Performance validity test failure in clinical populations—a systematic review

Performance validity test failure in clinical populations- a systematic review

Laura McWhirter MBChB*

Craig Ritchie PhD

Jon Stone PhD

Alan Carson MD

Centre for Clinical Brain Sciences, University of Edinburgh

* corresponding author

Word count: 4625

Abstract word count: 244

Supplementary Tables: 9

Figures: 2

References: 60

Disclosures

LM is funded by a University of Edinburgh Clinical Research Fellowship funded philanthropically by Baillie Gifford. LM provides independent medical testimony in court cases regarding patients with functional disorders. AC is a director of a limited personal services company that provides independent medical testimony in Court Cases on a range of neuropsychiatric topics on a 50% pursuer 50% defender basis, is an associate editor of the Journal of Neurology Neurosurgery and Psychiatry, and is the treasurer of the International Functional Neurological Disorder Society. JS reports personal fees from UptoDate, outside the submitted work, runs a self help website for patients with functional neurological symptoms (www.neurosymptoms.org) which is free and has no advertising, provides independent medical testimony in personal injury and negligence cases regarding patients with functional disorders, and is secretary of the International Functional Neurological Disorder Society. Professor Jon Stone is a Chief Scientists Office NHS Research Scotland Career Researcher. CR declares no conflicts of interest.

Author contributorship

LM performed the literature search, data extraction and collation, and writing. AC, JS, and CR contributed equally to the study design, and to subsequent review and revision of the manuscript.

Abstract

Objective

Performance validity tests (PVT) are widely used in attempts to quantify effort and/or detect negative response bias during neuropsychological testing. However, it can be challenging to interpret the meaning of poor PVT performance in a clinical context. Compensation-seeking populations predominate in the PVT literature. We aimed to establish base rates of PVT failure in clinical populations without known external motivation to underperform.

Methods

We searched MEDLINE, EMBASE, and PsycINFO for studies reporting PVT failure rates in adults with defined clinical diagnoses, excluding studies of active or veteran military personnel, forensic populations, or studies of participants known to be litigating or seeking disability benefits. Results were summarised by diagnostic group and implications discussed.

Results

Our review identified 69 studies, and 45 different PVTs or indices, in clinical populations with intellectual disability, degenerative brain disease, brain injury, psychiatric disorders, functional disorders, and epilepsy. Various pass/fail cut-off scores were described. PVT failure was common in all clinical groups described, with failure rates for some groups and tests exceeding 25%.

Conclusions

Performance validity test failure is common across a range of clinical conditions, even in the absence of obvious incentive to underperform. Failure rates are no higher in functional disorders than in other clinical conditions. As performance validity test failure indicates invalidity of other attempted neuropsychological tests, the finding of frequent and unexpected failure in a range of clinical conditions raises important questions about the degree of objectivity afforded to neuropsychological tests in clinical practice and research.

Background

Performance validity tests, also historically called effort tests, are used by clinical psychologists to try to detect inadequate effort and exaggerated or feigned impairment. Identifying invalid performance has critical implications for how the psychologist interprets the rest of the neuropsychological examination, and may also have clinical and medicolegal implications.

As clinicians in neuropsychiatry and neurology we often read neuropsychology reports which include reference to effort and validity measures. However, it can be difficult to interpret the significance of PVT failure in our patients, where complex combinations of neuropathological, cognitive, and emotional factors, including negative prior experiences with other health professionals, can influence symptom experience and behaviour in the consultation.

Moreover, the PVT literature is difficult to assimilate in a clinically meaningful way. This is in part due to the wide range of free-standing and embedded measures described in different studies, and in part due to the range of mixed clinical and litigating populations tested. In addition, descriptions of tests and cut-offs provided are often limited, in view of concerns about the possibilities of preparation or coaching in litigants undergoing neuropsychological assessment[1].

Previous reviews have discussed the application, meaning, and interpretation of validity tests results[2–4], have reviewed specific tests, or described PVT performance in specific groups. While some describe the proportion of examinees involved in seeking compensation, it is difficult to extract from these data a clear picture of performance in individuals who are ill and/or impaired and are not seeking compensation.

We identified a clinical need for a clear summary of the rates of PVT failure in distinct clinical groups: i.e. by diagnosis. In our view, better understanding of how people with different clinical diagnoses perform in PVTs is an important preliminary to further research to understand what single or multiple factors we might be measuring when one of our patients ‘fails’ one or more PVTs.

Aim

Our primary aim was to summarise the available published data on performance validity test failure rates in clearly defined (by diagnosis) non-litigating, non-forensic, non-military, non-military-veteran, **clinical** populations. Secondly, we aimed to consider the implications of our findings in terms of the uses of performance validity tests (PVTs) in clinical practice.

Method

Search strategy and selection criteria

We systematically searched the published peer-reviewed English language literature in MEDLINE, Embase, and PsycINFO databases from inception to July 5th 2019. The search, screening, and data extraction were done by one author (LM), and the review was conducted in line with PRISMA guidelines[5]. The search terms used were ["performance validity test*" OR "symptom validity test*" OR "effort test*"]. We included studies reporting the results of performance validity tests (not symptom validity questionnaires) in one or more individuals with a recorded clinical diagnosis of a specific medical disorder. We excluded studies of mixed clinical populations, in which performance by diagnosis was not reported. We also excluded studies of children and adolescents (<16), forensic populations, studies in which ≥ 50% of participants were known to be involved in litigation or seeking welfare benefits, studies of active military personnel or military veterans, and studies involving assessments of individuals with possible Attention Deficit Hyperactivity Disorder (ADHD) or Post-traumatic Stress Disorder (PTSD). The reason for exclusion of these groups was that they are substantially more likely to be undergoing assessment where there is a potential incentive for financial compensation or other social advantages. However, it should be noted that it is also likely that the included studies included individuals with incentives to underperform which were unknown to the investigators. Studies describing attempts to assess the validity of self-reported symptoms were excluded, as they were considered outside the scope of the paper.

Following the initial search and collation of data, additional title and keyword searches were performed on 15th January 2020, for the eight most frequently identified PVTs in the studies identified in the initial search. This search yielded an additional 11 eligible studies.

Data were extracted independently by author LM using Excel, and synthesised into tables of test failure rate by diagnosis, with the aim of examining pooled failure rates for specific disorders in the context of a narrative review.

Search results and screening

[*Figure 1 – Selection of included studies*]

45 different PVTs or indices were identified (supplementary table 1), and within these indices a range of cut-off scores were reported for many tests. The majority of results identified were for free-standing validity tests.

Many of the validity tests identified (including the three most frequently reported tests: the Word Memory Test (WMT), Test of Memory Malingering (TOMM), and Medical Symptom Validity Test (MSVT)) used a forced choice paradigm. In a forced choice PVT, the examinee is asked to recognise previously seen words, pictures or numbers mixed with unseen foils in a 1:1 ratio. If the examinee correctly recognises significantly fewer than half (<18/50 in the TOMM, on the basis of 90% confidence intervals), as would be expected if they were selecting answers at random, they are assumed to be preferentially selecting incorrect answers (intentionally or unintentionally). Of note, however, the cut-off scores for these tests were consistently much higher than the chance level, and the proportion of individuals scoring below the chance level was infrequently reported. The relevance of the use of a forced-choice paradigm was therefore unclear.

Other tests used the 'floor effect': a cut-off score which it seems improbable that any individual applying full effort will score below. Reliable Digit Span (the fourth most commonly reported test, consisting summed maximum forward and backward digit span) and the Rey 15-item test, are examples of 'floor effect' validity tests.

A small number of tests used an 'atypical pattern' principle. For example, in the dot counting test, examinees are expected to count grouped collections of dots more quickly than ungrouped dots and the absence of such a discrepancy (or reversed discrepancy) is taken as an indicator of invalid performance.

Twenty-seven studies stated either that no litigating or compensation-seeking examinees were included. In 40 studies, presence of litigation was not reported, but the population was recruited from a clinical or clinical research (rather than medicolegal) setting. In one study participants were informed that test results would not

be made available and so could not be used to support compensation claims. Finally, one study examined adults seeking to regain custody of their children, who were presumably motivated to perform well[6].

Quality of included studies

The potential for selection bias in included studies was significant. Many studies were conducted retrospectively on previously collected data. The methods for ascertaining clinical diagnoses were not always clearly described. Importantly, it should be assumed (in the absence of any evidence otherwise) that those examiners undertaking the validity tests were not blinded to clinical history and/or diagnosis of the examinees. Variable cut-off scores were used, as can be seen in the detailed discussion of results below and in the supplementary tables. Finally, despite our efforts to minimise possible influence of external incentives, the presence of undetected external (or internal) incentive cannot be excluded.

Intellectual disability (supplementary table 2)

Three studies described PVT performance in adults with intellectual disability. In Goldberg and Miller, 6/16 (38%) adults with mean IQ 63.9 failed (<9) the Rey 15-item test[7]. In the largest study included, 6 of 276 (2%) adults with intellectual deficits but full-scale IQ >70 seeking to regain custody of their children (and therefore expected to be motivated to pass) failed the Medical Symptom Validity Test (criterion A) and 11 of 223 (5%) failed the Word Memory Test[6]. In the same study, 14% (2) of 14 individuals in the same circumstances but with FSIQ ≤70 failed the Word Memory Test and 0 of 17 failed the Medical Symptom Validity Test[6].

