Psychological Injury and Law (2024) 17:439–449
https://doi.org/10.1007/s12207-024-09525-4

REVIEW ARTICLE

The (Small) World of Validity Test Research According to Leonhard and Leonhard (2024): A Critique

Harald Merckelbach[1] · **Brechje Dandachi-FitzGerald**[1,2]

Received: 13 August 2024 / Accepted: 24 October 2024 / Published online: 9 November 2024

Abstract

The paper "Neuropsychological malingering determination: The illusion of scientific lie detection" by Chunlin Leonhard and Christoph Leonhard (2024) critically assesses the use of symptom and performance validity tests (SVTs/PVTs) in forensic settings. The authors argue that the research community's lack of critical examination leads to a flawed peer review process and scientifically dubious SVTs/PVTs, making them unsuitable for expert testimony. We comment on the arguments presented by Leonhard and Leonhard, questioning the scientific rigor of their approach, the limited scope of their literature review, their mischaracterization of validity tests as "malingering tests," and their naive reliance on medical standards for evaluating SVTs/PVTs. We assert that referring to validity tests as "malingering tests" is a straw-man argument, as SVTs/PVTs are designed not to detect malingering per se, but to identify symptom overreporting or cognitive underperformance, respectively. In contrast to Leonhard and Leonhard's stance on the inadmissibility of validity tests, other researchers offer a more balanced perspective, indicating that SVTs/PVTs, despite their limitations, receive favorable reviews and general acceptance in the field. In conclusion, we find Leonhard and Leonhard's analysis unconvincing and argue that their questionable arguments undermine the credibility of symptom and performance validity research. The potential consequences of this include diminished funding prospects. We emphasize that SVTs/PVTs provide valuable insights into symptom overreporting and cognitive underperformance, which are crucial for accurate diagnosis and treatment.

Keywords Symptom validity tests · Malingering · Feigning · Daubert · SVT · PVT

The paper titled "Neuropsychological malingering determination: The illusion of scientific lie detection" by Chunlin Leonhard and Christoph Leonhard (2024) seeks to offer a comprehensive critique of the role of symptom and performance validity tests in the forensic domain, presenting three key conclusions:

1. The community of researchers producing papers on validity tests[1] lacks critical scrutiny, resulting in a deficient peer review process.
2. Consequently, the scientific robustness of these tests is dubious, making them unsuitable for basing expert testimony.
3. Therefore, validity tests cannot meet Daubert standards in US courts, and expert testimonies relying on such tests should be inadmissible; instead, triers of fact should consider it their privilege to assess the credibility of defendants or plaintiffs.

In this paper, we critically evaluate these assertions from the standpoint of European researchers interested in symptom and performance validity assessment. We emphasize that we have no formal affiliations with the US researchers mentioned by Leonhard and Leonhard nor have we served as expert witnesses in US court cases or supported opinions

✉ Harald Merckelbach
H.Merckelbach@maastrichtuniversity.nl

[1] Section Forensic Psychology, Faculty of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD Maastricht, Netherlands

[2] Department of Clinical Psychology, Faculty of Psychology, Open University, Heerlen, Netherlands

[1] To enhance readability, we use the term "validity test(s)" broadly, encompassing both tools that assess the validity of subjective symptom reports (i.e., symptom validity tests (SVTs)) and tasks that evaluate the validity of test performance (i.e., performance validity tests (PVTs); Larrabee, 2012).

through legal briefs to US judges on matters related to symptom and performance validity.

Before addressing the key conclusions of Leonhard and Leonhard (2024), we highlight four preliminary concerns. First, we question the scientific rigor with which these authors approach their subject matter. Second, we critically address the breadth of literature they engage with. Third, we take issue with their use of "malingering tests" as a strawman term for validity tests. Fourth, we criticize their naïve reliance on medical standards for evaluating validity tests.

Questionable Scientific Rigor

Leonhard and Leonhard's paper exhibits lapses in scientific rigor at various points. Phrases such as "advanced imaging technology allows scientists to observe brain waves" (Leonhard & Leonhard, 2024, p. 495) do not inspire confidence. The same applies to the authors' discussions on the psychometrics of validity tests. The extensive argument presented by Leonhard and Leonhard (2024; pp. 524, 525) regarding the difference between construct validity and predictive validity—claiming that predictive validity is not demonstrated for validity tests—appears outdated and superficial. This is largely because the authors position themselves as the most authoritative source on the subject while overlooking recent literature. This newer research indicates that the core concept in this context should be "low-quality data," and ignoring this concept has significant implications (e.g., Chandler et al., 2020). The key feature here could be referred to as "consequential validity" (Iliescu & Greiff, 2021). Ultimately, the established value of validity tests lies in their ability to identify validity issues—essentially, "dirty data"—in scores obtained from other tests. We will explore this topic further below.

Even more concerning is their tendency to both overinterpret and misinterpret existing literature. For instance, Leonhard and Leonhard (2024; p. 522 and footnote 179; p. 538 and footnote 254) misrepresent a study by Suhr and Spickard (2012), suggesting that these authors observed a relationship between cogniphobia and poor performance on validity tests, when in reality, their study found no difference in cogniphobia status between those who failed and those who passed a PVT. In passing, we note that even if cross-sectional research were to find cogniphobia to be significantly associated with failure on a validity test (e.g., Henry et al., 2018), the causal meaning of such correlation would be unclear. Cogniphobia could potentially lead to abnormal scores on validity tests, but it might also be the case that the tendency to exaggerate symptoms, assessed by these tests, extends to overreporting on measures of cogniphobia, which could then be wrongly interpreted as cogniphobia. And, of course, both causal pathways could exist simultaneously. Nevertheless, Leonhard and Leonhard overinterpret the null findings of Suhr and Spickard (2012) in one direction (i.e., cogniphobia fosters symptom overreporting), molding their interpretation to align with their agenda.

One instance of misinformation arises in Leonhard and Leonhard's (2024; p. 541 and footnote 272) discussion regarding the notion of a cry for help as a primary factor in a validity test failure. They seem to imply that an article by Dandachi-FitzGerald et al. (2022) supports this view. However, in reality, this article critically evaluates the concept and finds scant evidence to substantiate it.

The lack of scientific rigor becomes most apparent when Leonhard and Leonhard (2024) fail to provide empirical support for their claims. They argue that validity tests are flawed instruments leading to incorrect legal outcomes. This assertion is open to empirical evaluation. For instance, how often are validity tests utilized in criminal or civil cases? Is there a relationship between their usage and appeals? In documented instances of miscarriages of justice, did validity tests contribute to those errors? These questions can be addressed with available data sets, yet Leonhard and Leonhard made no effort in this regard. In fact, the authors do not present even one compelling case in which validity tests contributed to an incorrect judicial decision. Why is that?

