptom and performance validity during routine psychological assessments. Clinicians frequently refrain from reporting on validity testing or tend to explain away fails. The implications of validity test fails are often not fully considered when other test results are interpreted, conclusions drawn, and treatment recommendations given. While some improvements, such as implementing a structured report format for listing the test battery and consistent result reporting (e.g., Guilmette et al., 2020), may be relatively straightforward, our findings underscore the importance of implementing systematic and comprehensive education to enhance understanding of validity testing.

We readily acknowledge some important limitations to our study. For one thing, the clinical psychological and neuropsychological reports that we analyzed were crafted approximately ten years ago. During the intervening period,

there has been a conceptual evolution in our understanding of symptom and performance validity tests, viewing them more as measures of behavior, specifically symptom over-reporting and cognitive underperformance, rather than solely as indicators of malingering (Merten et al., 2022). Hence, it is conceivable that our results could be outdated due to the evolving understanding and conceptual refinement in the field over the past decade. Nevertheless, the clinicians involved in this study were likely to be more aware of symptom and performance validity testing than the average Dutch clinician. For instance, both the first and fifth authors, affiliated with one of the largest contributing hospitals, provided consultations on the topic to colleagues upon request. Additionally, the willingness of colleagues in various hospitals to participate in a study on symptom and performance validity in routine clinical assessments indicates a genuine interest in the topic and its importance.

A restriction that is related to the previous point is that, due to the anonymization of our data set, we could not determine whether disregarding SVT/PVT results was typical across all psychologists from the five different hospitals involved in this research project. The unit of analysis was the reports themselves, rather than the individuals who authored them. However, our findings suggest the need for future research to systematically compare what psychologists report about their use of SVTs/PVTs with how they actually document these tools in their reports (see also MacAllister et al., 2019).

A second limitation is that our study took place in The Netherlands, and its findings may therefore not be directly generalizable to other countries. As far as the European continent is concerned, The Netherlands is recognized for its advanced practices in symptom and performance validity assessment (Merten et al., 2022). Consequently, our findings might compare favorably to those in other countries. However, the extent to which our findings align with or diverge from practices in the United States remains unknown.

A third limitation is that our focus was solely on validity test results. This test-centered approach may overlook other factors, such as age, education, medical diagnoses, and indicators of invalid performance (e.g., compelling inconsistencies between interviews, behavioral observations, and cognitive test scores), necessary to determine whether a score in the invalid range on a PVT/SVT represents a true or false positive finding, and to understand the overall validity of the test profile. Furthermore, the reliance on a single SVT and/or PVT may not provide a sufficient basis for accurate determinations on the validity of self-reported symptoms and cognitive test performance. Although we excluded patients with obvious cognitive impairment to minimize the risk of false positives, we cannot entirely rule out this possibility. Future studies might adopt a more comprehensive approach to interpreting validity test results, such as utilizing an expert-panel approach to evaluate the complete file. This aligns with the consensus that "in situations in which an examinee produces one or more invalid range or indeterminate range score(s), it is ultimately the clinician who is responsible for judging, based on the totality of information available, what those scores mean and how they should be interpreted" (Guilmette et al., 2020, p. 449). As an additional point, while our primary focus was on how clinicians communicate validity test results, we note that in cases of multiple failures on PVTs/SVTs indicating invalid test data, further criteria, such as the presence of incentives, should be considered for a determination of possible causes (if required), such as malingering (Sherman et al., 2020) or factitious disorder (Chafetz et al., 2020).

Fourth, and related to the previously discussed limitations, clinical psychological and neuropsychological reports differed in their approaches: the former used only one validity assessment tool, while the latter relied on two. It could be argued that invalid scores on two validity tests are more difficult to overlook than a failure on just one, which may explain the observed differences between these two types of reports. Therefore, the disparities between the two types of reports might partly be attributed to this methodological difference.

Yet another limitation is that we focused on whether the results of validity tests were given adequate consideration in clinical psychological and neuropsychological reports. We did not check to what extent other types of clinical information, such as intelligence and personality test scores, were accurately portrayed in these reports (see for critical discussions, Allard & Faust, 2000; Styck & Walsh, 2016). Future studies might want to compare how different types of psychometric information including those from validity assessment tools, are represented in reports.