Mild cognitive impairment (MCI) (supplementary table 3)

Nine studies reported PVT performance in mild cognitive impairment (MCI) or minor neurocognitive disorder, constructs in which measurable cognitive impairment is present which is not severe enough to merit diagnosis of dementia and which is not associated with functional impairment. The highest reported failure rates were 42% (153 of 365) individuals with amnestic MCI in Loring et al.; 36% (29) of 80 with minor neurocognitive disorder failed the Rey 15-item test (cut-off <20) in Fazio et al.; 27% (1462) of 5414 with MCI failed the logical memory test (cut-off <14) and 25% (1354) of 5414 failed semantic word generation (cut-off <13) in Daviset al, and 22% (13) of 60 individuals with 'probable MCI' in Green et al.[8–11]. Of note, 11 of the 13 MCI individuals in Green et al. 2011 who failed criterion A of the Word Memory Test did not meet criterion B (easy – hard difference

<30) and so had a possible dementia profile[11]. Pooled failure rates for Reliable Digit Span in MCI were 16% (83 of 533) at a cut-off of ≤7[8,12], and 1% (6/613) at a cut-off of ≤5[8,9,12].

Functional disorders (supplementary table 4)

Eleven studies described PVT performance in people with functional disorders, including for the purposes of this review those conditions termed 'medically unexplained', somatoform or 'nonorganic'. Where possible, PVT failure rates were pooled by specific condition. In two studies of individuals with fibromyalgia, 8% (8) of 104 failed the TOMM[13,14]. In three studies of psychogenic non-epileptic seizures (PNES, also called dissociative seizures), 10% (13) of 132 failed the TOMM[15–17]. In two other studies of PNES, 44% (25) of 57 met criterion A (therefore failed) on the standard Word Memory Test[18,19]. In two studies of individuals with chronic fatigue syndrome, 25% (374) of 1526 failed the Amsterdam Short Term Memory Test (scoring <86/100)[20,21].

Failure rates higher than 25% were reported by Tyson et al in 33 individuals with psychogenic non-epileptic seizures on Reliable Digit Span (cut-off ≤7), vocabulary – digit span (≥3), forced choice recall on the CVLT (≤15), and the Boston Naming Test[17].

Epilepsy (supplementary table 5)

Eleven studies reported PVT performance in people with epilepsy. In five studies including 246 people with epilepsy, 13% (31) failed the TOMM[15–17,22,23]. In three studies including a total of 74 people with epilepsy, 19% met criterion A of the standard version of the Word Memory Test[19,24,25]. Two studies reported Reliable Digit Span results in people with epilepsy. Maiman et al. reported a failure rate of 23% (14/63) at a ≤7 cut-off and 10% (6/63) at a ≤5 cut-off, and Tyson et al. reported a failure rate of 45% (32/72) at a ≤7 cut-off; the two studies producing a pooled RDS failure rate in epilepsy of 34% at a ≤7 cut-off.

Notably, Tyson et al. reported higher failure rates in epilepsy compared with a group with Psychogenic Non-Epileptic Seizures (see supplementary table 4) in six of eight tests included (TOMM, RDS, digit span, Boston naming test, complex ideational material, logical memory recognition trial) with failure rates higher in PNES than epilepsy only in vocabulary – digit span, and the forced choice test of CVLT. Of the two other studies comparing these groups, Cragar et al. reported higher failure rates in PNES than epilepsy (14% vs 2% on TOMM), as did

Drane et al. (48% vs 8%), but Hoskins reported similar failure rates on the standard Word Memory Test in epilepsy and PNES (31% and 29% respectively).

Acquired brain injury (supplementary table 6)

The studies included in supplementary table 6 describe PVT performance in clinical groups falling under a broad acquired brain injury definition: irreversible but non-progressive structural brain injury, including traumatic and hypoxic brain injury, stroke, and Korsakoff's syndrome.

Eight studies described PVT performance after mild Traumatic Brain Injury (TBI). Results in this group as a whole were highly variable, suggesting between-group differences. Most studies in mild TBI reported low PVT failure rates (<20%). In contrast, however, Novitski et al. reported failure rate of 52% (13/25) on RBANS digit span (cut-off <9) in 25 individuals who had sustained a mild TBI more than six months previously, and Erdodi et al. 2017 reported failure using a liberal cut-off on the TOMM in 53% of 20 adults after mTBI [22,26]. Similarly, Sherer et al. reported 25% of 118 people with mild TBI failed on criterion A of the Word Memory Test: the same failure rate (25%, or 38/150) as that reported in the severe TBI population described in the same study[27].

Grouping together moderate and severe brain injuries in what we consider a clinically relevant way (communication impairments prevent testing in those with the most severe injuries), three studies reported Word Memory Test results after moderate and severe brain injury, resulting in a pooled failure rate of 28% (63 of 228; 95% CI 22-34%)[25,27,28]. Results of other tests studied in moderate and severe brain injury were heterogeneous. Macciochi et al. in 2006 reported 0% failures on the Victoria Symptom Validity Test in 71 adults a mean 43.4 days after severe brain injury[29]. The same group in 2017 reported poor performance on the delayed recall (failure in 5/9), and consistency (4/9) components of the Medical Symptom Validity Test during the post-traumatic amnesia phase after brain injury but lower failure rates after resolution of post-traumatic Amnesia[30]. Erdodi et al. reported high failure rates on validity indices derived from the WAIS[31].

A study reporting validity test performance after stroke with initial aphasia found low failure rates on the (standard, pictorial) TOMM measures (7% (1/15 failing trial 2 and 0 failing the retention trial, but high failure rates on the Rey 15-item, RDS (<7) and reliable spatial span (60%, 73% and 40% respectively)[32].

One study described a single case of surgical removal of medial temporal lobe structures, and another described three cases of bilateral hippocampal atrophy after anoxic brain injury; none of these four individuals failed the Word Memory Test[33,34]. Oudman et al. reported that 2 of 20 individuals (10%) with Korsakoff Amnesia failed the 2nd trial of the TOMM[35].

Neurodegenerative disease (supplementary table 7)

Neurodegenerative disorders featured in 20 included studies – a greater number than any other group of conditions. The wide range of disorders, severities, tests, and test cut-off scores prevented calculation of meaningful pooled failure rates, although in general, failure rates were high (supplementary table 7, figure 2).

The Word Memory and Medical Symptom Validity Tests were most frequently described. Green et al. reported high failure rates in clinically defined 'probable, mild, and moderate' dementia on the Word Memory Test (71% of 42) and Medical Symptom Validity Test (48% of 23), but reported that all who failed met the 'dementia or severe impairment profile', a profile of results defined by the test author as typical of dementia or severe impairment rather than non-credible performance[11]. Howe et al. reported failure rates of 38% on the Medical Symptom Validity Test in 13 with mild dementia, all of whom met the 'dementia profile', and 83% (of 18) in advanced dementia of whom 15 met the 'dementia profile'[36]. 18 of 20 (90%) mild Alzheimer's dementia examinees in Merten et al's study failed the delayed recall component of the Word Memory Test, even though a cut-off (34%) significantly lower than the standard cut-off(45%) was applied[37]. Rudman et al. and Singhal et al. both reported high failure rates (73% of 22, and 100% of 10) in advanced dementia on the Medical Symptom Validity Test[38,39].

Two studies reported validity test results in individuals with Parkinson's disease undergoing testing in the workup for possible deep brain stimulation[40,41]. Here, failure rates were reasonably low – at most 5 of 47 (10%) failed the Medical Symptom Validity Test in Wodushek et al - but this 10% might also be considered a rather higher failure rate than expected in individuals without gross cognitive impairment in whom there is an incentive (in the form of access to a potentially beneficial treatment) to perform well on neuropsychological testing[40].

Psychiatric disorders (supplementary table 8)

Studies of schizophrenia, schizoaffective disorder, and other psychotic disorders generally reported relatively high failure rates on a range of validity tests. The highest failure rate reported was in 72% of 64 individuals with schizophrenia on the Word Memory Test[42]. In contrast, Schroeder et al's study of 104 individuals with a 'psychotic psychiatric disorder' reported low failure rates on a range of embedded tests, including 4% failure on RDS with a <=6 cut-off and 3% failure on finger-tapping[43]. Whearty et al's 2015 study of 60 individuals with schizophrenia or schizoaffective disorder reported that 28% failed Reliable Digit Span ≤6 and 36% failed finger-tapping[44].

Two studies examined performance validity in depression, Lee et al. reporting low failure rates(<=5%) on the Rey 15-item and dot counting tests and Rees et al. reporting no failures on the TOMM in 26 inpatients with depression[45,46].

Dandachi-Fitzgerald compared Amsterdam Short-Term Memory test performance in different psychiatric diagnoses: failure rates were 31% of 16 with personality disorders, 25% of 8 with psychotic disorders, 18% with substance abuse/dependence, 16% with ASD and 14% with ADHD[47]. Price et al. reported no failures on the TOMM in 71 individuals with methamphetamine dependence[48].

Other conditions (supplementary table 9)

Heintz et al. reported 23% of 13 individuals with Gilles de la Tourette syndrome failed the ASTM[49]. Two studies reported validity results in people with HIV – in one study 15% of 111 people with HIV (stable on antiretroviral therapy) failed trial 1 of the TOMM (note, TOMM is usually scored on trial 2 or a delayed trial); and in another 17% of 30 failed the Amsterdam Short-Term Memory test [50,51]. A study of neuropsychological performance in adults with sickle cell disease reported low failure rates on the TOMM and on RDS <=6, but 33% of 43 failed Reliable Digit Span with a <=7 cut-off[52]. In Rossetti et al. 2 of 10 deep brain stimulation candidates with essential tremor failed the Word Memory Test[41].

Comparative analysis of PVT results between groups

The heterogeneity of populations, tests, and in some cases cut-off scores used, makes comparisons difficult.

Failure rates (with confidence intervals), by study, in the most frequently reported validity tests are displayed graphically, by diagnostic heading, in **Figure 2**. Error margins are wide due to the small numbers in most studies. Allowing for this, however, it is clear that PVT failure is common in a range of clinical groups.