Limited Scope of Literature Review

Leonhard and Leonhard (2024) predominantly focus on a highly selective collection of papers from North American sources, neglecting studies from other regions (e.g., Merten et al., 2013, 2022). This narrow perspective leads them to overlook relevant empirical and review papers that could enrich their analysis and challenge their assertions regarding the scientific quality of symptom and performance validity research. For instance, Leonhard and Leonhard (2024) propose that authors within the field have insufficiently recognized the interpretational difficulties surrounding the term "malingering." Conversely, UK authors like Rix and Tracy (2017) have extensively argued for cautious use of this term in forensic contexts. They advocate focusing on arguments regarding whether or not a defendant or plaintiff exhibits specific disorders, impairments, or symptoms. Within this framework, these authors emphasize the added value of validity tests, albeit acknowledging that such tests offer only one source of pertinent information (also see Bass & Halligan, 2014; Bush et al., 2014; Rubenzer, 2020). Arguably, assessing whether the symptoms presented by defendants or plaintiffs warrant a particular diagnosis does not encroach upon the prerogative of triers of fact to determine the honesty of the defendant or plaintiff.

Another illustration comes from the research of German-based scholar Thomas Merten, which opposes the assertion

by Leonhard and Leonhard (2024) that professionals in the field have massively overlooked the constraints of validity tests. Merten (2023) showed that a particular performance validity test frequently employed to assess the exaggeration of post-traumatic stress disorder could yield an unacceptably high rate of false positives (34%) in older patients with neurological conditions. Leonhard and Leonhard (2024) disregard this paper and numerous others (e.g., Dandachi-FitzGerald & Merckelbach, 2013; Lilienfeld et al., 2013; McWhirter et al., 2020) that acknowledge a robust understanding of the limitations and hazards associated with validity tests within the field. In fact, the diagnostic potential of validity testing has been critically examined in the field for some time, including by US-based researchers such as Bianchini et al. (2001). Therefore, the assertion by Leonhard and Leonhard (2024) to the contrary does not accurately reflect the current state of the field (see also Davis, 2018; Moore et al., 2021; Sweet et al., 2021).

Yet another example is the investigation conducted by Dandachi-FitzGerald et al., (2017) regarding the ability of neuropsychologists to anticipate distorted symptom presentation. These researchers highlighted the corrective role of validity tests in cases where clinicians—or forensic experts—initially suspect symptom exaggeration, but subsequent tests reveal no basis for such concerns (also see Niesten et al., 2022). We emphasize this study because Leonhard and Leonhard (2024; p. 542/543) overlook this line of research on non-deviant outcomes of validity tests, instead merely accusing experts "to cherry pick results to support a malingering determination."

We could continue to present numerous additional examples to emphasize that by confining their analysis to a narrowly selected subset of primarily US-based research papers and court decisions, Leonhard and Leonhard (2024) portray a distorted representation of research in the domain of symptom and performance validity. Their assertion (Leonhard & Leonhard, 2024; footnote 312) that "(T)here are some dissenting voices, but those voices are clearly overwhelmed by the enthusiastic endorsement" lacks the necessary nuance.

Malinger Tests as Straw-man Argument

Leonhard and Leonhard (2024) consequently refer to validity tests as malingering tests. It is true that the names of older tests, like the Structured Inventory of Malingered Symptomatology (SIMS; Smith & Burger, 1997) and the Test of Memory Malingering (TOMM; Tombaugh, 1997) have fostered the misconception that validity tests are designed to detect malingering. However, their capability—though imperfect—is limited to identifying instances of symptom overreporting or underperformance on cognitive tasks. This point has been made repeatedly by multiple authors in early (e.g., Larrabee, 2012) as well as recent papers (e.g., Basso et al., 2024). Already a decade ago, Bush et al. (2014) stressed this point in their position paper, writing that "(B)ecause no tests have been developed that can capture malingering per se in a reliable and valid manner, caution is indicated when inferring meaning from scores that fall in ranges that are consistent with invalid performance." In a similar vein, Merten and Merckelbach (2013; p. 122) summarized the historical trajectory and prevailing consensus within this field as follows: "In the 1980s and early 1990s, symptom validity tests (SVTs) were conceived as malingering tests. In the process of conceptual clarification and refinement that followed, this idea was largely abandoned. Many experts now would agree that.

(a) SVTs may help to clarify the nature of certain symptom constellations
(b) symptom validity assessment comprises both self-report measures that tap over-endorsement of symptoms and tasks (i.e., "effort tests") that tap cognitive underperformance (Larrabee, 2012, referred to these cognitive SVTs as performance validity tests)
(c) symptom over-endorsement and/or cognitive underperformance represent two aspects of negative response bias. In some cases, they occur together, in other cases only one of the two aspects is present (Iverson, 2006)
(d) malingering is considered to be only one possible source of negative response bias."

Note that since the publication of this paper, the terminology has been further refined, with the superordinate term SVTs now more strictly reserved for self-report symptom validity tests while cognitive underperformance tests referred to as PVTs, providing a more accurate description of what these distinct tests measure. This development reflects the field's progress toward a more precise and nuanced understanding of validity assessments.

Leonhard and Leonhard (2024; p. 551) ignore this conceptual refinement and assert that labels such as "symptom validity tests" conceal "the true nature of those tests," which, according to these authors, is "to support a positive finding of malingering." This argumentative tactic allows the authors to draw two extravagant inferences. The first is that validity tests "are statistically speaking redundant tests, i.e., they are functionally the same tests with different labels" (Leonhard & Leonhard, 2024, p. 519). This conclusion is contradicted by factor analytic studies (e.g., Nelson et al., 2007; Van Dyke et al., 2013; for a more recent paper, see Ord et al., 2021). These studies show that SVTs and PVTs load onto distinct constructs, with the most consistently replicated constructs being symptom overreporting and cognitive underperformance.

The second dubious inference is related to Leonhard and Leonhard's (2024) ill-founded idea that SVTs and PVTs would only demonstrate validity when these tests accurately predict (future) instances of malingering. Since studies substantiating such predictive accuracy are lacking—owing to the sheer impossibility to create an independently verified reference standard (i.e., individuals who later on admit to malingering)—Leonhard and Leonhard (2024; p. 596) dismiss these tests as mere "subjective speculations."

