BJPsych Advances (2026), vol. 32, 221–230 doi: 10.1192/bja.2025.10177

The place of validity tests in psychiatric diagnosis: beyond common misconceptions

ARTICLE

Harald Merckelbach, **Brechje Dandachi FitzGerald** & **Jan Wise**

Harald Merckelbach is a professor of legal psychology in the Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands.
Brechje Dandachi FitzGerald is a professor of psychotherapy in the Department of Clinical Psychology, Open University, Heerlen, The Netherlands, and an associate professor in the Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands; she is a licensed clinical psychologist and psychotherapist and a director of the postdoctoral psychotherapy programme at RINO South, Eindhoven, The Netherlands.
Jan Wise is a consultant in general adult psychiatry at the Central and North West London NHS Foundation Trust, London, UK.
Correspondence Harald Merckelbach.
Email: h.merckelbach@maastrichtuniversity.nl

First received 24 Jun 2025
Revised 1 Oct 2025
Accepted 7 Oct 2025

Copyright and usage
© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal College of Psychiatrists. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

SUMMARY

Validity tests are used in both forensic and clinical settings, but their application in clinical practice is often hindered by misconceptions. These include the assumptions that validity tests imply a medico-legal dimension, primarily detect feigning or malingering, and provide minimal actionable information to clinicians. The authors critically discuss these misconceptions and argue that validity tests may offer significant value in clinical practice by assessing whether patients can describe their symptoms, complaints and impairments with reasonable accuracy, which has important implications for diagnosis and treatment planning. Importantly, in clinical practice, when interpreting validity tests, neutral terminology such as 'over-reporting' and 'underperformance' is often preferable to – and better to substantiate than – terms like 'feigning' and 'malingering', which can evoke moral judgements, creating an unnecessary barrier to using these valuable clinical tools.

LEARNING OBJECTIVES

After reading this article you will be able to:
- explain the type of information that validity tests provide to clinicians and how this information supports clinical decision-making
- explain the difference between 'symptom over-reporting', 'underperformance', 'feigning' and 'malingering'
- explain the distinct functions and significance of validity tests in forensic versus non-forensic settings.

KEYWORDS

Symptom validity tests (SVTs); performance validity tests (PVTs); symptom over-reporting; underperformance; feigning.

When patients undergo psychiatric assessment, they are routinely asked to describe their symptoms, difficulties and functional limitations. Although self-reported information is central to clinical assessment, clinicians may at times question whether patients' symptom reports are within a plausible range and align with data from other sources. In such cases, validity tests can provide valuable guidance, as they help to identify distorted symptom expressions (Tracy 2014).

Two types of validity test

Validity tests fall into two broad categories (Young 2025): symptom validity tests (SVTs) and performance validity tests (PVTs). SVTs assess the accuracy of self-reported symptoms by detecting patterns suggestive of symptom over-reporting. Examples include the Structured Inventory of Malingered Symptomatology (SIMS) and the validity scales embedded within the Minnesota Multiphasic Personality Inventory (MMPI-2/MMPI-2-RF). Typically, SVTs combine both common and rare symptoms. The rare symptoms are those shown by research to occur only infrequently in clinical populations. Illustrative items include 'I have headaches so severe my feet hurt', 'The buzzing in my ears keeps switching from left to right' or 'I often black out when I sit down'. When patients endorse these symptoms at rates exceeding empirically established thresholds, this is taken as an indicator of symptom over-reporting.

PVTs are tasks designed to assess cognitive functions such as memory and reasoning. Notable examples are the Amsterdam Short-Term Memory Test (ASTM) and the Test of Memory Malingering (TOMM; Martin 2020). Although PVTs may seem challenging at first glance, most individuals find them easy to complete. For example, in a typical PVT trial, the person is first shown five closely related words – say, trousers, skirt, shirt, sweater, coat – and asked to read them aloud while memorising them. Next, the person completes a simple distraction task – for instance, 12 + 7 =. Finally, the person is presented with another set of five words – suit, trousers, skirt, sweater, pyjamas – three of which also appeared in the first list. The task is to identify the three words from the original list. Scores falling below an established

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

threshold – typically derived from individuals with genuine cognitive impairments – are considered indicative of performing below one's true cognitive ability.

Thus, whereas SVTs focus on exaggerated symptom reports, PVTs are designed to assess suboptimal performance on cognitive tasks. Although these tendencies overlap to some extent – both reflecting distorted symptom presentation – they are distinct constructs: individuals who exaggerate symptoms do not necessarily underperform, and vice versa. That said, the cut-offs for both SVTs and PVTs in detecting symptom distortion are generally set to prioritise specificity over sensitivity (Series 2025). This approach aims to minimise false positives – misclassifying genuine symptom presentations as distorted – while accepting a potential reduction in true positives, or the accurate identification of distorted symptom presentations. Specificities >0.95 and sensitivities >0.50 are not uncommon (Dandachi-FitzGerald 2022). Below, scores on SVTs and/or PVTs that exceed commonly accepted cut-off values will be referred to as 'deviant validity test scores'.

Misconceptions regarding validity tests

Despite their established role in clinical and medico-legal contexts (Tracy 2014), validity tests are often misunderstood. Misconceptions about their function, interpretation and limitations persist in psychiatric practice. This article examines common misunderstandings regarding validity testing and clarifies the appropriate use and clinical utility of these instruments.