[*Figure 2 - failure rates in the 12 most frequently reported tests by diagnosis. Each point represents reported failure rate, in a particular test (indicated by colour), as reported by an individual study. Points are grouped along the x axis in the same test (colour) order in each plot, so as to allow visual comparison between plots. Vertical lines indicate the asymptotic 95% confidence interval for each reported failure rate*.]

Discussion

Our review suggests that failure of performance validity tests during neuropsychological assessment is not a rare phenomenon, but is common in many clinical groups. Of note, validity test failure is particularly likely in moderate and severe traumatic brain injury, and both mild and moderate-severe dementia (where the 'severe impairment' profile on the Word Memory Test often applies). Of note, whilst some individuals with functional disorders fail PVTs, failure rates are no higher than in a range of other diverse conditions, including epilepsy, and mild cognitive impairment.

Remarkably few studies in the very large validity test literature describe performance by clinical diagnosis. Even some studies which appear to do so often group together different illness or injury severities in a way that renders the data difficult to apply to clinical practice. For example, studies of validity tests in traumatic brain injury populations mixed those with mild, moderate and severe injuries, in whom vastly different cognitive and symptom profiles would be expected. These studies were excluded from our review on this basis, but it is likely that there is still a degree of heterogeneity in the included studies.

We aimed to select studies of individuals without clear external incentives to fail. It is of course possible that these factors were present in some cases, unknown to the investigators. Indeed, we would argue that a range of external motivators and internal factors influence how people behave during the majority of conscious encounters in most areas of healthcare. One possibility to explain our results, therefore, is that many patients do not apply the degree of effort that we would like them to apply, intentionally or unintentionally, for reasons that we cannot always immediately perceive or understand.

It seems much more likely, however, that PVTs, using commonly-applied cut-offs, are in fact not only measuring deficient effort but a whole range of factors, including memory impairment, apathy, fatigue, or attention deficit due to pain or other cognitive or somatic symptoms. People who have symptoms of any sort, in any condition, are liable to divert attention towards those symptoms. If attention is conceptualised as a finite resource (more accurately, attentional processes govern use of finite processing capacity), we suggest it is possible to fail almost any 'floor-level' test if there is not enough spare attention available to allocate to the task.

Many of the tests reported by included studies are based on a 'forced choice' paradigm. Scoring comfortably below the level of chance in a forced choice validity test has been used as evidence of deliberate exaggeration of impairment – intention to fail – which most would acknowledge is qualitatively different from, rather than on a spectrum with, not applying sufficient effort. In our experience there is a widely-held view that less-than-chance performance is precisely what PVTs are used to detect. However, our review demonstrates is that this is not really the case. Without exception, the cut-off scores used in PVTs are much higher than chance (defined as 50% or ideally lower, to allow for error): most test cut-offs are between 80% and 90%. We suggest that using a forced choice paradigm with cut-off scores greatly exceeding chance makes the forced choice element redundant, and that the test instead functions as a 'floor level' test, vulnerable to functional attentional deficit in people with symptoms of any sort. We feel it is important to point out that failure at accepted cut-off levels on commonly-used forced choice tests – the TOMM, the Word Memory Test, and the Medical Symptom Validity Test – does **not** demonstrate intention to fail.

Inadequate attentional focus on a PVT might sometimes result from diversion of attention in adaptation to symptoms and associated disability. In other situations, however, excessive focus on the task may be an intrinsic feature of the disorder being tested. In functional neurological disorders, clinical experience and experimental evidence show that excessive or misdirected effort interfere with normal performance. For example, patients with functional motor disorders who are unable to walk may be able to walk backwards, or to run – essentially when engaged in tasks which divert attention away from deliberate and effortful processes so that automatic movement-control processes to take over. Similarly, people with functional cognitive disorders can struggle and underperform when trying hard on cognitive tests but demonstrate intact cognition by providing effortless and detailed descriptions of memory lapses[53,54]. We wonder if individuals with functional neurological disorders might in some cases paradoxically fail PVTs because of an excessive degree of effort, where the harder they try,

the worse their performance. Hoover's sign of functional leg weakness depends on demonstrating impaired 'effort' in hip extension which returns to normal with contralateral hip flexion. Our clinical experience with patients with functional leg weakness is that the more they try the weaker their movements are.

Our experience of screening studies for this review illustrates some of the problems and difficulties that have arisen in validating performance validity tests.

The poor quality of the PTV evidence base examined here, with a lack of blinding to diagnosis and potential for selection bias, is in itself a key finding of the review.

The majority of excluded studies reported validity test from mixed groups of people with a wide range of different conditions attending for neuropsychological assessment, and did not report test results by diagnosis. The reason for this clumping is of course that the question investigators have been interested in is not 'How do people with different clinical conditions perform in PVTs?' but 'How can I identify a non-credible performance regardless of clinical condition?' Mixed groups are either compared with simulators, or split into 'credible' and 'noncredible' groups for the purposes of a known-groups design. Slick, Sherman and Iverson's criteria for 'probable malingered neurocognitive dysfunction', or similar definitions, are frequently used to define 'noncredible': a) motive to feign symptoms (litigation or seeking disability compensation), b) failure on two independent performance validity tests, and c) evidence of inconsistency between self-reported symptoms and observed behaviour[55].

Examination of these criteria quickly makes apparent some of the difficulties in establishing a 'gold standard' for invalid performance. Firstly, the presence of an external incentive, particularly in the form of seeking disability benefit, while it may increase the chance of invalid performance, also selects out a group of people who are 'ill' and have a range of other reasons to perform poorly. While this review did not include studies of primarily litigating or disability-benefit seeking populations in order to minimise the influence of major external influences on performance, we suggest that there are many reasons for people with 'external incentives' to fail PVTs other than inadequate effort or intention to fail.

The second 'malingered neurocognitive dysfunction' criterion[55], failure on two independent PVTs, relies on an assumption that those tests are indeed measuring something akin to effort. Alternatively, we suggest that failure on multiple PVTs indicates that 'something' is going on, but does not tell us that that 'something' is inadequate

effort or wilful exaggeration. The assumption that PVTs primarily measure effort is pervasive in the PVT literature and is reinforced by reporting of sensitivity and specificity metrics, with use of the term 'false positive' to describe failure in a 'credible' participant.

Finally, inconsistency between cognitive scores and level of function in activities of daily living is in our experience common in functional neurological disorders, and also in certain psychiatric disorders.

An important question is, therefore, why is it so difficult to find a 'gold standard' here? We suggest firstly that inadequate effort – 'not trying hard enough' – is highly subjective, is not a binary variable with a single dimension, and depends on a mixture of cognitive and emotional processes. Importantly, we consider that 'inadequate effort' is qualitatively different from deliberate exaggeration or intentional failure (as defined by Slick et al.[55]). And yet, by using these criteria to divide examinees into credible and non-credible groups, researchers use a definition for the latter (malingered dysfunction) to establish cut-offs for the former (inadequate effort).

Importantly, the manner in which we have described PVT failure rates does not necessarily reflect how they are used in practice by skilled clinical neuropsychologists, although where there is certainly expertise there is little consensus[56]. Published guidance documents for neuropsychologists are clear to point out limitations, including various reasons for test failure, and limited evidence in clinical populations[57,58]. Guidance documents recommend that multiple performance validity measures should be used, including both free-standing and embedded indicators, and emphasise that PVTs should be interpreted as part of the wider context of the assessment.

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[59]. One analogy is of movement artefact on an MRI scan; there are many reasons that a person might move during an MRI scan, but a single common end result: degradation of the images so that they are difficult or impossible to interpret. While PVT failure tells us that there is a problem with the image drawn by the other neuropsychological tests, it is not always possible to fully understand the reasons for that

interference. We suggest that future research in clinical groups with a range of symptom and impairment complexes is one possible route to better understanding of the factors influencing performance.

References

1 Bauer L, Mccaffrey RJ. Coverage of the Test of Memory Malingering, Victoria Symptom Validity Test, and Word Memory Test on the Internet: Is test security threatened? *Arch Clin Neuropsychol* 2006; **21**: 121–6.

2 Larrabee GJ. Assessment of malingering. In: Forensic neuropsychology: A scientific approach. New York: Oxford University Press, 2012: 116–59.

3 Bigler ED. Symptom validity testing, effort, and neuropsychological assessment. J. Int. Neuropsychol. Soc. 2012; **18**: 632–42.

4 Davis JJ, Millis SR. Examination of performance validity test failure in relation to number of tests administered. *Clin Neuropsychol* 2014; **28**: 199–214.

5 Moher D, Liberati A, Tetzlaff J, *et al.* Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Med. 2009; **6**. DOI:10.1371/journal.pmed.1000097.

6 Green P, Flaro L. Results From Three Performance Validity Tests (PVTs) in Adults With Intellectual Deficits. *Appl Neuropsychol Adult* 2015; **22**: 293–303.

7 Goldberg JO, Miller HR. Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. *J Clin Psychol* 1986; **42**: 792–5.

8 Loring DW, Goldstein FC, Chen C, *et al.* False-Positive Error Rates for Reliable Digit Span and Auditory Verbal Learning Test Performance Validity Measures in Amnestic Mild Cognitive Impairment and Early Alzheimer Disease. *Arch Clin Neuropsychol* 2016; **31**: 313–31.

9 Fazio RL, Faris AN, Yamout KZ. Use of the Rey 15-Item Test as a performance validity test in an elderly population. *Appl Neuropsychol Adult* 2019; **26**: 28–35.

10 Davis JJ. Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. *J Clin Exp Neuropsychol* 2018; **40**: 1013–21.

11 Green P, Montijo J, Brockhaus R. High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. *Appl Neuropsychol* 2011; **18**: 86–94.

12 Zenisek R, Millis SR, Banks SJ, Miller JB. Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Arch Clin Neuropsychol* 2016; **31**: 426–33.

13 Iverson GL, Le Page J, Koehler BE, Shojania K, Badii M. Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. *Clin Neuropsychol* 2007; **21**: 532–46.