While acknowledging these limitations, the distinctive strength of our study lies in the accessibility of the raw scores obtained from the tests administered during routine assessments. Additionally, at the time of writing the reports, clinicians were unaware that their reports would be subject to further inspection. Thus, our study offers insight into real-world practices and may serve as a valuable reference point for future studies, delving into the actual practices of communicating symptom and performance validity assessment results.

The interpretation and communication of validity test failures represent an ongoing concern that merits attention in future research. Declaring test data as invalid is an initial step, yet it is crucial to recognize that providing invalid data in clinical and rehabilitation settings should be seen as behavior requiring clinical attention (e.g., Carone & Bush, 2018). In this vein, Boone (2007) remarked that "recognition of feigning behaviors may prove to be the first therapeutic step in understanding the patient's actual needs" (p. 11).

This holds true only when validity test failure is properly understood and not hastily dismissed or interpreted using unsubstantiated explanations. In moving forward, researchers should address the existing gap in both research and skills training concerning how to effectively handle validity test failures, shifting the focus from mere detection to comprehensive studies on providing feedback, communicating findings, and understanding the clinical implications of such failures.

Appendix

Table 7 Rating list psychological reports with the initial agreements of the independent ratings

Variables	Scoring categories	kappa (number of disagreements)
Methods		
Is the test battery mentioned in the report?	0 = no 1 = yes	0.978 (1)
Are the validity tests mentioned in the test battery?	0 = yes, complete (ASTM, SIMS) 1 = acronyms 2 = incomplete (ASTM or SIMS) 3 = not mentioned 4 = one by acronym, other in full name	0.848 (14)
Observations		
Is there a statement that the patient is cooperative with the assessment and puts forth good effort?	0 = statement present 1 = not mentioned 2 = cooperation / effort is suboptimal / fluctuates 3 = no observations in the report	0.647 (34)
Results		
Is there a separate result section present?	0 = no 1 = yes	0.581 (27)
Is the outcome of the validity test(s) mentioned?	0 = complete, including the obtained score(s) 1 = the classification of pass or fail on the test 2 = summarized outcome of validity testing 3 = not mentioned 4 = incomplete (one test is mentioned, the other not) 5 = incorrect interpretation of a pass or fail on one or both tests	0.551 (50)
In case of test validity failure(s), is there a statement that:		
(a) the test invalidity has been discussed with the patient?	0 = no 1 = yes	* (1)
(b) the assessment has been adapted?	0 = no 1 = yes –> open text: nature of the adaptation	1.00 (0)
Conclusion		
Is there a statement that:		
(a) the test results are inconsistent with the severity of the injury or condition?	0 = no 1 = yes	* (3)
(b) the test results suggest or indicate underperformance and/or symptom overreporting?	0 = no 1 = yes	0.765
(c) no conclusions can be drawn?	0 = no 1 = yes	0.487 (4)
(d) the test results are invalid?	0 = no 1 = yes	0.764 (10)
(e) the test results suggest or indicate malingering?	0 = no 1 = yes	* (1)

Table 7 (continued)

Variables	Scoring categories	kappa (number of disagreements)
(f) the test results suggest or indicate a valid assessment?	0 = no 1 = yes	0.557 (12)
Is there an interpretation of the scores on the other tests?	0 = no 1 = Yes, without taking validity test failure into account 2 = Yes, with warning "interpret with caution" 3 = Yes, while explaining away validity test failure (e.g., due to pain, fatigue, depression) 4 = Yes, in case of passing the validity test(s) 5 = Yes, with taking validity test failure(s) into account 6 = other	0.591 (26)
Is there an explanation given for the validity test(s) failure(s)?	0 = no explanation 1 = yes, a substantiated explanation 2 = Yes, explaining away validity test failure (e.g., due to pain, fatigue, depression) 3 = no validity test(s) failure(s) 4 = other	0.659 (24)
Are treatment recommendations given?	0 = no treatment recommendation 1 = yes, taking the validity test(s) failure(s) into account 2 = yes, without taking the validity test(s) failure(s) into account 3 = yes in case of passing validity test(s) 4 = other	0.729 (26)

* kappa not calculated because one rater scored this category as a constant

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12207-024-09519-2.