Misconception 1: validity tests imply a medico-legal dimension

A common misconception is that validity tests inherently imply a medico-legal dimension in a case. This misunderstanding likely stems from the fact that many validity tests were originally developed for forensic contexts and bear names that reflect this origin. Although it is true that these tests are frequently used in forensic settings – such as in evaluations of criminal defendants or litigants – they also serve an important role in clinical assessments beyond legal contexts.

Medico-legal versus clinical contexts

Validity tests often serve distinct functions in medico-legal and clinical settings, primarily differing in their temporal focus. As a rule of thumb, one could say that in medico-legal contexts, these tests help to assess whether reported symptoms may be reliably linked to past events at the centre of a legal dispute. For example, when a perpetrator claims that imperative hallucinations motivated a violent crime, validity tests can help determine whether the individual truly suffers from psychotic symptoms or whether there are grounds to doubt this. Similarly, in personal injury cases, when plaintiffs report memory impairments following an accident, these tests assess the plausibility of their reported cognitive difficulties.

In contrast, clinical assessments often apply validity tests prospectively, ensuring that reported symptoms can be relied on for accurate diagnosis and treatment planning. Here, deviant validity test results act as warning signals, highlighting potential distortions in symptom presentation. A useful analogy is found in movement artefacts on a magnetic resonance imaging (MRI) scan (McWhirter 2020): just as radiologists reject movement-corrupted MRIs, clinicians should exercise caution when interpreting clinical data based on patients' self-reports if validity measures indicate compromised accuracy. In such cases, it is worth considering whether the remaining diagnostic information sufficiently supports treatment planning, or whether additional sources – such as independent informant reports – are required.

In medico-legal contexts, deviant SVT or PVT scores may inform triers of fact about the alleged causal links between self-reported symptoms and legally relevant incidents. In contrast, clinical practice tends to view such scores as prompts for further inquiry. Regardless of the context, some authors interpret deviant SVT or PVT scores as evidence that a defendant, litigant or patient is engaging in 'non-credible' symptom reporting. As Rix & Tracy (2017) have eloquently noted, this qualifier is problematic – even in legal settings – where the assessment of an individual's credibility is the prerogative of the trier of fact. In clinical contexts, the term is nearly meaningless, as many psychiatric symptoms – such as delusional ideas or Bonnet hallucinations – are non-credible in the everyday sense of the word.

Prevalence of deviant validity test scores in clinical settings

In medico-legal assessments, the base rate of symptom over-reporting on SVTs and/or cognitive underperformance on PVTs is relatively high, even when conservative criteria are employed. For example, one study found the proportion of individuals ($n = 1300$) undergoing forensic evaluations and failing at least three SVTs and/or PVTs to be approximately 24% (Svete 2025). Darzi et al (2025) reviewed 44 studies on distorted symptom presentation in patients undergoing independent

BJPsych Advances (2026), vol. 32, 221–230 doi: 10.1192/bja.2025.10177

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

medical examinations ($n = 9794$) and estimated its prevalence at 35%. When validity tests are administered in mental healthcare, hospitals or rehabilitation centres, the prevalence of individuals who over-report and/or underperform is typically lower, but still not trivial. For example, Roor et al (2024) conducted a meta-analysis to determine the prevalence of underperformance on single PVTs in clinical contexts. Aggregating data from 47 studies encompassing nearly 6500 patients, they found a prevalence rate of 16%.

Consequences matter as much as causes

Overlooking distorted symptom presentations can create a discrepancy between psychiatric diagnoses and patients' actual problems, potentially impeding therapeutic progress. Not surprisingly, research has shown that symptom over-reporting as flagged by deviant scores on SVTs is associated with poor treatment adherence and higher healthcare utilisation (Marquardt 2024). Thus, in clinical settings the consequence of distorted symptom presentations is at least as important as their cause. Individuals exhibiting such patterns are more likely to terminate treatment prematurely, are perceived as more challenging by therapists and tend to seek help from multiple healthcare providers (Merckelbach 2025). Anticipating such challenges is one of the key advantages that validity tests offer clinicians.

Misconception 2: deviant validity test scores imply feigning

When individuals obtain deviant scores on validity tests, it signals that they present their symptoms in a distorted way. This is sometimes interpreted as feigning, faking bad, simulation or negative impression management, implying exaggeration of symptoms or impairments motivated by deceptive intent. We use the term 'intent' here in its clinical sense – the conscious fabrication of symptoms – rather than in its legal sense (i.e. the mental state accompanying a wilful criminal act). However, apart from below-chance performance (see below), validity tests cannot determine deceptive intent. Unfortunately, the belief that they can has been reinforced by the names of some of these tests.

Consider the 43-item scale derived from the MMPI, commonly known as the Fake Bad Scale (FBS). The FBS measures a patient's tendency to endorse an unusually high number of common, but also less common, symptoms such as 'my sleep is fitful and disturbed' and 'I sometimes feel that I am about to go to pieces'. Individuals who are faking bad are likely to have elevated FBS scores; however, the reverse is not necessarily true. Despite the scale's name, not everyone with an elevated FBS score is attempting to feign symptoms. For instance, individuals may respond carelessly owing to boredom, anger, frustration, lack of insight into the assessment's importance, or oppositional behaviour – none of which imply deceptive intent. In such cases, interpretive labels like 'faking bad' or 'feigning' would be inappropriate. High FBS scores can also be observed in individuals who are experiencing significant emotional distress, where specific impairments may lead to excessive symptom endorsement. In such cases, a lack of certain abilities (see below) – rather than intentional exaggeration – disrupts accurate symptom presentation (Merckelbach 2019; Edwards 2023). An elevated FBS score – or any deviant score on a validity measure, for that matter – should prompt the clinician to explore the underlying factors contributing to the result, rather than immediately interpret it as evidence of feigning. This is especially important because a deviant score, regardless of its origin, raises concerns about the overall validity of the test data, including results from other clinical scales administered to the patient.