14 Bar-On Kalfon T, Gal G, Shorer R, Ablin JN. Cognitive functioning in fibromyalgia: The central role of effort. *J Psychosom Res* 2016; **87**: 30–6.

15 Cragar DE, Berry DTR, Fakhoury TA, Cibula JE, Schmitt FA. Performance of Patients with Epilepsy or Psychogenic Non-Epileptic Seizures on Four Measures of Effort. *Clin Neuropsychol* 2006; **20**: 552–66.

16 Hill S, Ryan L, Kennedy C, Malamut B. The relationship between measures of declarative memory and the test of memory malingering in patients with and without temporal lobe dysfunction. *J Forensic Neuropsychol* 2003; **3**: 1–18.

17 Tyson B, Baker S, Greenacre M, *et al.* Differentiating epilepsy from psychogenic nonepileptic seizures using neuropsychological test data. *Epilepsy Behav* 2018; **87**: 39–45.

18 Drane DL, Williamson DJ, Stroup ES, *et al.* Cognitive impairment is not equal in patients with epileptic and

psychogenic nonepileptic seizures. *Epilepsia* 2006; **47**: 1879–86.

19 Hoskins LL, Binder LM, Chaytor NS, Williamson DJ, Drane DL. Comparison of oral and computerized versions of the word memory test. *Arch Clin Neuropsychol* 2010; **25**: 591–600.

20 Van Der Werf S, Prins J, Jongen P, Van Der Meer J, Bleijenberg G. Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment. A matched comparison between chronic fatigue syndrome and multiple sclerosis. *Neuropsychiatry, Neuropsychol Behav Neurol* 2000; **13**: 199–203.

21 Roor JJ, Knoop H, Dandachi-FitzGerald B, Peters MJ V, Bleijenberg G, Ponds RWHM. Feedback on underperformance in patients with chronic fatigue syndrome: The impact on subsequent neuropsychological test performance. *Appl Neuropsychol Adult* 2018; **27**: 188–96.

22 Erdodi L, Roth R. Low scores on BDAE Complex Ideational Material are associated with invalid performance in adults without aphasia. *Appl Neuropsychol Adult* 2017; **24**: 264–74.

23 Maiman M, Del Bene VA, MacAllister WS, *et al.* Reliable digit span: Does it adequately measure suboptimal effort in an adult epilepsy population? *Arch Clin Neuropsychol* 2019; **34**: 259–67.

24 Drane DL, Williamson DJ, Stroup ES, *et al.* Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. *Epilepsia* 2006; **47**: 1879–86.

25 Hampson NE, Kemp S, Coughlan AK, Moulin CJA, Bhakta BB. Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. *Appl Neuropsychol Adult* 2014; **21**: 183–94.

26 Novitski J, Steele S, Karantzoulis S, Randolph C. The Repeatable Battery for the Assessment of Neuropsychological Status Effort scale. *Arch Clin Neuropsychol* 2012; **27**: 190–5.

27 Sherer M, Davis L, Sander A, *et al.* Factors Associated with Word Memory Test Performance in Persons with Medically Documented Traumatic Brain Injury. *Clin Neuropsychol* 2015; **29**: 522–41.

28 Wu TC, Allen MD, Goodrich-Hunsaker NJ, Hopkins RO, Bigler ED. Functional neuroimaging of symptom validity testing in traumatic brain injury. *Psychol Inj Law* 2010; **3**: 50–62.

29 Macciocchi SN, Seel RT, Alderson A, Godsall R. Victoria Symptom Validity Test performance in acute severe traumatic brain injury: implications for test interpretation. *Arch Clin Neuropsychol* 2006; **21**: 395–404.

30 Macciocchi SN, Seel RT, Yi A, Small S. Medical Symptom Validity Test Performance Following Moderate-Severe Traumatic Brain Injury: Expectations Based on Orientation Log Classification. *Arch Clin Neuropsychol* 2017; **32**: 339–48.

31 Erdodi LA, Abeare CA, Lichtenstein JD, *et al.* Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. *Psychol Assess* 2017; **29**: 148–57.

32 Bodner T, Merten T, Benke T. Performance validity measures in clinical patients with aphasia. *J Clin Exp Neuropsychol* 2019; **41**: 476–83.

33 Goodrich-Hunsaker NJ, Hopkins RO. Word memory test performance in amnesic patients with hippocampal damage. *Neuropsychology* 2009; **23**: 529–34.

34 Carone DA, Green P, Drane DL. Word memory test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. *Appl Neuropsychol Adult* 2014; **21**: 155–60.

35 Oudman E, Krooshof E, van Oort R, Lloyd B, Wijnia JW, Postma A. Effects of korsakoff amnesia on performance and symptom validity testing. *Appl Neuropsychol Adult* 2019; : No-Specified.

36 Howe LLS, Anderson AM, Kaufman DAS, Sachs BC, Loring DW. Characterization of the Medical Symptom Validity

Test in evaluation of clinically referred memory disorders clinic patients. *Arch Clin Neuropsychol* 2007; **22**: 753–61.

37 Merten T, Bossink L, Schmand B. On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *J Clin Exp Neuropsychol* 2007; **29**: 308–18.

38 Rudman N, Oyebode JR, Jones CA, Bentham P. An investigation into the validity of effort tests in a working age dementia population. *Aging Ment Health* 2011; **15**: 47–57.

39 Singhal A, Green P, Shankar K, Gill D. High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. *Arch Clin Neuropsychol* 2009; **24**: 721–8.

40 Wodushek TR, Domen CH. Comparing two models of performance validity assessment in patients with parkinson's disease who are candidates for deep brain stimulation surgery. *Appl Neuropsychol Adult* 2018; : No-Specified.

41 Rossetti MA, Collins RL, York MK. Performance Validity in Deep Brain Stimulation Candidates. *Arch Clin Neuropsychol* 2018; **33**: 508–14.

42 Gorissen M, Sanz JC, Schmand B. Effort and cognition in schizophrenia patients. *Schizophr Res* 2005; **78**: 199–208.

43 Schroeder RW, Marshall PS. Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. *Clin Neuropsychol* 2011; **25**: 437–53.

44 Whearty K, Allen C, Lee B, Strauss G. The evaluation of insufficient cognitive effort in schizophrenia in light of low IQ scores. *J Psychiatr Res* 2015; **68**: 397–404.

45 Lee A, Boone K, Lesser I, Wohl M, Wilkins S, Parks C. Performance of older depressed patients on two cognitive malingering tests: False positive rates for the Rey 15-item memorization and Dot Counting Tests. *Clin Neuropsychol* 2000; **14**: 303–8.

46 Rees LM, Tombaugh TN, Boulay L. Depression and the Test of Memory Malingering. *Arch Clin Neuropsychol* 2001; **16**: 501–6.

47 Dandachi-FitzGerald B, Ponds RWHM, Peters MJ V, Merckelbach H. Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *Clin Neuropsychol* 2011; **25**: 812–28.

48 Price K, Desantis S, Simpson A, *et al.* The impact of clinical and demographic variables on cognitive performance in methamphetamine-dependent individuals in rural South Carolina. *Am J Addict* 2011; **20**: 447–55.

49 Heintz CEJ, van Tricht MJ, van der Salm SMA, *et al.* Neuropsychological profile of psychogenic jerky movement disorders: importance of evaluating non-credible cognitive performance and psychopathology. *J Neurol Neurosurg Psychiatry* 2013; **84**: 862–7.

50 Janssen MAM, Bertens D, Kessels L, Kessels RPC, Koopmans PP. A case-control pilot study on cognitive functioning, symptom validity and psychological wellbeing in HIV-1-infected patients in the Netherlands. *Int J STD AIDS* 2013; **24**: 387–91.

51 Paul R, Rhee G, Baker LM, Vaida F, Cooley SA, Ances BM. Effort and neuropsychological performance in HIV-infected individuals on stable combination antiretroviral therapy. *J Neurovirol* 2017; **23**: 725–33.

52 Dorociak K, Schulze E, Piper L, Molokie R, JK J. Performance validity testing in a clinical sample of adults with sickle cell disease. *Clin Neuropsychol* 2018; **32**: 81–97.

53 Jones D, Drew P, Elsey C, *et al.* Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders. *Aging Ment Health* 2016; **20**: 500–9.

54 McWhirter L, Ritchie C, Stone J, Carson A. Functional cognitive disorders: a systematic review. *The Lancet Psychiatry* 2020; **7**: 191–207.

55 Slick DJ, Sherman EMS, Iverson GL. Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *Clin Neuropsychol* 1999; **13**: 545–61.

56 Dandachi-FitzGerald B, Ponds RWHM, Merten T. Symptom Validity and Neuropsychological Assessment: A Survey of Practices and Beliefs of Neuropsychologists in Six European Countries. *Arch Clin Neuropsychol* 2013; **28**: 771–83.

57 British Psychological Society. Assessment of Effort in Clinical Testing of Cognitive Functioning For Adults. Leicester, 2009.

58 Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, Conference Participants1 C. American Academy of Clinical Neuropsychology Consensus Conference Statement on the Neuropsychological Assessment of Effort, Response Bias, and Malingering. *Clin Neuropsychol* 2009; **23**: 1093–129.

59 Green P. The Pervasive Influence of Effort on Neuropsychological Tests. Phys. Med. Rehabil. Clin. N. Am. 2007; **18**: 43–68.