Data Availability The anonymized database of the psychological report ratings used in this study can be obtained upon reasonable request by contacting the first author at b.fitzgerald@maastrichtuniversity.nl. The original psychological reports cannot be made available due to their sensitive nature.

Declarations

Conflict of Interest We have no known conflict of interest to disclose.

References

Allard, G., & Faust, D. (2000). Errors in scoring objective personality tests. *Assessment, 7*(2), 119–129. https://doi.org/10.1177/107319110000700203

Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology, 19*(2), 96–105. https://doi.org/10.1136/practneurol-2018-001950

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Boone, K. B. (Ed.). (2007). Assessment of feigned cognitive impairment: *A neuropsychological perspective.* Guilford Press

Brooks, B. L., Ploetz, D. M., & Kirkwood, M. W. (2016). A survey of neuropsychologists' use of validity tests with children and adolescents. *Child Neuropsychology, 22*(8), 1001–1020. https://doi.org/10.1080/09297049.2015.1075491

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. https://doi.org/10.1016/j.acn.2005.02.002

Carone, D. A., & Bush, S. S. (2018). *Validity assessment in rehabilitation psychology and settings*. Oxford University Press.

Carone, D. A., Iverson, G. L., & Bush, S. S. (2010). A model to approaching and providing feedback to patients regarding invalid test performance in clinical neuropsychological evaluations. *The Clinical Neuropsychologist, 24*(5), 759–778. https://doi.org/10.1080/13854041003712951

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., ... & Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist, 29*(6), 723–740. https://doi.org/10.1080/13854046.2015.1099738

Chafetz, M. D., Bauer, R. M., & Haley, P. S. (2020). The other face of illnessdeception: Diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The*

Clinical Neuropsychologist, *34*(3), 454–476. https://doi.org/10.1080/13854046.2019.1663265

Dandachi-FitzGerald, B., Ponds, R. W., Peters, M. J., & Merckelbach, H. (2011). Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *The Clinical Neuropsychologist, 25*(5), 812–828. https://doi.org/10.1080/13854046.2011.583280

Dandachi-FitzGerald, B., Duits, A. A., Leentjens, A. F., Verhey, F. R., & Ponds, R. W. (2020a). Performance and symptom validity assessment in patients with apathy and cognitive impairment. *Journal of the International Neuropsychological Society, 26*(3), 314–321. https://doi.org/10.1016/j.psychres.2016.02.061

Dandachi-FitzGerald, B., & Martin, P. K. (2022). Clinical judgment and clinically applied statistics: Description, benefits, and potential dangers when relying on either one individually in clinical practice. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 107–125). The Guilford Press.

Dandachi-FitzGerald, B., Merckelbach, H., Bošković, I., & Jelicic, M. (2020b). Do you know people who feign? Proxy respondents about feigned symptoms. *Psychological Injury and Law, 13*(3), 225–234. https://doi.org/10.1007/s12207-020-09387-6

Dandachi-FitzGerald, B., Merckelbach, H., & Merten, T. (2024). Cry for help as a root cause of poor symptom validity: A critical note. *Applied Neuropsychology: Adult, 31*(4), 527–532. https://doi.org/10.1080/23279095.2022.2040025

Dandachi-FitzGerald, B., Merckelbach, H., & Ponds, R. W. (2017). Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology, 39*(3), 257–264. https://doi.org/10.1080/13803395.2016.1223278

Dandachi-Fitzgerald, B., & Ponds, R. W. H. M. (2013). Onderpresteren en overrapporteren bij neuropsychologisch onderzoek: wat is de praktijk?, een enquête onder neuropsychologen in Nederland. [Underperformance and overreporting in neuropsychological assessment: What are the current practices? A survey of neuropsychologists in the Netherlands.] *Tijdschrift voor Neuropsychologie, 8*(1), 2–16.