Cry for help?

Alternative explanations for over-reporting on SVTs or underperformance on PVTs are both plausible and well documented, and should therefore be considered before attributing such behaviour to deceptive intent. These explanations may include a limited ability to articulate symptoms, as observed in conditions such as alexithymia or apathy (Dandachi-FitzGerald 2020). The presence or absence of such conditions can be further assessed through clinical observation, structured interviews and collateral information from family members.

There is little benefit in attributing deviant scores on validity tests to patients' self-reported psychopathology. Validity tests are used when clinicians have concerns about the plausibility of symptom reports, so invoking those same reports to explain away deviant test results would be circular reasoning (Merten 2013a). Likewise, interpreting deviant test results as an unconscious cry for help offers little meaningful clinical insight. It is reasonable to assume that many patients are in some way signalling a need for help – otherwise, they would not seek clinical evaluation. Moreover, unlike explanations based on factors such as boredom fuelling careless responding, indiscriminate symptom endorsement due to alexithymia, or lack of mental energy due to apathy, the notion of an unconscious cry for help is underspecified and unfalsifiable and therefore lacks clinical utility (Dandachi-FitzGerald 2024). It is true, though, that

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

patients on a waiting list may present with extreme versions of their symptoms in an attempt to be given priority. However, the potentially symptom-escalating effect of waiting lists is best seen as a contextual factor (Punton 2022).

Context

Indeed, contextual factors tend to be more informative than vague notions such as a 'cry for help'. One increasingly prominent influence is the role of social media platforms, which often disseminate caricatured or misleading information about various disorders. Such misinformation can shape how people experience symptoms and promote catastrophic over-reporting. A striking example is the surge in atypical tics among adolescents, linked to extensive use of platforms like TikTok, where sensationalised portrayals of Tourette syndrome are widespread (Olvera 2021). A recent study documented a similar effect in relation to the widespread misinformation about attention-deficit hyperactivity disorder (ADHD) circulating online, to which young people are heavily exposed (Schiros 2025). They may internalise inaccurate symptom information, which in turn may encourage help-seeking.

Another relevant contextual factor is repeated and detailed inquiry into symptoms, complaints, limitations and problems – especially when accompanied by suggestive feedback (van Helvoort 2020) – which can have an escalating effect on symptom reporting (Berthelot 2019). Arguably, the carry-over effects of such contextual factors on symptom presentation have little to do with deceptive intent.

Observation versus interpretation

The key point is that a deviant validity test score amounts to nothing more than an observation of distorted symptom expression. Whether this distortion is under the individual's conscious control, however, lies beyond the interpretive scope of validity test scores. Terms such as 'feigning', 'faking bad' and 'negative impression management' are interpretive labels with pejorative connotations, precisely because they attribute deceptive intent to the individual. Given this, such terms are often best avoided in clinical contexts, as they may undermine the potential for constructive follow-up conversations with patients (Stone 2002).

However, there are circumstances in which it is justifiable to interpret symptom over-reporting as an intentional act. For example, PVTs such as the TOMM and the coin-in-the-hand test use a forced-choice format, in which blind guessing due to severe cognitive impairment would be expected to yield roughly 50% correct responses. Performance that is significantly below that chance level can only occur through deliberate attempts to provide incorrect answers (Merten 2013b). Similarly, when an individual demonstrates marked discrepancies between deviant scores on multiple SVTs and PVTs, against the backdrop of intact and complex daily functioning, this raises concerns about intentional symptom distortion.

Misconception 3: deviant validity test scores imply malingering (or factitious disorder)

The term 'malingering' extends considerably beyond 'feigning'. While 'feigning' refers to deliberate exaggeration of symptoms, 'malingering' denotes symptom exaggeration motivated by the expectation of external gain ('secondary gain'). This constitutes a more specific interpretation that demands even stronger evidence. The names of certain tests – such as the Structured Inventory of Malingered Symptomatology and the Test of Memory Malingering – foster the misunderstanding that such evidence can be obtained from these instruments. This is not the case. These tests merely assess over-reporting of improbable symptoms (e.g. the SIMS) or underperformance on simple cognitive tasks (e.g. the TOMM).

Interpreting deviant scores on validity tests as malingering is justified only when two or more validity measures are failed, the symptom distortion is clearly under conscious control (i.e. not attributable to other conditions) and there is strong evidence of a clear motive for exaggerating symptoms or reducing effort. The literature provides well-articulated guidelines for such an interpretation (Rix 2017). Most importantly, identifying malingering requires demonstrable – rather than merely suggestive – evidence that symptom distortion is linked to identifiable external incentives. These incentives extend beyond financial or legal gain and may include, depending on the context, exemption from caregiving responsibilities, eligibility for academic accommodations (e.g. extended time on tests or assignments), approval for a residence permit on medical grounds, access to housing or exemption from military service. Crucially, the presence of such incentives cannot be inferred from SVT or PVT results; it must be corroborated by additional information within the patient's records.