Remarkably, Leonhard and Leonhard (2024) fail to adhere to their own reasoning. The lack of a "gold standard" would suggest that one cannot ascertain a case of malingering, but likewise, it would preclude the determination that malingering is overdiagnosed within a specific sample. However, they appear to align with the conclusion of the Madigan Army Medical Center (MAMC) that malingering was "over-diagnosed" among the soldiers treated at this facility (Leonhard & Leonhard, 2024; p. 523 and footnote 182).

Malingering is a harsh reality in both clinical and forensic contexts: some patients intentionally fabricate or exaggerate symptoms, and certain defendants or litigants do the same, potentially on a wider scale (e.g., Bass & Halligan, 2014; Matto et al., 2019; Pierre, 2019). While individuals who malinger exaggerate symptoms and/or underperform on cognitive tasks, the reverse is not true: not everyone who reports excessive symptoms and/or underperforms is a malingerer (Merckelbach et al., 2019). Validity tests often effectively identify instances of symptom overreporting or cognitive underperformance, but they cannot—as most neuropsychologists and forensic psychologists would agree (e.g., see consensus statement in Sweet et al., 2021)—discern the underlying reasons for such behavior. Not surprisingly, surveys among neuropsychologists indicate that the overwhelming majority are reluctant to use the term "malingering" in cases where examinees fail validity tests (Martin et al., 2015). Therefore, requiring validity tests to show predictive accuracy in identifying future cases of self-confessed malingering misses the point.

Naive Reliance on Medical Standards

Leonhard and Leonhard (2024) strongly associate malingering with the forensic field, mentioning it alongside lie detection. It is therefore all the more remarkable that they want to subject validity tests to quality assessments derived from medicine. Thus, these authors repeatedly insist that validity tests must adhere to the Standards for the Reporting of Diagnostic Accuracy (STARD) and the Quality Assessment of Diagnostic Studies (QUADAS), which are widely used by the medical community. These standards apply to laboratory tests and other technologies used to detect medical conditions such as tuberculosis, urinary tract infections, prostate cancer, shoulder pain, bacterial infections, lumbar fusion, multiple sclerosis, and osteoporosis. Since validity tests do not meet these standards—specifically, studies assessing their diagnostic accuracy are not based on large groups of self-confessed malingerers—Leonhard and Leonhard (2024, p. 511) conclude that validity tests "are not scientifically validated diagnostic tests." Their reasoning is flawed. Malingering is neither a somatic disease nor a psychiatric condition (Bass & Wade, 2019), a point well recognized by the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM–5; American Psychiatric Association, 2013; see also McDermott & Scott, 2015). Furthermore, SVTs/PVTs are designed to detect symptom overreporting and underperformance, respectively, rather than malingering. Consequently, predictive validity is indicated by the extent to which deviant scores on these instruments correlate with markedly abnormal scores on standard clinical scales and tasks. Numerous studies, many of European origin, have documented this type of interrelationship (e.g., Copeland et al., 2016; Dandachi-FitzGerald et al., 2011, 2016; Fuermaier et al., 2023; Kirkwood et al., 2012; Merten et al., 2020; Wisdom et al., 2014). Likewise, it is well established that the *number* of failed PVTs predicts abnormally low performance across various cognitive tests (e.g., Finley et al., 2024). In this context, the large-scale study by Rohling et al. (2024), which included over 5000 participants, is worth mentioning. These authors observed a robust connection between the number of validity tests failed and overall lower test battery performance.

In their comprehensive critique of the literature on validity tests, McWhirter et al., (2020, p. 950; see also Rubenzer, 2020) acknowledge this point by stating: "Finally, it is important to remember that the key purpose of validity tests should be not to assess the validity of the person being tested, but the validity of the results of other neuropsychological tests. While what we are measuring in PVTs remains unclear, what is much clearer is that poor performance on PVTs renders other neuropsychological tests invalid." Leonhard and Leonhard (2024) fail to grasp the crux of symptom and performance validity testing, which becomes particularly evident when they compare it to the control strip of COVID-19 rapid test devices. If the control line on these devices does not appear, it indicates that the test device is faulty, but it does not affect the validity of other tests such as ECGs or blood pressure measurements. In contrast, when patients, defendants, or claimants consistently produce abnormal scores on validity tests, it casts doubt on the accuracy of their self-reported mental health assessments and neuropsychological task performance.

Peer Review is not a Closed Shop

Leonhard and Leonhard (2024) suggest that articles on validity tests are generated by a highly overlapping group of researchers, journal editors, and expert witnesses. Dissenting opinions would be rare among this group, whereas uncritical acceptance of validity tests would be the rule. Leonhard and Leonhard (2024, p. 545) conclude: "The peer review system the Daubert standard relied on did not function as intended in this case. Neuropsychologists published thousands of articles on the Malingering Tests and hundreds of studies purporting to validate the Malingering Tests, using seemingly scientific methods. This extensive publication record creates the illusion of peer review."

We would like to counter by noting that critical papers have long been a part of the tradition in this field (e.g., Bianchini et al., 2001). Comprehensive discussions on issues such as false-positive rates, multiple tests, criterion grouping in diagnostic accuracy studies, and the role of digital technology are abundant (Berthelson et al., 2013; Davis & Millis, 2014; Farkas et al., 2006; Finley, 2024; Larrabee, 2014; Rhoads et al., 2021; Schroeder et al., 2019; Soble et al., 2020). Additionally, there is self-reflection on the field's neglect of negative predictive power (e.g., Niesten et al., 2022) and base rate issues (Rosenfeld et al., 2000), as well as considerations of the risk of false positives in forensic populations (Finley et al., 2023) and problematic cross-cultural generalizations (e.g., Nijdam-Jones & Rosenfeld, 2017). These elements collectively characterize a research domain with ample room for diverse opinions and criticism.

While it is true that "some neuropsychologists who testify regularly as experts are also members of the editorial boards of most journals" (Leonhard & Leonhard, 2024, p. 548 and footnote 310), it would be incorrect to interpret this as evidence that the typical review process in this field operates in a closed-shop atmosphere. The editorial board of *Psychological Injury and Law* is a case in point. It not only lists Christoph Leonard himself as a contributing editor but also Anselm Fuermaier (The Netherlands). Among the section head editors of this journal, we find names such as Esteban Puente López (Spain), Thomas Merten (Germany), Irena Boskovic (The Netherlands), and Francesca Ales (Italy). One of the associate editors is Luciano Giromini (Italy). It seems safe to assume that these European scholars will rarely if ever appear as expert witnesses in US courts of law. It is equally safe to assume that they have no vested interest in the outcomes of US legal disputes regarding validity tests and their admissibility. Leonhard and Leonhard's (2024) claim that the peer review system in this research domain is substandard is a serious one. Therefore, it requires solid empirical evidence, not just opinions, anecdotal evidence, or idiosyncratic reasoning. Leonhard and Leonhard (2024) do not provide such evidence, which is disconcerting when one considers that they are publishing in a legal journal, where editors and reviewers are expected to set requirements for substantiating opinions.