Dandachi-FitzGerald, B., Ponds, R. W. H. M., & 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*(8), 771–783. https://doi.org/10.1093/arclin/act073

Dandachi-FitzGerald, B., van Twillert, B., van de Sande, P., van Os, Y., & Ponds, R. W. (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. https://doi.org/10.1016/j.psychres.2016.02.061

Dutch Association of Psychologists. (2017). Guidelines for the use of tests. *Dutch Association of Psychologists*. https://issuu.com/communicatienip/docs/4132.31_nip_brochure_engels_2_def_v

Faust, D. (1989). Data integration in legal evaluations: Can clinicians deliver on their premises. *Behavioral Sciences & the Law, 7*(4), 469–484. https://doi.org/10.1002/bsl.2370070405

Faust, D., Hart, K., Guilmette, T. J., & Arkes, H. R. (1988). Neuropsychologists' capacity to detect adolescent malingerers. *Professional Psychology: Research and Practice, 19*(5), 508–515. https://doi.org/10.1037/0735-7028.19.5.508

Faust, D., & Furman, A. (2022). When clinical judgment and science conflict, how does one decide? The epistomological status of learning from experience vs. Science. In C.L. Cobb, S. J. Lynn, & W. O'Donohue, W. (Eds.), *Towards a science of clinical psychology: A tribute to the life and works of Scott O. Lilienfeld* (pp. 71–104). Springer

Green, P., & Merten, T. (2013). Noncredible explanations of noncredible performance on symptom validity tests. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 73–100). Springer.

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen, L.M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain injury, 15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Guilmette, T. J., Sweet, J. J., Hebben, N., Koltai, D., Mahone, E. M., Spiegler, B. J., Stucky, K., Westerveld, M., & Participants, C. (2020). American Academy of Clinical Neuropsychology consensus conference statement on uniform labeling of performance test scores. *The Clinical Neuropsychologist, 34*(3), 437–453. https://doi.org/10.1080/13854046.2020.1722244

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. https://doi.org/10.1093/arclin/acx009

Institute of Medicine. (2015). Psychological testing in the service of disability determination. *National Academies Press*

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*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

MacAllister, W. S., Vasserman, M., & Armstrong, K. (2019). Are we documenting performance validity testing in pediatric neuropsychological assessments? *A Brief Report. Child Neuropsychology, 25*(8), 1035–1042. https://doi.org/10.1080/09297049.2019.1569606

Martin, P. K., Schroeder, R. W., Boone, K. B., & Bush, S. S. (2022). Ethical considerations associated with validity assessment in clinical evaluations. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 86–103). The Guilford Press.

Martin, P. K., & Schroeder, R. W. (2022). A framework for providing clinical feedback when patients invalidate testing. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 47–69). The Guilford Press.

Martin, P. K., & Schroeder, R. W. (2021). Feedback with patients who produce invalid testing: Professional values and reported practices. *The Clinical Neuropsychologist, 35*(6), 1134–1153. https://doi.org/10.1080/13854046.2020.1722243

Martin, P. K., Schroeder, R. W., & Odland, A. P. (2015). Neuropsychologists' validity testing beliefs and practices: A survey of North American professionals. *The Clinical Neuropsychologist, 29*(6), 741–776. https://doi.org/10.1080/13854046.2015.1087597

McCarter, R. J., Walton, N. H., Brooks, D. N., & Powell, G. E. (2009). Effort testing in contemporary UK neuropsychological practice. *The Clinical Neuropsychologist, 23*(6), 1050–1066. https://doi.org/10.1080/13854040802665790

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. https://doi.org/10.1016/S0887-6177(01)00191-3

Merckelbach, H., Dandachi-FitzGerald, B., van Helvoort, D., Jelicic, M., & Otgaar, H. (2019). When patients overreport symptoms: More than just malingering. *Current Directions in Psychological Science, 28*(3), 321–326. https://doi.org/10.1177/0963721419837681

Merten, T., Dandachi-FitzGerald, B., Hall, V., Bodner, T., Giromini, L., Lehrner, J., González-Ordi, H., Santamaría, P., Schmand,