A similar line of reasoning applies to factitious disorder (including Munchausen syndrome and Munchausen syndrome by proxy), which requires significantly more robust evidence than merely deviant validity test scores (Edwards 2023). Criteria related to tests, background and context have been proposed by Chafetz et al (2020). Briefly,

BJPsych Advances (2026), vol. 32, 221–230 doi: 10.1192/bja.2025.10177

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

substantiating evidence is needed to demonstrate that the individual fabricates symptoms motivated by a desire to assume the patient role. The willingness of individuals to undergo invasive medical procedures and accept treatments with harmful side-effects – a motivation not typically seen in malingering – serves as a red flag for factitious behaviour, alongside a range of other indicators.

Misconception 4: normal scores on validity tests are not informative

It is sometimes mistakenly assumed that the informational value of validity tests is asymmetrical – meaning that although a deviant score may be informative, a normal score is meaningless. This assumption overlooks the negative predictive value of validity tests. When a psychiatrist begins the diagnostic process with a suspicion that a patient may be distorting their symptom presentation, but the patient subsequently scores entirely within normal limits on SVTs and/or PVTs, there is good reason to reconsider that initial impression – especially when the tests used have high negative predictive values (Tracy 2014). In this context, negative predictive value refers to the strength with which normal test scores support the conclusion that the individual is *not* engaging in symptom overreporting and/or cognitive underperformance. The higher the negative predictive value, the more confidence clinicians can place in normal scores as reliable evidence against symptom distortion. Such a finding is far from trivial: it is not uncommon for clinicians to initially harbour doubts about symptom accuracy, only for those doubts to be refuted by normal validity test results (Dandachi-FitzGerald 2017). When patients perform within normal limits on SVTs and/or PVTs, clinicians can have strong confidence in diagnostic conclusions drawn from self-report data, which can then inform appropriate treatment recommendations.

Misconception 5: if the patient obtains deviant scores there is nothing you can do

Some clinicians hesitate to use validity tests, fearing these measures might place them in a difficult position: once it is established that a patient presents health complaints inaccurately, how can that conclusion be meaningfully addressed? This concern is misguided for two reasons.

First, and as said above, the identification of overreporting and/or underperformance should prompt a critical reassessment of all other diagnostic data. Validity test results are not isolated findings; they have direct implications for the interpretation of the entire diagnostic picture.

Second, in a clinical setting, deviant validity scores should ideally serve as a starting point for dialogue with the patient (and their significant others). The main concern should not be solely – or even primarily – about whether the individual is feigning symptoms, malingering or displaying factitious behaviour. A more relevant question is whether such scores indicate a diminished ability to accurately describe symptoms owing to factors such as alexithymia, apathy or lack of insight. This underlying issue can be further explored through collateral information. Contextual factors also warrant exploration (Berthelot 2019; Punton 2022): has prolonged time on waiting lists exacerbated health complaints? Have stereotypical portrayals on social media influenced symptom presentation? Has repeated symptom assessment inflated symptom reporting?

The literature on therapeutic assessment provides valuable guidance on involving patients in such conversations (Brown 2016). A key consideration is ensuring that the final report is clear and accessible to both the patient and other healthcare providers. This transparency helps prevent an endless cycle of repetitive diagnostic assessments and successive, yet ineffective, treatment attempts (Onofrj 2021).

The place of validity tests

Validity tests are sometimes presented as superior to the clinical interview, but such framing overlooks an important fact: administering tests – including validity tests – makes little sense in isolation. Their use presupposes that the clinician has already developed an understanding of the patient and their background. In this respect, the clinical interview is invaluable, precisely because it enables the clinician to evaluate whether the patient's symptom presentation is consistent with the medical record and with established scientific knowledge regarding the typical manifestation and progression of symptoms. Empirical research indicates that inconsistencies in symptom reporting during interviews can provide valuable insights. For example, Keesler et al (2017) interviewed individuals with mild traumatic brain injury following an accident. Those who gave inconsistent accounts of the date of the accident and/or loss of consciousness were found to display a qualitative indicator of distorted symptom presentation. These individuals also produced significantly more aberrant scores on PVTs compared with those who consistently reported the accident date and loss of consciousness.

The core idea is that identifying distorted symptom presentation requires an integration of information from both the clinical interview and validity testing (Burchett 2014). Validity tests may