Scientific Robustness of Symptom and Performance Validity Tests

Leonhard and Leonhard (2024, p. 511) argue that research on symptom and performance validity assessment is plagued with statistical and methodological flaws, leading them to conclude that "the Malingering Tests are not scientifically validated diagnostic tests." They reiterate this conclusion throughout their article, primarily supported by frequent references to two commentaries by Christoph Leonhard *lui-même* (Leonhard, 2023a, b; see Young & Erdodi, 2024 for a critique). We agree that there are problems that limit the generalizability of results obtained in single studies on symptom overreporting and/or cognitive underperformance. However, the body of knowledge in symptom and performance validity assessment is derived from three types of research designs (Rogers et al., 1993; Schroeder et al., 2019): experimental simulation studies, differential prevalence studies, and studies with variously defined criterion groups. Although each design has its restrictions, meta-analytic studies that synthesize and evaluate results from all three design types provide reasonably robust data on their psychometric qualities. Such meta-analyses can also offer valuable insights to legal system actors (Blumenthal, 2007). Without aiming for completeness, meta-analytic evaluations have been conducted for various validity tests. These include the Structured Inventory of Malingered Symptomatology (SIMS; Van Impelen et al., 2014; Shura et al., 2022), the Miller Forensic Assessment of Symptoms (M-Fast; Detullio et al., 2019), the Personality Assessment Inventory (PAI; Hawes & Boccaccini, 2009; Kurtz & Pintarelli, 2024), the Inventory of Problems-29 (IOPs-29; Puente-López et al., 2023), the validity scales of the MMPI (Aparcero et al., 2023), the Advanced Clinical Solutions Word Choice Test (Bernstein et al., 2021), and the Effort Index of the Repeatable Battery for the Assessment of Neuropsychological Status (Goette & Goette, 2019; Shura et al., 2018). Other meta-analytic studies have compared data from various validity tests (e.g., Sollman & Berry, 2011; Vickery et al., 2001). Additionally, some have focused on populations with a higher probability of false-positive outcomes, such as patients with psychotic symptomatology (Ruiz et al., 2020), while others have attempted to estimate base rates of validity test failure in different clinical settings

(Roor et al., 2024). Recent studies in this domain commit to the values of open science by preregistering in public databases such as Prospero (Roor et al., 2024; Ruiz et al., 2020) and OSF (e.g., Robinson et al., 2023). Together and viewed in conjunction, these meta-analytic studies justify the conclusion that various validity tests have a reasonable track record and are well anchored in the scientific literature. Claiming that "these purportedly objective scientific tests turn out to be only scientific fiction" (Leonhard & Leonhard, 2024; p. 562) does not do justice to this collection of meta-analytic studies.

Admissibility Under Daubert

As European researchers, we do not have an opinion on whether results obtained with validity tests should be admissible as evidence in American courtrooms. We lack the knowledge and authority regarding Daubert-like guidelines and how they should be applied. Leonhard and Leonhard (2024) do have a clear stance on this matter: because of their poor scientific underpinning, the results of validity tests should not be admissible, yet the Daubert rule has so far failed to keep such results out of American courtrooms. Other authors (e.g., Greiffenstein & Kaufmann, 2012; Kaufmann, 2008; laDuke et al., 2018; Vallabhajosula & van Gorp, 2001) have also commented on this issue, albeit in a considerably more balanced and empirically inspired tone. It is notable—and not particularly indicative of a scholarly attitude—that Leonhard and Leonhard (2024) fail to mention these authors, let alone engage in a discussion with those who have published more recent papers on this issue. For example, looking at general acceptance in the field and psychometric reviews, Neal et al. (2019) found many weaknesses in the tests that psychologists use to address legal issues and the way courts assess those tests. However, these authors also conclude that a number of tests, including some validity tests, have favorable reviews, are generally accepted in the field, and might, in principle, be able to pass an admissibility challenge in courts (Neal et al., 2019; see also Kurtz & Pintarelli, 2024).

Concluding Remarks

Leonhard and Leonhard (2024) conclude their article by asserting that validity tests lack scientific support and, as a result, should have no place in courts of law. They argue that relying on such tests poses significant risks of injustice. Instead, they suggest that determining whether defendants or claimants are presenting their symptoms or impairments accurately should be left to the judgment of jury members, rather than relying on experts' validity assessment tools. In essence, Leonhard and Leonhard (2024) revisit an old debate in psychology: the superiority of actuarial versus intuitive decision-making (Dawes et al., 1989; Merten et al., 2022). We were under the impression that this debate had been empirically settled, with research indicating that the error rate of actuarial decision-making—meaning here reliance on validity test outcomes—is lower than that of intuitive decision-making, which in this context refers to clinical impressions of symptom distortion (Dandachi-Fitzgerald et al., 2017).

For the reasons discussed above, we find the analysis by Leonhard and Leonhard (2024) unconvincing. They will undoubtedly argue that their focus was on the American literature as it pertains to the forensic context. But that is precisely the point: when it comes to the quality of validity tests, the boundaries between the forensic and clinical domains, as well as between American and European literature, are highly artificial. For impartial scholars, such boundaries hold no relevance. Meanwhile, with their questionable analysis, Leonhard and Leonhard (2024) tarnish the reputation of symptom and performance validity research. We would like to highlight the potential consequences of this, some of which extend beyond the courtroom.

To begin with, the article by Leonhard and Leonhard (2024) perpetuates the notion that symptom and performance validity is an academically poor research area, undeserving of grant funding. Our view is that, due to existing prejudices, misconceptions, and taboos (Jelicic et al., 2018), this field is already underfunded. In Europe, we are aware of numerous large grants for research on depression, schizophrenia, anxiety disorders, eating disorders, and similar conditions. However, we know of no grants specifically for symptom and performance validity research, despite the efforts of many colleagues to secure such funding. An article like the one by Leonhard and Leonhard (2024) further diminishes the prospects for obtaining such grants.