Supplementary tables

Table 1 - PVTs / performance validity tests in included studies

Test name (acronym)	Free-standing / embedded	Type of test	N studies reporting test
Word memory test (WMT)	Free-standing	Forced choice	18
Test of memory malingering (TOMM)	Free-standing	Forced choice	16
Medical Symptom Validity Test (MSVT)	Free-standing	Forced choice	11
Reliable Digit Span (RDS)	Free-standing or embedded (in WAIS)	Floor effect	10
Amsterdam Short Term Memory Test (ASTM)	Free-standing	Forced choice	5
Victoria Symptom Validity Test (VSVT)	Free-standing	Forced choice	5
Rey 15-item Test	Free-standing	Floor effect	4
RBANS Effort Index	Embedded	Floor effect	4
Coin-in-the-hand Test	Free-standing	Forced choice	3
Dot counting	Free-standing	Atypical pattern	3
Finger tapping	Free-standing	Floor effect	3
Vocabulary - digit span	Embedded (WAIS)	Atypical pattern	3
California Verbal Learning Test II forced choice	Embedded (CVLT)	Forced choice	2
Digit Symbol Coding	Embedded (WAIS)	Floor effect	2
Rey Word Recognition Test	Free-standing	Forced choice OR Atypical pattern (with RAVLT recall)	2
Visual Association Test-Extended	Free-standing	Forced choice	2
Logical Memory	Embedded(WMS)	Floor effect	2
Mental Control test	Embedded (WAIS)	Floor effect	2
Autobiographical Memory Inventory	Free-standing	Floor effect	2
Digit span	Embedded (WAIS)	Floor effect	2
Rey-Osterrieth Complex Figure Test equation: copy score + [(true positive recognition – atypical recognition errors) x 3	Embedded (ROCFT)	Atypical pattern + floor effect	2
Hiscock Digit Memory Test / Hiscock forced choice test	Free-standing	Forced choice	2
Validity Indicator Profile (VIP) verbal, Symbol Search, Portland Digit Recognition Test, b-Test, Rarely missed index, Sentence repetition, Rey Auditory Verbal Learning Test equation, Camden memory test for faces, Camden Pictorial Recognition Memory Test, Wechsler Adult Intelligence Scale (WAIS) processing speed index, Digit Memory Test (DMT), Semantic word generation raw score, Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) Effort Scale, Short Test of Mental Status (STMS), Rey Complex Figure Test (RCFT), Letter Memory Test (LMT), Trail Making Test B:A ratio, reading subtest of Wide Range Achievement Test, fourth edition (WRAT-4), elements of the Auditory Verbal Learning Test (AVLT), Reliable spatial span, Coding age-corrected scaled score, Wechsler Adult Intelligence Scale (WAIS) effort index, Warrington Words			1 each

Tables (2-9) summarising reported failure rates (percentages) by diagnosis (percentages highlighted in red indicate > 25% failure rate)

Table 2 - Intellectual disability (percentages ≥25% highlighted in red)

Intellectual disability				
Study*	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Goldberg and Miller 1986	"intellectually deficient individuals": IQ 40-69 (mean 63.9)	Rey 15-item test (< 9)	16	38%
Hoskins et al. 2010	learning disability	WMT (criterion A)	5	20%
		WMT oral (criterion A)	6	0%
Green and Flaro 2015	adults with intellectual deficits (full-scale IQ (FSIQ) **≤70**) seeking to regain custody of their children	WMT (criterion A)	14	14%
		MSVT (criterion A)	17	0%
		NV-MSVT (criterion A)	4	0%
	*adults with intellectual deficits (FSIQ **>70**) seeking to regain custody of their children*	*WMT (criterion A)*	*223*	*5%*
		MSVT (criterion A)	*276*	*2%*

* References for all included studies are available in the supplementary file 'List of included studies'

Table 3 - Mild cognitive impairment (MCI) (percentages ≥25% highlighted in red

Mild cognitive impairment (MCI)				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Howe et al. 2007	MCI	MSVT (criterion A)	16	13%
Duff et al. 2011	amnestic MCI	RBANS Effort Index (>3)	72	0%
Green et al. 2011	possible MCI	WMT (criterion A)	60	22%*
Walter et al. 2014	MCI	TOMM trial 2 (≤45)	31	10%
Loring et al. 2016	amnestic MCI	RDS (≤5)	365	1%
		RDS (≤7)		14%
		AVLT recognition (≤9/15)		42%
Zenisek et al. 2016	MCI	RDS (≤5)	168	1%
		RDS (≤6)		5%
		RDS (≤7)		19%
Meyer et al. 2017	MCI	VAT-E (Visual Association Test-Extended) IR (≤21)	76	0%
		VAT-E DR (≤20)		1%
		VAT-E CNS (≤21)		4%
		VAT-E FR-MC - (≥7 - ≤9)		7%
Davis 2018	MCI	Digit Symbol Coding AASS (<6)	5414	3%
		Digit Span AASS (<6)		4%
		Logical memory (<14)		27%
		Semantic word generation (<13)		25%
		Trail Making Test B:A ratio (<1.5)		3%
Fazio et al. 2019	Minor neurocognitive disorder	Rey 15-Item Test (recall <20)	80	36%
		RDS (≤5)		0%
*11/13 'possible dementia profile' – profile of results suggestive of failure due to dementia rather than invalid performance				

Table 4 - Functional and somatoform disorders (percentages ≥25% highlighted in red)

Functional and somatoform disorders				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail**
Bar-On Kalfon et al. 2016	fibromyalgia	TOMM (≤45, assume on trial 2 or retention)	50	16%
Cragar et al. 2006	Psychogenic non-epileptic seizures ()	LMT (<93%)	21	23%
		DMT (<90%)		5%
		PDRT-27 (<54%)		14%
		TOMM trial 2 (≤45)		14%
		TOMM retention (≤45)		14%
	both epilepsy and psychogenic non-epileptic seizures (PNES)	LMT (<93%)	18	5%
		DMT (<90%)		5%
		PDRT-27 (<54%)		0%
		TOMM trial 2 (≤45)		0%
		TOMM retention (≤45)		5%
Drane et al. 2006	Psychogenic non-epileptic seizures	WMT (criterion A)	43	48%
Heintz et al. 2013	Psychogenic movement disorder with jerk-like movements	ASTM (≤85)	26	24%
Hill et al. 2003	Psychogenic non-epileptic seizures	TOMM (≤45 trial 2 or retention trial)	57	11%
Hoskins et al. 2010	Psychogenic non-epileptic seizures	WMT oral (criterion A)	16	44%
		WMT (criterion A)	14	29%
Iverson et al. 2007	Fibromyalgia	TOMM trial 1 (not stated)	54	0%
		TOMM trial 2 (not stated)		0%
		TOMM retention (not stated)		0%
Kemp et al. 2008	patients with medically unexplained symptoms (20 psychogenic non-epileptic seizures, 14 functional movement disorder/paralysis, 4 nonorganic sensory deficit, 2 functional blindness, 1 fibromyalgia, 1 nonorganic neuropsychological complaints)	MSVT IR (≤85)	43	12%
		MSVT DR (≤85)		12%
		Coin-in-hand test (≤7/10)		9%
		Autobiographical Memory Index (≤9)		5%
		Camden Pictorial Recognition Memory Test (<5th age-related centile using upper limit sample)		19%
		Mental Control Test (<5th age-related centile using upper limit sample)		16%
Van der Werf et al. 2000	Chronic fatigue syndrome	ASTM (<86)	144	29%
Roor et al. 2018	Chronic fatigue syndrome	ASTM (≤85)	1382	24%
Tyson et al. 2018	Psychogenic non-epileptic seizures	TOMM (trial 1 ≤39 or trial 2 ≤44)	33	13%
		RDS (≤7)		27%
		Digit span age-corrected scaled score (≤6)		22%
		vocabulary – digit span (≥3)		26%
		Forced choice recall test of CVLT (≤15)		32%

		FAS and animals verbal fluency (≤33)		24%
		Boston Naming Test (≤37)		25%
		Complex Ideational Material (≤29)		10%
		Logical Memory Recognition trial (≤20)		13%

Table 5 – Epilepsy (percentages ≥25% highlighted in red)

Epilepsy				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Cragar et al. 2006	epilepsy	LMT (<93%)	41	17%
		DMT (<90%)		5%
		PDRT-27 (<54%)		2%
		TOMM trial 2 (≤45)		2%
		TOMM retention (≤45)		2%
	both epilepsy and psychogenic non-epileptic seizures	LMT (<93%)	18	5%
		DMT (<90%)		5%
		PDRT-27 (<54%)		0%
		TOMM trial 2 (≤45)		0%
		TOMM retention (≤45)		5%
Drane et al. 2006	epilepsy	WMT criterion A	41	8%
Grote et al. 2000	epilepsy	VSVT(<16/24 difficult correct)	30	7%
Erdodi et al. 2017 (2)	epilepsy	TOMM trial 2 (≤48)	22	9%
Hampson et al. 2014	epilepsy	WMT-IR	16	6%
		WMT-DR		13%
		WMT-CR		38%
		WMT criterion A		38%
		Coin-in-hand test (ns)		6%
		Autobiographical memory index (ns)		0%
		digit-symbol coding (not stated)		25%
		Camden memory test for faces (ns)		6%
		Mental Control Test (ns)	15	27%
Hill et al. 2003	epilepsy (temporal lobe)	TOMM (≤45)	48	4%
Hoskins et al. 2010	epilepsy	WMT oral (criterion A)	14	14%
		WMT (criterion A)	17	31%
Keary et al. 2013	medically intractable focal epilepsy	VSVT (<18/24 hard items)	404	5%
Loring et al. 2005	epilepsy	VSVT (<18/24 hard items)	120	12%
Tyson et al. 2018	epilepsy	TOMM (<45)	72	35%
		RDS (≤7)		45%
		Digit span age-corrected scaled score (≤6)		45%
		vocabulary – digit span (≥3)		21%
		Forced choice recall test of CVLT (≤15)		12%
		FAS and animals verbal fluency (≤33)		51%
		Boston Naming Test (≤37)		68%

		Complex Ideational Material (≤29)		31%
		Logical Memory Recognition trial (≤20)		18%
Maiman et al. 2019	epilepsy or suspected seizures	RDS (≤6)	63	15%
		RDS (≤7)		23%
		RDS (≤5)		10%
		TOMM trial 1 (≤45)		35%
		TOMM trial 2 (≤45)		2%