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

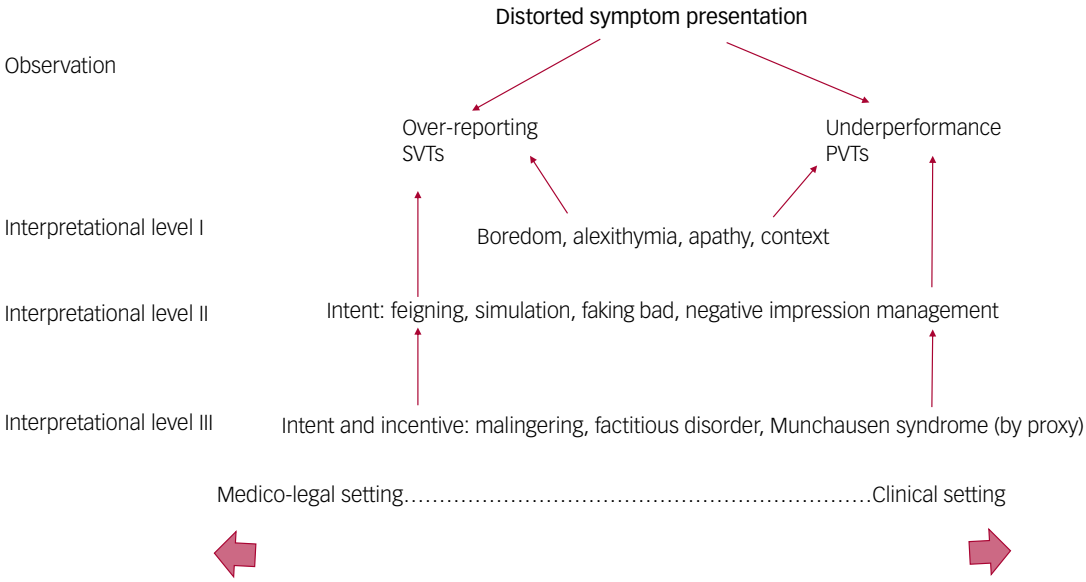


FIG 1 Three increasingly complex levels of interpreting distorted symptom expression on symptom and performance validity tests. SVTs, symptom validity tests; PVTs, performance validity tests.

provide a quantitative foundation for the qualitative indicators that emerge during the interview.

Interpreting distorted symptom presentation

Figure 1 summarises our perspective on validity tests. SVTs and PVTs aim to detect distorted symptom expression. These tests are redundant when it is obvious that the individual's symptom expression is distorted owing to acute psychosis, intoxication or severe cognitive impairment. However, in less obvious cases, SVTs and PVTs can be informative, especially when clinicians suspect that the person's symptom reports fall outside a plausible range. In such situations, these tests can either refute or support that suspicion by revealing distorted symptom expression through symptom over-reporting (SVTs) and/or cognitive underperformance (PVTs). Nevertheless, although these manifestations can be detected, their underlying causes require further interpretation. Validity tests alone do not provide this interpretation; instead, multiple sources of information at different levels are necessary.

A first level, which can be reasonably assessed through case file information and additional tests or structured interviews, concerns the ability to accurately identify and verbalise symptoms, which is often compromised in individuals with alexithymia or apathy. A similar consideration applies to contextual factors, such as exposure to social media examples of disorders, repeated testing and having been placed on a waiting list (Punton 2022).

A second and more complex level involves intentionality: is the individual deliberately attempting to exaggerate symptoms? Assessing intent is challenging, as it is not directly observable and can only be inferred from surrounding circumstances. For example, an individual may demonstrate significant underperformance on the Amsterdam Short-Term Memory Test yet actively participate in a local bridge club, indicating a major discrepancy between test performance and everyday functioning.

The third and most profound level involves attributing motives to the individual's actions. Is the person presenting distorted symptoms for secondary gain (malingering)? Or is there a pathological drive for medical attention (factitious disorder)? These considerations involve two inherently non-observable factors: intent and motive. It is only when deviant validity test scores are accompanied by a series of robust, documentable indicators that employing such terminology becomes justifiable (Rix 2017; Chafetz 2020).

Such indicators are often more thoroughly documented in medico-legal case files than in clinical records. For example, a perpetrator may claim total amnesia for a violent crime that is well documented, such as by multiple eyewitness accounts (Peters 2013). If such an individual fails multiple validity tests – performing below chance levels – and their claimed amnesia significantly deviates from known patterns of genuine memory loss, is inconsistent with their medical history and clearly serves the purpose of securing a reduced sentence, then labelling the behaviour as malingering is justified. Such a constellation may occasionally occur in clinical settings as well. For instance, consider a patient highly motivated to obtain an

BJPsych Advances (2026), vol. 32, 221–230 doi: 10.1192/bja.2025.10177

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

ADHD diagnosis who repeatedly fails validity tests, frequently requests stimulant medication and is known to resell it (Patel 2023). Nevertheless, in clinical practice, it often serves little purpose to frantically search for evidence to classify symptom distortion as malingering or factitious disorder. Instead, clinicians typically focus on understanding the potential consequences of symptom distortion to prevent overdiagnosis and overtreatment.

Clinical and legal domains may overlap

The clinical and medico-legal domains are not mutually exclusive. Social adversity, economic hardship and health problems often cluster in vulnerable populations. As a result, defendants, plaintiffs or appellants are frequently also patients, and vice versa (Pleasence 2008). Legal issues and symptom expression may interact in complex ways. For example, van der Heide et al (2020) describe a young asylum seeker referred for in-patient psychiatric assessment who presented with trauma-related and psychotic symptoms. She was treated with psychotropic medication and ultimately granted asylum on medical grounds. Collateral sources later revealed that she had fabricated her trauma history and symptoms. To appear convincing, she adhered to the treatment, which led to serious side-effects. Clinicians did not administer SVTs and/or PVTs, which might have supported a more cautious and individualised approach. In cases such as these, the priority should lie not in labelling symptom over-reporting as feigning or malingering, but in identifying and addressing it to prevent harm.