Second, Leonhard and Leonhard (2024) argue that validity tests have no added value, yet they do not propose any alternatives. This stance implicitly suggests that we should refrain from critically evaluating the symptoms reported by patients, suspects, and claimants and instead accept their accounts at face value (see also Kats et al., 2024). Such an approach overlooks the fact that overdiagnosis and overtreatment are prevalent in certain areas, such as attention-deficit/hyperactivity disorder (ADHD) (Sadek, 2022). The rising prevalence rates of ADHD can partly be attributed to the ease with which its symptoms can be exaggerated. Exaggerating symptoms can yield significant benefits, such as obtaining stimulant medication for recreational use or sale on the black market, as well as gaining access to additional study facilities or less stringent exam requirements. Although this is an uncomfortable topic (Patel, 2023), it cannot be ignored. An article like that of Leonhard and

Leonhard (2024) encourages the tendency to completely dismiss the issue under the premise that the patient is always right. Conversely, we would argue that validity tests can improve diagnostic accuracy, laying a more robust foundation for effective therapeutic interventions (see also Horner et al., 2014; Marquardt et al., 2023).

A third consequence relates to the quality of research data. Dismissing validity tests and thereby ignoring distorted symptom presentations may obscure research outcomes. A good example comes from authors who investigated the relationship between hippocampal volume and memory performance among elderly people who attended a memory clinic (Rienstra et al., 2013). The expected relationship between hippocampal atrophy and poor memory performance was present among participants with non-deviant scores on a performance validity test but absent among participants with deviant scores. Arguably, collapsing the data of these two groups would have led to an underestimation of the link between hippocampal volume and memory performance (see also Fox, 2011). Similarly, Meyer et al. (2018) studied emotion regulation and brain asymmetry in trauma victims. They found that left frontal activation in response to negative stimulus material was related to less intense intrusions, and this connection became considerably more pronounced when the authors corrected for deviant responses on a symptom validity test. This suggests that symptom overreporting, as indicated by validity tests, may dampen true associations. Likewise, not correcting for such overreporting might lead to inflated prevalence rates of dissociative symptoms (Merckelbach et al., 2017), underestimation of the dose–response relationship between trauma intensity and post-traumatic stress symptoms (Merckelbach et al., 2014), and possibly underestimation of positive therapy outcomes (Pfeiffer et al., 2017; but see also Roor et al., 2022).

Validity tests make symptom overreporting and cognitive underperformance observable phenomena. Many experts in the field (e.g., Bigler, 2014), long before Leonhard and Leonhard (2024), have recognized that these phenomena can be caused by various factors, such as factitious motives (e.g., Chafetz et al., 2020), certain personality traits like alexithymia (Brady et al., 2017), and careless responding (e.g., Merckelbach et al., 2019), in addition to malingering. We emphasize this because Leonhard and Leonhard (2024) present the idea that deviant scores on validity tests could arise from causes other than malingering as if it were their own unique insight. They overstate their originality. Additionally, we note that many of the potential antecedents these authors mention—such as cogniphobia, distraction, exhaustion, and hostility—lack empirical evidence as sources of deviant performance on validity tests. Focusing on specifically performance validity tests, Schroeder and Martin (2022; p. 25/26) reviewed the extant literature and concluded as follows: "The empirical research also indicates that many previously proposed explanations of PVT failure (i.e., apathy, fatigue/daytime sleepiness, pain, medication effects, depression or anxiety, a "cry for help") do not actually cause PVT failure in the vast majority of cases, although rare exceptions might potentially occur when there are extreme presentations."

In high-stakes situations, where no alternative explanations exist and discrepancies or anomalies are evident, repeated failures on validity tests—especially when they involve performance below chance levels (Bianchini et al., 2005)—may suggest malingering as a probabilistic interpretation (Bush et al., 2014). This heuristic is central to the well-considered multidimensional criteria of Sherman et al. (2020). In discussing the older version of these criteria (Slick et al., 1999), Leonhard and Leonhard (2024) argue that such criteria hold little value unless they have been independently verified by self-confessed malingerers. This argument is akin to claiming that criteria for diagnosing dementia are worthless until cases are confirmed by post-mortem histopathological examination.

One of the key ideas in the Leonhard and Leonhard (2024; p. 554, 555) article is that validity tests appeal to biases and prejudices of judges and lawyers: "Fear about malingering of mental conditions is deeply rooted in Anglo-American law and culture (..). People commonly believe that human beings tend to lie to protect their own interest. The need to be able to tell when a party is lying and malingering is great. The Malingering Tests thus offer a much desired solution consistent with judges' and lawyers' world view."

With statements like these, the authors suggest that malingering only exists in the minds of biased people. However, the defendant who feigns dissociative amnesia certainly exists (Zago et al., 2024), as does the repeat offender who pretends to suffer from psychotic symptoms (Evan Jaffe & Sharma, 1998), or the person who exaggerates their cognitive problems in the context of disability benefits evaluations (Roor et al., 2016). Pretending that this is all a matter of prejudice and bias leads to malingering remaining a taboo subject, complicating clinical and legal decision-making. Both clinicians (e.g., Beach et al., 2017) and forensic experts (e.g., Van der Heide et al., 2020) have warned that this taboo can, and sometimes does, have dire consequences. Unless, of course, one finds it perfectly acceptable that people who malinger, for example, psychosis, suicidal ideation, or AIDS/HIV for that matter, are hospitalized and/or are given unnecessary and potentially harmful medication (see, for illustrations: Rumschik & Appel, 2019; Ryan et al., 1995; Waite & Geddes, 2006; Wang & Rehman, 2021). It is precisely because of these dire consequences that we thought it to be important to write this critical evaluation of the Leonhard and Leonhard (2024) paper.

Acknowledgements The authors are grateful to the anonymous reviewers for their critical comments and valuable suggestions.

Funding The authors received no funding.

Declarations

Conflict of Interest The authors have no conflict of interest to declare.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Aparcero, M., Picard, E. H., Nijdam-Jones, A., & Rosenfeld, B. (2023). Comparing the ability of MMPI-2 and MMPI-2-RF validity scales to detect feigning: A meta-analysis. *Assessment, 30*(3), 744–760.

Bass, C., & Halligan, P. (2014). Factitious disorders and malingering: Challenges for clinical assessment and management. *The Lancet, 383*(9926), 1422–1432.

Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology, 19*(2), 96–105.

Basso, M. R., Whiteside, D. M., & Combs, D. (2024). Introduction to the special issue on performance validity: What are we doing? What should we do? *Journal of Clinical and Experimental Neuropsychology, 46*(1), 1–5.

Beach, S. R., Taylor, J. B., & Kontos, N. (2017). Teaching psychiatric trainees to "think dirty": Uncovering hidden motivations and deception. *Psychosomatics, 58*(5), 474–482.