B., & Di Stefano, G. (2022). Symptom and performance validity assessment in European countries: An update. *Psychological Injury and Law, 15*(2), 116–127. https://doi.org/10.1007/s12207-021-09436-8

Merten, T., Kaminski, A., & Pfeiffer, W. (2020). Prevalence of overreporting on symptom validity tests in a large sample of psychosomatic rehabilitation inpatients. *The Clinical Neuropsychologist, 34*(5), 1004–1024. https://doi.org/10.1080/13854046.2019.1694073

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*(2), 122–137. https://doi.org/10.1007/s12207-013-9155-x

Moore, P., Bunnage, M., Kemp, S., Dorris, L., & Baker, G. (2021). Guidance on the assessment of performance validity in neuropsychological assessment. British Psychological Society

Ng, W., Mattos, L. A., Coffey, C. A., Molina, S. M., Gottfried, E. D., & Glassmire, D. M. (2021). The association between clinicians' initial judgments of feigning and outcomes on symptom validity measures among pretrial forensic psychiatric inpatients. *International Journal of Law and Psychiatry, 76*, 101698. https://doi.org/10.1016/j.ijlp.2021.101698

Rogers, R., Hinds, J. D., & Sewell, K. W. (1996). Feigning psychopathology among adolescent offenders: Validation of the SIRS, MMPI-A, and SIMS. *Journal of Personality Assessment, 67*(2), 244–257. https://doi.org/10.1207/s15327752jpa6702_2

Roor, J. J., Peters, M. J. V., Dandachi-FitzGerald, B., & Ponds, R. W. H. M. (2024). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychological Review, 34*(1), 299–319. https://doi.org/10.1007/s11065-023-09582-7

Schroeder, R. W., & Martin, P. K. (2022a). Explanations of performance validity test failure in clinical settings. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 11–30). Guilford Press.

Schroeder, R. W., & Martin, P. K. (2022b). Incorporating of validity assessment and validity findings when writing clinical reports. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 70–85). Guilford Press.

Schmand, B., & Lindeboom, J. (2005). *Amsterdam short-term memory test: Manual*. Psychologische Instrumenten, Tests en Services.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213–223. https://doi.org/10.1016/j.acn.2006.12.004

Sherman, E. M., Slick, D. J., & Iverson, G. L. (2020). Multidimensional malingering criteria for neuropsychological assessment: A 20-year update of the malingered neuropsychological dysfunction criteria. *Archives of Clinical Neuropsychology*, *35*(6), 735–764. https://doi.org/10.1093/arclin/acaa019

Shura, R. D., Ord, A. S., & Worthen, M. D. (2022). Structured Inventory of Malingered Symptomatology: A psychometric review. *Psychological Injury and Law, 15*(1), 64–78. https://doi.org/10.1007/s12207-021-09432-y

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *The Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189. https://doi.org/10.1007/s12207-021-09432-y

Styck, K. M., & Walsh, S. M. (2016). Evaluating the prevalence and impact of examiner errors on the Wechsler scales of intelligence: A meta-analysis. *Psychological Assessment, 28*(1), 3–17. https://doi.org/10.1037/pas0000157

Sweet, J. J., Heilbronnner, R. L., Morgan, J. E., Larrabee, G. J., Rohling, M. L., Boone, K. B., Kirkwood, M. W., Schroeder, R. W., Suhr, J. A., & Participants, C. (2021). American Academy of Clinical Neuropsychology (AACN) 2021 consensus statement on validity assessment: Update of the 2009 AACN consensus conference statement on neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 35*(6), 1053–1106. https://doi.org/10.1080/13854046.2021.1896036

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. https://doi.org/10.1080/13854046.2014.984763

Wedding, D., & Faust, D. (1989). Clinical judgment and decision making in neuropsychology. *Archives of Clinical Neuropsychology, 4*(3), 233–265. https://doi.org/10.1016/0887-6177(89)90016-4

Young, G. (2022). Interpreting symptom validity test fails in forensic disability and related assessments: When the cry for help for one fail makes sense. *Applied Neuropsychology: Adult*. Advance online publication. https://doi.org/10.1080/23279095.2022.2107929