Systematic research by Van Egmond et al (2005) involving patients ($n = 99$) from a psychiatric outpatient clinic revealed that approximately 40% had expectations from treatment beyond merely improving their health. These individuals sought assistance in securing benefits, resolving debts or handling legal disputes. These authors further demonstrated that clinicians are often unaware of such expectations, despite their potential to negatively affect treatment outcomes. Validity tests can alert clinicians to the possibility of intertwined external interests and facilitate open discussions with patients regarding these factors. Importantly, the presence of such expectations is not in itself evidence that the patient is malingering (Edwards 2023).

Under-reporting symptoms

The preceding discussion may have inadvertently reinforced the notion that validity tests are concerned only with detecting exaggerated presentations of symptoms and/or cognitive impairments. This perspective overlooks the equally significant literature on the under-reporting of health complaints (Picard 2023). Several validity tests have been developed that may help clinicians identify symptom under-reporting. One example is the Supernormality Scale (Cima 2003), which includes statements such as 'I have never had mental problems' and 'Most people I live with are clearly less mentally healthy than I am'. The more an individual endorses such statements, the more likely that the person is attempting to downplay psychological difficulties.

Embedded validity scales tailored to capture under-reporting, such as the L subscale of the MMPI, aim to measure similar phenomena. Within the MMPI framework, interpretive terms like 'fake good' and 'positive impression management' are frequently applied (Picard 2023). Such terminology should be employed only when well substantiated. Appropriate contexts may include legal disputes over custody and visitation (Valerio 2017), evaluations of inmates' eligibility for parole (Toohey 2016) or pre-surgical assessments for bariatric procedures (Stivaleti Colombarolli 2023). Nevertheless, numerous factors can lead to symptom under-reporting without any intent to manipulate, including feelings of shame (Strother 2012), lack of illness insight (DiMauro 2013) or a tendency towards superlative self-presentation (De Page 2021). Again, it is generally preferable – unless there are compelling reasons to do otherwise – to use neutral terminology and avoid interpretive labels that suggest a deceptive motive. More broadly, validity tests, whether designed to identify symptom over- or under-reporting, assess the accuracy of symptom reports rather than the patient's credibility.

Concluding remarks

This article explored five common misconceptions about validity tests. We do not mean to imply that these misunderstandings arise from a lack of foresight. Over the past two decades, research on validity tests has grown significantly, along with the corresponding professional literature. Currently, about a quarter of articles in neuropsychological journals focus on developing, standardising and calibrating validity tests (Martin 2020). More than 50 SVTs and PVTs are now backed by solid psychometric documentation (Institute of Medicine 2015). Given the rapid expansion of research and literature in this field, it is entirely understandable that clinicians may find it challenging to stay up to date.

This challenge is further compounded by the fact that some of these tests have outdated names, and older article titles often conflate key concepts such as feigning, over-reporting, malingering and non-credible symptoms in an imprecise manner. Such

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

inconsistencies can create additional confusion, making it even more difficult for clinicians to navigate the evolving landscape of validity tests.

The importance of neutral terminology

Our central argument is that the interpretation of validity test outcomes should generally employ the most neutral terminology possible, avoiding terms such as feigning, malingering, factitious disorder, faking bad and faking good – unless there are well-founded reasons to use them. It is important to recognise that deviant test results alone do not provide sufficient justification for such labels. This position is not new; for example, surveys among European neuropsychologists and their American counterparts indicate that many professionals share this perspective (Martin 2024). Nonetheless, it is worthwhile to restate this point explicitly, if only because such terms are often part of the aura surrounding validity tests. That these tests have acquired such connotations is understandable, given their historical roots in forensic contexts, where these labels can more often be substantiated. In clinical settings, however, this forensic aura may inadvertently dissuade practitioners from employing validity tests altogether – a missed opportunity.

Increasing test use in psychiatric practice

We do not believe that the use of validity tests should be reserved for psychologists. As long as test users have a solid understanding of psychometrics and are aware of the scope and limitations of these tests, there is no reason why psychiatrists should not also incorporate them into their assessments. In fact, we see significant advantages in doing so, particularly for the future refinement of validity tests.

After all, these tests are based on theories about how the average patient – compared with someone presenting symptoms in an atypical way – performs on them. Nevertheless, all validity tests have a certain, albeit small, rate of false positives. Thus, patients who describe their symptoms accurately may occasionally attain deviant scores on SVTs, just as well-motivated patients may sometimes obtain poor PVT scores (Merten 2007). A systematic description of such cases could help fine-tune test items, and in this regard, the insights of practising psychiatrists are highly valuable.

Further questions

We readily acknowledge that our discussion of validity tests has been somewhat rudimentary. Several important questions remain unaddressed: is it always advisable to use these tests? If so, how many should be administered? Which validity tests are available, and is their cross-cultural validity sufficient for use with people from different cultural backgrounds? To what extent should national test committee evaluations influence test selection? And how should clinicians communicate deviant test results to patients? These issues have been thoroughly examined in consensus documents authored by American experts (Sweet 2021). It would be beneficial for European countries to organise similar consensus meetings to establish shared guidelines.

Data availability

Data availability is not applicable to this article as no new data were created or analysed in this study.

Acknowledgement

We thank Dr Thomas Merten, senior neuropsychologist and researcher, Department of Neurology, Vivantes Clinic, Berlin, Germany, for providing us with critical feedback. Any remaining errors are solely our responsibility.

Author contributions

H.M. drafted the first version of this article. B.D.F. and J.W. edited and annotated several subsequent versions.