Bernstein, M. T., Resch, Z. J., Ovsiew, G. P., & Soble, J. R. (2021). A systematic review and meta-analysis of the diagnostic accuracy of the advanced clinical solutions word choice test as a performance validity test. *Neuropsychology Review, 31*(2), 349–359.

Berthelson, L., Mulchan, S. S., Odland, A. P., Miller, L. J., & Mittenberg, W. (2013). False positive diagnosis of malingering due to the use of multiple effort tests. *Brain Injury, 27*(7–8), 909–916.

Bianchini, K. J., Mathias, C. W., & Greve, K. W. (2001). Symptom validity testing: A critical review. *The Clinical Neuropsychologist, 15*(1), 19–45.

Bianchini, K. J., Greve, K. W., & Glynn, G. (2005). On the diagnosis of malingered pain-related disability: Lessons from cognitive malingering research. *The Spine Journal, 5*(4), 404–417.

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury, 28*(13–14), 1623–1638.

Blumenthal, J. A. (2007). Meta-analysis: A primer for legal scholars. *Temple Law Review, 80*, 201–235.

Brady, R. E., Bujarski, S. J., Feldner, M. T., & Pyne, J. M. (2017). Examining the effects of alexithymia on the relation between posttraumatic stress disorder and over-reporting. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(1), 80–87.

Bush, S. S., Heilbronner, R. L., & Ruff, R. M. (2014). Psychological assessment of symptom and performance validity, response bias, and malingering: Official position of the Association for Scientific Advancement in Psychological Injury and Law. *Psychological Injury and Law, 7*(3), 197–205.

Chafetz, M. D., Bauer, R. M., & Haley, P. S. (2020). The other face of illness-deception: Diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 34*(3), 454–476.

Chandler, J., Sisso, I., & Shapiro, D. (2020). Participant carelessness and fraud: Consequences for clinical research and potential solutions. *Journal of Abnormal Psychology, 129*(1), 49–55.

Copeland, C. T., Mahoney, J. J., Block, C. K., Linck, J. F., Pastorek, N. J., Miller, B. I., & Sim, A. H. (2016). Relative utility of performance and symptom validity tests. *Archives of Clinical Neuropsychology, 31*(1), 18–22.

Dandachi-FitzGerald, B., & Merckelbach, H. (2013). Feigning≠ feigning a memory deficit: The Medical Symptom Validity Test as an example. *Journal of Experimental Psychopathology, 4*(1), 46–63.

Dandachi-FitzGerald, B., Ponds, R. W., Peters, M. J., & Merckelbach, H. (2011). Cognitive underperformance and symptom overreporting in a mixed psychiatric sample. *The Clinical Neuropsychologist, 25*(5), 812–828.

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.

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.

Dandachi-FitzGerald, B., Merckelbach, H., & Merten, T. (2022). Cry for help as a root cause of poor symptom validity: A critical note. *Applied Neuropsychology: Adult., 31*(4), 527–553.

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.

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist, 28*(2), 199–214.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*(4899), 1668–1674.

Detullio, D., Messer, S. C., Kennedy, T. D., & Millen, D. H. (2019). A meta-analysis of the Miller Forensic Assessment of Symptoms Test (M-FAST). *Psychological Assessment, 31*(11), 1319.

Evan Jaffe, M., & Sharma, K. K. (1998). Malingering uncommon psychiatric symptoms among defendants charged under California's "Three Strikes and You're Out" law. *Journal of Forensic Sciences, 43*(3), 549–555.

Farkas, M. R., Rosenfeld, B., Robbins, R., & van Gorp, W. (2006). Do tests of malingering concur? Concordance among malingering measures. *Behavioral Sciences & the Law, 24*(5), 659–671.

Finley, J. C. A. (2024). Performance validity testing: The need for digital technology and where to go from here. *Frontiers in Psychology, 15*, 1452462.

Finley, J. C. A., Brook, M., Kern, D. M., Reilly, J. L., & Hanlon, R. E. (2023). Profile of embedded validity indicators in criminal defendants with verified valid neuropsychological test performance. *Archives of Clinical Neuropsychology, 38*(4), 513–524.

Finley, J. C. A., Leese, M. I., Roseberry, J. E., & Hill, S. K. (2024). Multivariable utility of the memory integrated language and making change test. *Applied Neuropsychology*: *Adult.* Advance online publication.

Fox, D. D. (2011). Symptom validity test failure indicates invalidity of neuropsychological tests. *The Clinical Neuropsychologist, 25*(3), 488–495.

Fuermaier, A. B., Dandachi-Fitzgerald, B., & Lehrner, J. (2023). Validity assessment of early retirement claimants: Symptom overreporting on the Beck Depression Inventory–II. *Applied Neuropsychology: Adult.* Advance online publication.

Goette, W. F., & Goette, H. E. (2019). A meta-analysis of the accuracy of embedded performance validity indicators from the repeatable battery for the assessment of neuropsychological status. *The Clinical Neuropsychologist, 33*(6), 1044–1068.

Greiffenstein, M. F., & Kaufmann, P. M. (2012). Neuropsychology and the law: Principles of productive attorney-neuropsychologist relations. In G. J. Larrabee (Ed.), *Forensic neuropsychology: A scientific approach* (2nd ed., pp. 23–69). Oxford University Press.

Hawes, S. W., & Boccaccini, M. T. (2009). Detection of overreporting of psychopathology on the Personality Assessment Inventory: A meta-analytic review. *Psychological Assessment, 21*(1), 112–124.

Henry, G. K., Heilbronner, R. L., Suhr, J., Gornbein, J., Wagner, E., & Drane, D. L. (2018). Illness perceptions predict cognitive performance validity. *Journal of the International Neuropsychological Society, 24*(7), 735–745.

Horner, M. D., VanKirk, K. K., Dismuke, C. E., Turner, T. H., & Muzzy, W. (2014). Inadequate effort on neuropsychological evaluation is associated with increased healthcare utilization. *The Clinical Neuropsychologist, 28*(5), 703–713.

Iliescu, D., & Greiff, S. (2021). On consequential validity. *European Journal of Psychological Assessment, 37*(3), 163–166.

Iverson, G. L. (2006). Ethical issues associated with the assessment of exaggeration, poor effort, and malingering. *Applied Neuropsychology, 13*(2), 77–90.

Jelicic, M., Merckelbach, H., & Bošković, I. (2018). Seven myths about feigning. In H. Otgaar & M. L. Howe (Eds.), *Finding the truth in the courtroom: Dealing with deception, lies, and memories* (pp. 227–242). Oxford University Press.

Kats, L., Hume, M., & Landes, J. (2024). Trial courts can reject unrebutted expert opinion. *The Journal of the American Academy of Psychiatry and the Law, 52*(3), 3–5.