Table 6 - Acquired brain injury (percentages ≥25% highlighted in red)

Acquired brain injury				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Rees et al. 1998	mild traumatic brain injury	TOMM (<45 trial 2)	10	0%
Allen et al. 2011	mild traumatic brain injury	WMT (criterion A)	1	0%
Erdodi et al. 2017	mild traumatic brain injury	WAIS processing speed index (≤68)	52	0%
		Coding age-corrected scaled score (≤4)		6%
		Symbol Search age-corrected scaled score (≤4)		2%
		WAIS EI 5 (Digit span, CVLT-II, WMS-IV Logical memory, letter and animal fluency) (≥5)		18%
		WAIS EI 5 (FCR) (≥4)		13%
		WAIS EI 5 (PSP) (≥4)		18%
	moderate-severe traumatic brain injury	WAIS processing speed index (≤68)	10	30%
		Coding age-corrected scaled score (≤4)		30%
		Symbol Search age-corrected scaled score (≤4)		20%
		WAIS EI 5 (Digit span, CVLT-II, WMS-IV Logical memory, letter and animal fluency) (≥2)		44%
		WAIS EI 5 (FCR) (≥4)		40%
		WAIS EI 5 (PSP) (≥4)		25%
Erdodi et al. 2017 (2)	mild traumatic brain injury	TOMM (≤48 trial 2 or retention)	20	53%
Hoskins et al. 2010	mild head trauma	WMT oral (criterion A)	10	50%
		WMT (criterion A)	11	27%
Macciocchi et al. 2006	acute severe traumatic brain injury (mean 43.4 days post injury)	VSVT combined scores (<30 invalid)	71	0%
Macciocchi et al. 2017	moderate-severe traumatic brain injuryin post-traumatic amnesia (orientation log 20-24)	MSVT IR (≤85)	9	11%
		MSVT DR (≤85)		55%
		MSVT CNS (≤85)		44%
	moderate-severe traumatic brain injurynot in post-traumatic amnesia (orientation log 25-29)	MSVT IR (≤85)	51	6%
		MSVT DR (≤85)		10%
		MSVT CNS (≤85)		26%
	moderate-severe traumatic brain injuryunimpaired on orientation log (30/30)	MSVT IR (≤85)	17	0%
		MSVT DR (≤85)		0%
		MSVT CNS (≤85)		12%
Novitski et al. 2012	mild traumatic brain injury, > 6/12 post injury	RBANS digit span (<9)	25	52%
Sherer et al. 2015	mild traumatic brain injury (GCS 13-15)	WMT (criterion A)	118	25%
	moderate traumatic brain injury (9-12)	WMT (criterion A)	47	28%
	severe traumatic brain injury (GCS 3-8)	WMT (criterion A)	150	25%
Wu et al. 2010	severe traumatic brain injury (GCS 3-8)	WMT (criterion A)	2	0%

Hampson et al. 2014	brain injury (acute moderate-severe (post-traumatic amnesia >24h, GCS <12/15))	WMT-IR	11	27%
		WMT (criterion A)	10	30%
	brain injury (in community residential care, moderate / severe (post-traumatic amnesia >24h, GCS <12/15))	WMT (criterion A)	19	45%
		coin-in-hand test (ns)	20	5%
		autobiographical memory index (ns)	18	78%
		digit-symbol coding (ns)	17	20%
		mental control (ns)	19	26%
		Camden memory test for faces (<5th age-related percentile for oldest normative age group)	18	28%
Terry et al. 2015	former high school footballers with >2 concussions >15 years prior	MSVT (criterion A)	25	0%
Bodner et al. 2019	acute stroke with first manifestation of aphasia (mild to severe)	TOMM 2nd trial (≤45)	15	7%
		TOMM retention trial (≤45)		0%
		Rey 15-item test pass/fail) (<8)		60%
		RDS (<7)		73%
		Reliable spatial span (<7)		40%
Oudman et al. 2019	Korsakoff amnesia	TOMM 2nd trial (not stated)	20	10%
		VAT-E IR (not stated)		5%
		VAT-E DR (not stated)		5%
		VAT-E CNS (not stated)		0%
Goodrich-Hunsaker and Hopkins 2009	bilateral hippocampal atrophy secondary to anoxic brain injury	WMT (criterion A)	3	0%
Carone et al. 2014	surgical removal of left anterior hippocampus and parahippocampal gyrus	WMT (criterion A)	1	0%

Table 7 - Degenerative brain disease (percentages ≥25% highlighted in red)

Degenerative brain disease				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Teichner et al. 2004	dementia	TOMM (<45 trial 2 or retention trial)	21	76%
Carone et al. 2014	non-specific progressive dementia	MSVT (criterion A)	1	100%*
		WMT (criterion A)		100%*
		RDS (not stated: assume ≤7)		0%
Davis 2018	dementia	Digit Symbol Coding (age-adjusted scaled score) (<6)	5761	16%
		Digit Span (age-adjusted scaled score) (<6)		11%
		Logical memory (<14)		68%
		Semantic word generation raw score (<13)		60%
		Trail Making Test B:A ratio (<1.5)		2%
Dean et al. 2009	dementia	Digit Span (age-adjusted scaled score) (≤5)	172	27%
		RDS pass/fail (≤6)		30%
		Three digits timed (>2s)	50	18%
		Four digits timed (>4s)	48	10%
		Vocabulary - digit span (>5)	149	3%
		Dot counting (escore <17)	80	50%
		TOMM trial 2 (≤45)	20	55%
		Warrington words (<33)	39	41%
		Rey 15-item test free recall (<9)	105	74%
		Rey 15-item test recognition equation (<20)	50	86%
		Logical memory RMI (≤136)	43	23%
		Finger tapping (men ≤35, women ≤28)	55	31%
		b-Test (≥160)	34	53%
		Rey word recognition (men ≤5, women ≤7)	32	22%
		Rey word recognition equation (≤9)	32	44%
		RAVLT equation (≤12)	64	87%
		Rey-Osterreith equation (≤47)	51	63%
Duff et al. 2011	probable Alzheimer's Disease	RBANS Effort Index (>3)	126	33%
Fazio et al. 2019	dementia (major neurocognitive disorder)	Rey 15-Item (<20 on recall & recognition)	52	90%
		RDS pass/fail (≤5)		9%
Green et al. 2011	dementia (probable, mild, and moderate: CDR 0.5 - 2)	WMT (criterion A)	42	71%*
		MSVT (criterion A)	23	48%*
Howe et al. 2007	dementia (early)	MSVT (criterion A)	13	38%*
	dementia (advanced)	MSVT (criterion A)	18	83%**
Loring et al. 2016	early Alzheimer's dementia (MMSE 20-26,+NINCDS/ARDRA criteria probable)	RDS (≤5)	176	3%
		RDS (≤7)		34%
		AVLT recognition		70%
Merten et al. 2007	mild Alzheimer's dementia (mean MMSE score 22.2, SD 2.9)	ASTM (<85)	20	90%
		WMT IR (<34)		90%

		WMT DR (<34)		90%
		WMT consistency (<34)		95%
		TOMM 2nd trial (<45)		30%
		TOMM delay trial (<45)		50%
Meyer et al. 2017	mild Alzheimer's dementia	VAT-E IR (≤20)	26	0%
		VAT-E DR (≤19)		0%
		VAT-E CNS (≤19)		4%
Rudman et al. 2011	mild dementia diagnosed before 65 (CAMCOG)	coin in hand (ns)	20	0%
		dot counting time (grouped > ungrouped)		0%
		dot counting errors (ns)		10%
		Rey 15-item test (ns)		15%
		TOMM (ns)		5%
		NV-MSVT (ns)		50%
		MSVT (ns)		35%
	moderate/severe dementia diagnosed before 45 (CAMCOG)	coin in hand (ns)	22	23%
		dot counting time (grouped > ungrouped)		0%
		dot counting errors (ns)		32%
		Rey 15-item test		73%
		TOMM (ns)		64%
		NV-MSVT (ns)		77%
		MSVT (ns)		73%
Sieck et al. 2013	Huntington Disease	RBANS EI (>3)	121	18%
		RBANS ES (only the 43 scoring <19 list recognition and <9 digit span) (<12)	43	70%
		TOMM (<45 on trial 2)	36	8%
Singhal et al. 2009	advanced dementia (6 AD, 4 undetermined)	MSVT (criterion A)	10	100%*
		NV-MSVT (criterion A)		100%*
Walter et al.	moderate-severe dementia	TOMM trial 2 (≤45)	28	21%
Wodushek et al.	Parkinson's disease candidates for DBS	MSVT (criterion A)	47	10%***
		MSVT (criterion A)		6%
		RDS (≤6)		5%
		vocabulary – digit span (scaled score) (>5)		4%
		CVLT-II forced choice (<14)		0%
Zenisek et al. 2016	Alzheimer's dementia	RDS (≤7)	133	39%
		RDS (≤6)		20%
		RDS (≤5)		8%
	Vascular dementia	RDS (≤7)	8	63%
		RDS (≤6)		25%
		RDS (≤5)		0%
	Dementia with Lewy Bodies	RDS (≤7)	27	37%
		RDS (≤6)		15%
		RDS (≤5)		0%
	Frontotemporal dementia	RDS (≤7)	15	53%
		RDS (≤6)		27%
		RDS (≤5)		13%
	Parkinsonian syndromes	RDS (≤7)	20	35%
		RDS (≤6)		20%
		RDS (≤5)		5%