Funding

This work received no specific grant from any funding agency, commercial or not-for-profit sectors. Open access funding provided by Maastricht University.

Declaration of interest

None.

References

Berthelot JM, Nizard J, Maugars Y (2019) The negative Hawthorne effect: explaining pain overexpression. *Joint Bone Spine*, **86**: 445–9.

Brown JD, Morey LC (2016) Therapeutic assessment in psychological triage using the PAI. *Journal of Personality Assessment*, **98**: 374–81.

Burchett D, Bagby RM (2014) Multimethod assessment of response distortion: Integrating data from interviews, collateral records, and standardized assessment tools. In *Multimethod Clinical Assessment* (eds C Hopwood, R Bornstein): 345–78. Guilford Press.

Chafetz MD, Bauer RM, Haley PS (2020) The other face of illness-deception: diagnostic criteria for factitious disorder with proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, **34**: 454–76.

Cima M, Merckelbach H, Hollnack S, et al (2003) The other side of malingering: supernormality. *The Clinical Neuropsychologist*, **17**: 235–43.

Dandachi-FitzGerald B, Merckelbach H, Ponds RW (2017) Neuropsychologists' ability to predict distorted symptom presentation. *Journal of Clinical and Experimental Neuropsychology*, **39**: 257–64.

Dandachi-FitzGerald B, Duits AA, Leentjens AF, et al (2020) Performance and symptom validity assessment in patients with apathy and cognitive

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

impairment. *Journal of the International Neuropsychological Society*, **26**: 314–21.

Dandachi-FitzGerald B, Martin PK (2022) Clinical judgment and clinically applied statistics: description, benefits, and potential dangers when relying on either one individually in clinical practice. In *Neuropsychological Validity Assessment in Non-Forensic Clinical Practice* (eds RW Schroeder, PK Martin): 107–25. Guilford Press.

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**: 527–32.

Darzi AJ, Wang L, Riva JJ, et al (2025) Prevalence of symptom exaggeration among North American independent medical evaluation examinees: a systematic review of observational studies. *PLOS One*, **20**: e0324684.

De Page L, Merckelbach H (2021) Associations between supernormality ('faking good'), narcissism and depression: an exploratory study in a clinical sample. *Clinical Psychology & Psychotherapy*, **28**: 182–8.

DiMauro J, Tolin DF, Frost RO, et al (2013) Do people with hoarding disorder under-report their symptoms? *Journal of Obsessive-Compulsive and Related Disorders*, **2**: 130–6.

Edwards MJ, Yogarajah M, Stone J (2023) Why functional neurological disorder is not feigning or malingering. *Nature Reviews Neurology*, **19**: 246–56.

Institute of Medicine (2015) *Psychological Testing in the Service of Disability Determination*. National Academies Press.

Keesler ME, McClung K, Meredith-Duliba T, et al (2017) Red flags in the clinical interview may forecast invalid neuropsychological testing. *The Clinical Neuropsychologist*, **31**: 619–31.

Marquardt CA, Ferrier-Auerbach AG, Schumacher MM, et al (2024) MMPI-2-RF validity scales add utility for predicting treatment engagement during partial psychiatric hospitalizations. *Psychological Assessment*, **36**: 124–33.

Martin PK, Schroeder RW, Olsen DH, et al (2020) A systematic review and meta-analysis of the Test of Memory Malingering in adults: two decades of deception detection. *The Clinical Neuropsychologist*, **34**: 88–119.

Martin PK, Schroeder RW, Odland AP (2024) Neuropsychological validity assessment beliefs and practices: a survey of North American neuropsychologists and validity assessment experts. *Archives of Clinical Neuropsychology*, **40**: 201–23.

McWhirter L, Ritchie CW, Stone J, et al (2020) Performance validity test failure in clinical populations – a systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*, **91**: 945–52.

Merckelbach H, Dandachi-FitzGerald B, van Helvoort D, et al (2019) When patients overreport symptoms: more than just malingering. *Current Directions in Psychological Science*, **28**: 321–6.

Merckelbach H, Dandachi-FitzGerald, B (2025) Symptom overreporting and its consequences for treatment. *Current Opinion in Psychology*, **65**: 102091.

Merten T, Bossink L, Schmand B (2007) On the limits of effort testing: symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal of Clinical and Experimental Neuropsychology*, **29**: 308–18.

Merten T, Merckelbach H (2013a) Symptom validity testing in somatoform and dissociative disorders: a critical review. *Psychological Injury and Law*, **6**: 122–37.

Merten T, Merckelbach H (2013b) Forced-choice tests as single-case experiments in the differential diagnosis of intentional symptom distortion. *Journal of Experimental Psychopathology*, **4**: 20–37.

Olvera C, Stebbins GT, Goetz CG, et al (2021) TikTok tics: a pandemic within a pandemic. *Movement Disorders Clinical Practice*, **8**: 1200–5.

Onofrj M, Digiovanni A, Ajdinaj P, et al (2021) The factitious/malingering continuum and its burden on public health costs: a review and experience in an Italian neurology setting. *Neurological Sciences*, **42**: 4073–83.

Patel GRJ (2023) Feigning ADHD: a necessary exploration of an uncomfortable topic. *Journal of the New Zealand College of Clinical Psychologists*, **33**: 61–71.

Peters MJ, van Oorsouw KI, Jelicic M, et al (2013) Let's use those tests! Evaluations of crime-related amnesia claims. *Memory*, **21**: 599–607.