Kaufmann, P. M. (2008). Admissibility of neuropsychological evidence in criminal cases: Competency, insanity, culpability, and mitigation. In R. L. Denney & J. P. Sullivan (Eds.), *Clinical neuropsychology in the criminal forensic setting* (pp. 55–90). Guilford Press.

Kirkwood, M. W., Yeates, K. O., Randolph, C., & Kirk, J. W. (2012). The implications of symptom validity test failure for ability-based test performance in a pediatric sample. *Psychological Assessment, 24*(1), 36.

Kurtz, J. E., & Pintarelli, E. M. (2024). The Daubert standards for admissibility of evidence based on the Personality Assessment Inventory. *Psychological Injury and Law, 17*, 105–116.

LaDuke, C., Barr, W., Brodale, D. L., & Rabin, L. A. (2018). Toward generally accepted forensic assessment practices among clinical neuropsychologists: A survey of professional practice and common test use. *The Clinical Neuropsychologist, 32*(1), 145–164.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society, 18*, 625–631.

Larrabee, G. J. (2014). False-positive rates associated with the use of multiple performance and symptom validity tests. *Archives of Clinical Neuropsychology, 29*(4), 364–373.

Leonhard, C. (2023a). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part I: Statistical issues. *Neuropsychology Review, 33*(3), 581–603.

Leonhard, C. (2023b). Review of statistical and methodological issues in the forensic prediction of malingering from validity tests: Part II—Methodological issues. *Neuropsychology Review, 33*(3), 604–623.

Leonhard, C., & Leonhard, C. (2024). Neuropsychological malingering determination: The illusion of scientific lie detection. *Georgia Law Review, 58*(2), 483–571.

Lilienfeld, S. O., Thames, A. D., & Watts, A. L. (2013). Symptom validity testing: Unresolved questions, future directions. *Journal of Experimental Psychopathology, 4*(1), 78–87.

Marquardt, C. A., Ferrier-Auerbach, A. G., Schumacher, M. M., & Arbisi, P. A. (2023). MMPI-2-RF validity scales add utility for predicting treatment engagement during partial psychiatric hospitalizations. *Psychological Assessment, 36*(2), 124–133.

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.

Matto, M., McNiel, D. E., & Binder, R. L. (2019). A systematic approach to the detection of false PTSD. *The Journal of the American Academy of Psychiatry and the Law, 47*(3), 325–334.

McDermott, B., & Scott, C. (2015). DSM-5 and malingering. In C. Scott (Ed.), *DSM-5® and the law: Changes and challenges* (pp. 242–267). Oxford University Press.

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations—A systematic review. *Journal of Neurology, Neurosurgery & Psychiatry, 91*(9), 945–952.

Merckelbach, H., Langeland, W., de Vries, G., & Draijer, N. (2014). Symptom overreporting obscures the dose–response relationship between trauma severity and symptoms. *Psychiatry Research, 217*(3), 215–219.

Merckelbach, H., Boskovic, I., Pesy, D., Dalsklev, M., & Lynn, S. J. (2017). Symptom overreporting and dissociative experiences: A qualitative review. *Consciousness and Cognition, 49*, 132–144.

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.

Merten, T. (2023). Comparative data for the Morel emotional numbing test: High false-positive rate in older bona-fide neurological patients. *Psychological Injury and Law, 16*(1), 49–60.

Merten, T., & Merckelbach, H. (2013). Symptom validity testing in somatoform and dissociative disorders: A critical review. *Psychological Injury and Law, 6*, 122–137.

Merten, T., Dandachi-FitzGerald, B., Hall, V., Schmand, B. A., Santamaríae, P., & González-Ordi, H. (2013). Symptom validity assessment in European countries: Development and state of the art. *Clínica y Salud, 24*(3), 129–138.

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.

Merten, T., Dandachi-FitzGerald, B., Hall, V., Bodner, T., Giromini, L., Lehrner, J., & Di Stefano, G. (2022). Symptom and performance validity assessment in European countries: An update. *Psychological Injury and Law, 15*(2), 116–127.

Meyer, T., Quaedflieg, C. W., Weijland, K., Schruers, K., Merckelbach, H., & Smeets, T. (2018). Frontal EEG asymmetry during symptom provocation predicts subjective responses to intrusions in survivors with and without PTSD. *Psychophysiology, 55*(1), e12779.

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

Neal, T. M., Slobogin, C., Saks, M. J., Faigman, D. L., & Geisinger, K. F. (2019). Psychological assessments in legal contexts: Are courts keeping "junk science" out of the courtroom? *Psychological Science in the Public Interest, 20*(3), 135–164.

Nelson, N. W., Sweet, J. J., Berry, D. T., Bryant, F. B., & Granacher, R. P. (2007). Response validity in forensic neuropsychology: Exploratory factor analytic evidence of distinct cognitive and psychological constructs. *Journal of the International Neuropsychological Society, 13*(3), 440–449.

Niesten, I. J., Merckelbach, H., Dandachi-FitzGerald, B., Jutten-Rooijakkers, I., & van Impelen, A. (2022). Experts' failure to consider the negative predictive power of symptom validity tests. *Frontiers in Psychology, 13*, 789762.

Nijdam-Jones, A., & Rosenfeld, B. (2017). Cross-cultural feigning assessment: A systematic review of feigning instruments used with linguistically, ethnically, and culturally diverse samples. *Psychological Assessment, 29*(11), 1321.

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology, 35*(3), 241–251.

Patel, G. R. J. (2023). Feigning ADHD: A necessary exploration of an uncomfortable topic. *Journal of the New Zealand College of Clinical Psychologists, 33*(1), 61–71.

Pfeiffer, W., Gutermann, D., & Bassler, M. (2017). Beschwerdenvalidierungin der psychosomatischen Rehabilitation: Der SFSS in einer klinischen Stichprobe [Symptom validation in psychosomatic rehabilitation: The SIMS in a clinical sample] [Poster presentation]. World Congress of Psychiatry.

Pierre, J. M. (2019). Assessing malingered auditory verbal hallucinations in forensic and clinical settings. *The Journal of the American Academy of Psychiatry and the Law, 47*(4), 448–456.

Puente-López, E., Pina, D., López-Nicolás, R., Iguacel, I., & Arce, R. (2023). The Inventory of Problems–29 (IOP-29): A systematic review and bivariate diagnostic test accuracy meta-analysis. *Psychological Assessment, 35*(4), 339.