Rossetti et al. 2018	Parkinson's disease – deep brain stimulation surgical candidates	WMT (criterion A)	20	5%
Woods et al. 2003	HIV-associated neurocognitive disorders	Hiscock Digit Memory Test (<90%)	82	2%
Van der Werf et al. 2000	Multiple sclerosis	ASTM (<86)	40	13%
* all who failed had a dementia / severe impairment profile (profile of results suggestive of failure due to dementia rather than invalid performance) **13/15 who failed had dementia / severe impairment profile *** examinees with dementia / severe impairment profile excluded				

Table 8 - Psychiatric disorders (percentages ≥25% highlighted in red)

Psychiatric disorders				
Study	**Clinical definition**	**Test (cut-off)**	**N**	**% to fail test**
Back et al. 1996	schizophrenia	Rey 15-item test (<9)	30	13%
		Rey dot-counting (mean grouped-dot counting time > 4.8x AND grouped time:ungrouped time ≤ 2:1		13%
		Hiscock Forced Choice, 18-trial version (<90%)		27%
Gorissen et al. 2005	schizophrenia	WMT (criterion A)	64	72%
Moore et al. 2013	schizophrenia or schizoaffective disorder	RBANS EI (> 3)	128	23%
Hunt et al. 2014	schizophrenia (63%) or schizoaffective disorder (37%)	Validity Indicator Profile (VIP) verbal (ns)	53	60%
		VIP non-verbal (ns)	54	83%
		TOMM trial 2		28%
		TOMM retention		17%
		STMS (short test of mental status) (≤29)		35%
		reading subtest of WRAT-4 (≤79)		22%
Stevens et al. 2014	schizophrenia	WMT (criterion A)	70	26%
Strauss et al. 2015	schizophrenia or schizoaffective disorder	VSVT	97	1%
		WMT (criterion A)	46	15%
Morra et al. 2015	schizophrenia (289), schizoaffective disorder (32) or another psychotic disorder (9)	RBANS Effort Index (>3)	330	9%
Whearty et al. 2015	schizophrenia (47) or schizoaffective disorder (13)	RDS (≤6)	60	28%
		Finger tapping (≤35 male, ≤28 female)		36%
Schroeder et al. 2011	psychotic psychiatric disorder	sentence repetition (≤10)	104	2%
		RDS (≤ 7)		17%
		RDS (≤ 6)		4%
		CVLT-II forced choice (≤14)		8%
		rarely missed index (≤ 136)		10%
		finger tapping (≤ 35 males, ≤ 28 females)		3%
		dot counting (≥ 20)		3%
		dot counting (≥ 17)		3%
		RCTF (≤3 true positive or > 4 false positive)		4%
Dandachi-Fitzgerald et al. 2011	personality disorders	ASTM	16	31%
	mood and anxiety disorders	ASTM	34	24%
	Autism spectrum disorder	ASTM	25	16%
	substance abuse/dependence	ASTM	11	18%
	Attention deficit hyperactivity disorder	ASTM (<85)	56	14%
	psychotic disorder	ASTM	8	25%
Ruocco 2016	borderline personality disorder	VSVT hard items (≤ 15/24)	50	2%
Lee et al. 2000	major depressive disorder (middle aged or elderly)	Rey 15-item test (<9 OR spatial score < 9)	64	5%
		Rey dot-counting (mean grouped counting time ≥ mean ungrouped dot counting time OR > 3 errors OR ungrouped time > 180s OR grouped time > 130s		0%

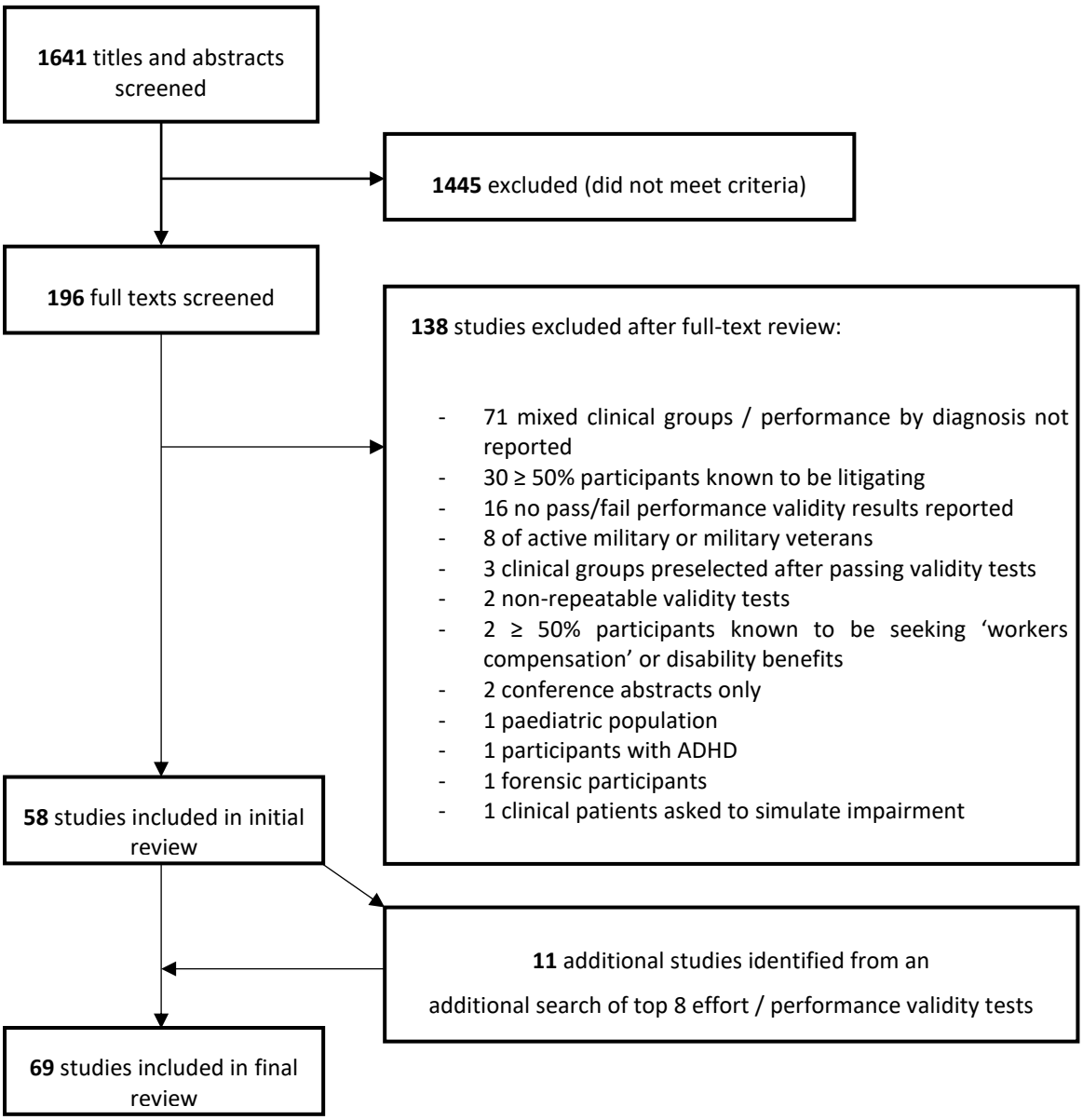
Rees et al. 2001	depression (psychiatric inpatients)	TOMM (<45 trial 2 or retention trial)	26	0%
Price et al. 2011	methamphetamine dependence	TOMM ('published cut-off score')	71	0%

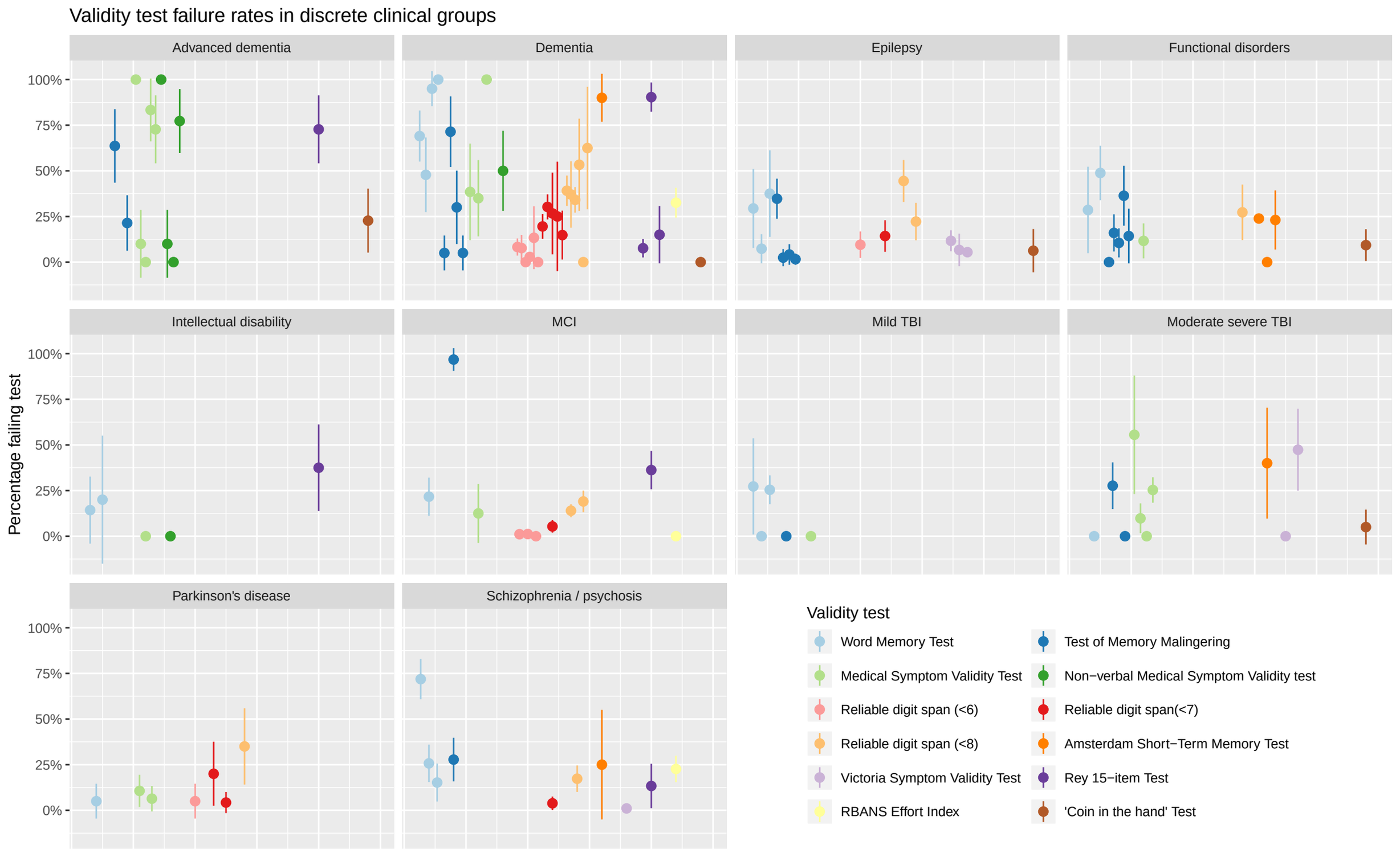
Table 9 - Other conditions (percentages ≥25% highlighted in red)