Picard E, Aparcero M, Nijdam-Jones A, et al (2023) Identifying positive impression management using the MMPI-2 and the MMPI-2-RF: a meta-analysis. *The Clinical Neuropsychologist*, **37**: 545–61.

Pleasence P, Balmer NJ, Buck A (2008) The health cost of civil-law problems: further evidence of links between civil-law problems and morbidity, and the consequential use of health services. *Journal of Empirical Legal Studies*, **5**: 351–73.

Punton G, Dodd AL, McNeill A (2022) 'You're on the waiting list': an interpretive phenomenological analysis of young adults' experiences of waiting lists within mental health services in the UK. *PLOS One*, **17**: e0265542.

Rix JB, Tracy DK (2017) Malingering mental disorders: medicolegal reporting. *BJPsych Advances*, **23**: 115–22.

Roor JJ, Peters MJ, Dandachi-FitzGerald B, et al (2024) Performance validity test failure in the clinical population: a systematic review and meta-analysis of prevalence rates. *Neuropsychology Review*, **34**: 299–319.

Schiros A, Bowman N, Antshel K (2025) Misinformation mayhem: the effects of TikTok content on ADHD knowledge, stigma, and treatment-seeking intentions. *European Child & Adolescent Psychiatry* [Epub ahead of print] 5 Jun 2025. Available from: https://doi.org/10.1007/s00787-025-02769-8.

Series H, Burns A (2025) Cognitive testing and the hazards of cut-offs. *BJPsych Advances*, **31**: 20–7.

Stivaleti Colombarolli M, Giromini L, Pasian SR (2023) Self-reports do not tell the whole story: a study of candidates for bariatric surgery using a multimethod approach. *Psychological Injury and Law*, **16**: 249–63.

Stone J, Wojcik W, Durrance D, et al (2002) What should we say to patients with symptoms unexplained by disease? The 'number needed to offend'. *BMJ*, **325**: 1449–50.

Strother E, Lemberg R, Stanford SC, et al (2012) Eating disorders in men: underdiagnosed, undertreated, and misunderstood. *Eating Disorders*, **20**: 346–55.

Svete LJ, Tindell WW, McLouth CJ, et al (2025) A retrospective analysis of rates of malingering in a forensic psychiatry practice. *Journal of the American Academy of Psychiatry and the Law*, **53**: 26–34.

Sweet JJ, Heilbronnner RL, Morgan JE, et al (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**: 1053–60.

Toohey MJ, Tirnady R, McCabe J, et al (2016) Underreporting of anger in men on probation. *Journal of Offender Rehabilitation*, **55**: 39–50.

Tracy DK (2014) Evaluating malingering in cognitive and memory examinations: a guide for clinicians. *Advances in Psychiatric Treatment*, **20**: 405–12.

Valerio C, Beck CJ (2017) Testing in child custody evaluations: an overview of issues and uses. *Journal of Child Custody*, **14**: 260–80.

Van Egmond J, Kummeling I, van Balkom T (2005) Secondary gain as hidden motive for getting psychiatric treatment. *European Psychiatry*, **20**: 416–21.

van Helvoort D, Otgaar H, Merckelbach H (2020) Worsening of self-reported symptoms through suggestive feedback. *Clinical Psychological Science*, **8**: 359–65.

van der Heide D, Boskovic I, van Harten P, et al (2020) Overlooking feigning behavior may result in potential harmful treatment interventions: two case reports of undetected malingering. *Journal of Forensic Sciences*, **65**: 1371–5.

Young G, Giromini L, Erdodi L, et al (2025) Invalid response set and malingering-related assessments in psychological injury: definitions and a hierarchy of terms. *Psychological Injury and Law*, **18**: 3–18.

MCQ answers
1 d 2 c 3 b 4 e 5 a

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press

MCQs

Select the single best option for each question stem

1 The primary purpose of validity tests in clinical settings is to:
a detect malingering in patients seeking financial compensation
b replace comprehensive clinical interviews
c identify patients who are intentionally feigning symptoms
d assess whether patients can report symptoms with reasonable accuracy
e determine which patients require forensic evaluation.

2 According to recent meta-analysis, the proportion of patients in mental health-care settings who show deviant scores on validity tests is:
a less than 5%
b between 5 and 10%
c between 10 and 20%
d between 30 and 40%
e over 50%.

3 Which of the following is *not* a well-documented cause of distorted symptom presentation?
a limited capacity to articulate symptoms (alexithymia)
b a cry for help
c influence of social media and misinformation about disorders
d repeated detailed inquiry into symptoms having an escalating effect
e apathy affecting symptom reporting.

4 Patients who show patterns of over-reporting or underperformance on validity tests are more likely to:
a have higher IQ scores
b respond better to medication interventions
c have fewer comorbid conditions
d require shorter treatment durations
e terminate treatment prematurely and seek help from multiple providers.

5 When discussing validity test results with patients, the most appropriate approach is to:
a use neutral terminology and engage in collaborative dialogue
b directly confront patients about apparent inconsistencies
c avoid discussing the results altogether to prevent damaging rapport
d focus primarily on potential financial motivations
e refer immediately to a forensic specialist.

BJPsych Advances (2026), vol. 32, 221–230 doi: 10.1192/bja.2025.10177

https://doi.org/10.1192/bja.2025.10177 Published online by Cambridge University Press