Rhoads, T., Neale, A. C., Resch, Z. J., Cohen, C. D., Keezer, R. D., Cerny, B. M., & Soble, J. R. (2021). Psychometric implications of failure on one performance validity test: A cross-validation study to inform criterion group definition. *Journal of Clinical and Experimental Neuropsychology, 43*(5), 437–448.

Rienstra, A., Groot, P. F., Spaan, P. E., Majoie, C. B., Nederveen, A. J., Walstra, G. J., & Schmand, B. (2013). Symptom validity testing in memory clinics: Hippocampal-memory associations and relevance for diagnosing mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 35*(1), 59–70.

Rix, K. J., & Tracy, D. K. (2017). Malingering mental disorders: Medicolegal reporting. *Bjpsych Advances, 23*(2), 115–122.

Robinson, A., Finley, J., Basurto, K., Cerny, B. M., Ovsiew, G. P., Phillips, M. S., & Soble, J. R. (2023). *Racial and ethnic representation of U.S. adults within PVT classification accuracy studies: A systematic review*. Retrieved from staging.osf.io/kc8rf

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13*(3), 255–274.

Rohling, M. L., Binder, L. M., Larrabee, G. J., & Langhinrichsen-Rohling, J. (2024). Forced choice test score of p≤. 20 and failures on≥ six performance validity tests results in similar Overall Test Battery Means. *The Clinical Neuropsychologist, 38*(5), 1193–1209.

Roor, J. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2016). A case of misdiagnosis of mild cognitive impairment: The utility of symptom validity testing in an outpatient memory clinic. *Applied Neuropsychology: Adult, 23*(3), 172–178.

Roor, J. J., Dandachi-FitzGerald, B., Peters, M. J., Knoop, H., & Ponds, R. W. (2022). Performance validity and outcome of cognitive behavior therapy in patients with chronic fatigue syndrome. *Journal of the International Neuropsychological Society, 28*(5), 473–482.

Roor, J. J., Peters, M. J., Dandachi-FitzGerald, B., & Ponds, R. W. (2024). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review, 34*(1), 299–319.

Rosenfeld, B., Sands, S. A., & Gorp, W. G. V. (2000). Have we forgotten the base rate problem? Methodological issues in the detection of distortion. *Archives of Clinical Neuropsychology, 15*(4), 349–359.

Rubenzer, S. (2020). The case for assessing for negative response bias, not malingering. *Journal of Forensic Psychology Research and Practice, 20*(4), 323–340.

Ruiz, I., Raugh, I. M., Bartolomeo, L. A., & Strauss, G. P. (2020). A meta-analysis of neuropsychological effort test performance in psychotic disorders. *Neuropsychology Review, 30*, 407–424.

Rumschik, S. M., & Appel, J. M. (2019). Malingering in the psychiatric emergency department: Prevalence, predictors, and outcomes. *Psychiatric Services, 70*(2), 115–122.

Ryan, J., Taylor, C. B., Bryant, G. D., Nayagam, A. T., & Sanders, J. (1995). HIV malingering in the accident and emergency department. *Emergency Medicine Journal, 12*(1), 59–61.

Sadek, J. (2022). Malingering and stimulant medications abuse, misuse and diversion. *Brain Sciences, 12*(8), 1004.

Schroeder, R. W., & Martin, P. K. (2022). 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). Guildford Press.

Schroeder, R. W., Martin, P. K., Heinrichs, R. J., & Baade, L. E. (2019). Research methods in performance validity testing studies: Criterion grouping approach impacts study outcomes. *The Clinical Neuropsychologist, 33*(3), 466–477.

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.

Shura, R. D., Brearly, T. W., Rowland, J. A., Martindale, S. L., Miskey, H. M., & Duff, K. (2018). RBANS validity indices: A systematic review and meta-analysis. *Neuropsychology Review, 28*(3), 269–284.

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.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist, 13*(4), 545–561.

Smith, G. P., & Burger, G. K. (1997). Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). *Journal of the American Academy of Psychiatry and the Law, 25*(2), 183–189.

Soble, J. R., Alverson, W. A., Phillips, J. I., Critchfield, E. A., Fullen, C., O'Rourke, J. J., & Marceaux, J. C. (2020). Strength in numbers or quality over quantity? Examining the importance of criterion measure selection to define validity groups in performance validity test (PVT) research. *Psychological Injury and Law, 13*, 44–56.

Sollman, M. J., & Berry, D. T. (2011). Detection of inadequate effort on neuropsychological testing: A meta-analytic update and extension. *Archives of Clinical Neuropsychology, 26*(8), 774–789.

Suhr, J., & Spickard, B. (2012). Pain-related fear is associated with cognitive task avoidance: Exploration of the cogniphobia construct in a recurrent headache sample. *The Clinical Neuropsychologist, 26*(7), 1128–1141.

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

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260–268.

Vallabhajosula, B., & van Gorp, W. G. (2001). Post-Daubert admissibility of scientific evidence on malingering of cognitive deficits. *The Journal of the American Academy of Psychiatry and the Law, 29*(2), 207–215.

van der Heide, D., Boskovic, I., van Harten, P., & Merckelbach, H. (2020). Overlooking feigning behavior may result in potential harmful treatment interventions: Two case reports of undetected malingering. *Journal of Forensic Sciences, 65*(4), 1371–1375.

Van Dyke, S. A., Millis, S. R., Axelrod, B. N., & Hanks, R. A. (2013). Assessing effort: Differentiating performance and symptom validity. *The Clinical Neuropsychologist, 27*(8), 1234–1246.

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.

Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology, 16*(1), 45–73.

Waite, S., & Geddes, A. (2006). Malingered psychosis leading to involuntary psychiatric hospitalization. *Australasian Psychiatry, 14*(4), 419–421.

Wang, A. Y., & Rehman, U. H. (2021). An unusual presentation of catatonia-like behavior: Differentiating malingering from catatonia. *Case Reports in Psychiatry, 2021*(1), 1860757.

Wisdom, N. M., Pastorek, N. J., Miller, B. I., Booth, J. E., Romesser, J. M., Linck, J. F., & Sim, A. H. (2014). PTSD and cognitive functioning: Importance of including performance validity testing. *The Clinical Neuropsychologist, 28*(1), 128–145.

Young, G., & Erdodi, L. (2024). Forensic prediction of malingering from performance validity tests: Review of Leonhard (2023, a, b, c). *Psychological Injury and Law, 17*, 129–138.

Zago, S., Preti, A. N., Difonzo, T., D'Errico, A., Sartori, G., Zangrossi, A., & Bolognini, N. (2024). Two cases of malingered crime-related amnesia. *Topics in Cognitive Science*, *16*, 752–769.