Other conditions

Study	Clinical definition	Test (cut-off)	N	% to fail test
Heintz et al. 2013	Gilles de la Tourette syndrome	ASTM (<85)	13	23%
Janssen et al. 2013	HIV-1 infected patients	ASTM (<85)	30	17%
Paul et al. 2017	HIV-infected individuals on stable combination antiretroviral therapy	TOMM trial 1 (<45)	111	15%
Rossetti et al. 2018	Essential tremor – deep brain stimulation surgical candidates	WMT criterion A	10	20%*
Dorociak et al. 2018	Sickle cell disease	TOMM trial 1 (< 40)	54	4%
		TOMM trial 2 (< 45)	43	2%
		RDS (≤ 6)		9%
		RDS (≤ 7)		33%
*1/2 of those who failed had Mild Cognitive Impairment				

Figure 1 – Selection of included studies




Validity test failure rates in discrete clinical groups
Advanced dementia
Dementia
Epilepsy
Functional disorders
Intellectual disability
MCI
Mild TBI
Moderate severe TBI
Parkinson's disease
Schizophrenia / psychosis
Percentage failing test
100%
75%
50%
25%
0%
Validity test
Word Memory Test
Test of Memory Malingering
Medical Symptom Validity Test
Non−verbal Medical Symptom Validity test
Reliable digit span (<6)
Reliable digit span(<7)
Reliable digit span (<8)
Amsterdam Short−Term Memory Test
Victoria Symptom Validity Test
Rey 15−item Test
RBANS Effort Index
'Coin in the hand' Test

List of included studies

Allen MD, Wu TC, Bigler ED. Traumatic brain injury alters word memory test performance by slowing response time and increasing cortical activation: An fMRI study of a symptom validity test. Psychol Inj Law 2011; 4: 140–6.

Back C, Boone KB, Edwards C, Parks C, Al E. The performance of schizophrenics on three cognitive tests of malingering, Rey 15-Item Memory Test, Rey Dot Counting, and Hiscock Forced-Choice Method. Assessment 1996; 3: 449–57.

Bar-On Kalfon T, Gal G, Shorer R, Ablin JN. Cognitive functioning in fibromyalgia: The central role of effort. J Psychosom Res 2016; 87: 30–6.

Bodner T, Merten T, Benke T. Performance validity measures in clinical patients with aphasia. J Clin Exp Neuropsychol 2019; 41: 476–83.

Carone DA, Ben-Porath YS. Dementia does not preclude very reliable responding on the MMPI-2 RF: a case report. Clin Neuropsychol 2014; 28: 1019–29.

Carone DA, Green P, Drane DL. Word memory test profiles in two cases with surgical removal of the left anterior hippocampus and parahippocampal gyrus. Appl Neuropsychol Adult 2014; 21: 155–60.

Cragar DE, Berry DTR, Fakhoury TA, Cibula JE, Schmitt FA. Performance of Patients with Epilepsy or Psychogenic Non-Epileptic Seizures on Four Measures of Effort. Clin Neuropsychol 2006; 20: 552–66.

Dandachi-FitzGerald B, Ponds RWHM, Peters MJ V, Merckelbach H. Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. Clin Neuropsychol 2011; 25: 812–28.

Davis JJ. Performance validity in older adults: Observed versus predicted false positive rates in relation to number of tests administered. J Clin Exp Neuropsychol 2018; 40: 1013–21.

Dean A, Victor T, Boone K, Philpott L, Hess A. Dementia and effort test performance. Clin Neuropsychol 2009; 23: 133–52.

Dorociak K, Schulze E, Piper L, Molokie R, JK J. Performance validity testing in a clinical sample of adults with sickle cell disease. Clin Neuropsychol 2018; 32: 81–97.

Drane DL, Williamson DJ, Stroup ES, et al. Cognitive impairment is not equal in patients with epileptic and psychogenic nonepileptic seizures. Epilepsia 2006; 47: 1879–86.

Duff K, Spering CC, O'Bryant SE, et al. The RBANS effort index: Base rates in geriatric samples. Appl Neuropsychol 2011; 18: 11–7.

Erdodi LA, Abeare CA, Lichtenstein JD, et al. Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) processing speed scores as measures of noncredible responding: The third generation of embedded performance validity indicators. Psychol Assess 2017; 29: 148–57.

Erdodi L, Roth R. Low scores on BDAE Complex Ideational Material are associated with invalid performance in adults without aphasia. Appl Neuropsychol Adult 2017; 24: 264–74.

Fazio RL, Faris AN, Yamout KZ. Use of the Rey 15-Item Test as a performance validity test in an elderly population. Appl Neuropsychol Adult 2019; 26: 28–35.

Goldberg J, Miller H. Performance of psychiatric inpatients and intellectually deficient individuals on a task that assesses the validity of memory complaints. J Clin Psychol 1986; 42: 792–5.

Goodrich-Hunsaker NJ, Hopkins RO. Word memory test performance in amnesic patients with hippocampal damage. Neuropsychology 2009; 23: 529–34.

Gorissen M, Sanz JC, Schmand B. Effort and cognition in schizophrenia patients. Schizophr Res 2005; 78: 199–208.

Green P, Flaro L. Results From Three Performance Validity Tests (PVTs) in Adults With Intellectual Deficits. Appl Neuropsychol Adult 2015; 22: 293–303.

Green P, Montijo J, Brockhaus R. High specificity of the Word Memory Test and Medical Symptom Validity Test in groups with severe verbal memory impairment. Appl Neuropsychol 2011; 18: 86–94.

Grote CL, Kooker EK, Garron DC, Nyenhuis DL, Smith CA, Mattingly ML. Performance of compensation seeking and non-compensation seeking samples on the Victoria symptom validity test: cross-validation and extension of a standardization study. J Clin Exp Neuropsychol 2000; 22: 709–19.

Hampson NE, Kemp S, Coughlan AK, Moulin CJA, Bhakta BB. Effort test performance in clinical acute brain injury, community brain injury, and epilepsy populations. Appl Neuropsychol Adult 2014; 21: 183–94.

Heintz CEJ, van Tricht MJ, van der Salm SMA, et al. Neuropsychological profile of psychogenic jerky movement disorders: importance of evaluating non-credible cognitive performance and psychopathology. J Neurol Neurosurg Psychiatry 2013; 84: 862–7.

Hill S, Ryan L, Kennedy C, Malamut B. The relationship between measures of declarative memory and the test of memory malingering in patients with and without temporal lobe dysfunction. J Forensic Neuropsychol 2003; 3: 1–18.

Hoskins LL, Binder LM, Chaytor NS, Williamson DJ, Drane DL. Comparison of oral and computerized versions of the word memory test. Arch Clin Neuropsychol 2010; 25: 591–600.

Howe LLS, Anderson AM, Kaufman DAS, Sachs BC, Loring DW. Characterization of the Medical Symptom Validity Test in evaluation of clinically referred memory disorders clinic patients. Arch Clin Neuropsychol 2007; 22: 753–61.

Hunt S, Root JC, Bascetta BL. Effort testing in schizophrenia and schizoaffective disorder: validity indicator profile and test of memory malingering performance characteristics. Arch Clin Neuropsychol 2014; 29: 164–72.

Iverson GL, Le Page J, Koehler BE, Shojania K, Badii M. Test of Memory Malingering (TOMM) scores are not affected by chronic pain or depression in patients with fibromyalgia. Clin Neuropsychol 2007; 21: 532–46.

Janssen MAM, Bertens D, Kessels L, Kessels RPC, Koopmans PP. A case-control pilot study on cognitive functioning, symptom validity and psychological wellbeing in HIV-1-infected patients in the Netherlands. Int J STD AIDS 2013; 24: 387–91.

Kanser RJ, Rapport LJ, Bashem JR, Hanks RA. Detecting malingering in traumatic brain injury: Combining response time with performance validity test accuracy. Clin Neuropsychol 2019; 33: 90–107.

Keary T, Frazier T, Belzile C, et al. Working memory and intelligence are associated with victoria symptom validity test hard item performance in patients with intractable epilepsy. J Int Neuropsychol Soc 2013; 19: 314–23.

Kemp S, Coughlan AK, Rowbottom C, Wilkinson K, Teggart V, Baker G. The base rate of effort test failure in patients with medically unexplained symptoms. J Psychosom Res 2008; 65: 319–25.

Lee A, Boone K, Lesser I, Wohl M, Wilkins S, Parks C. Performance of older depressed patients on two cognitive malingering tests: False positive rates for the Rey 15-item memorization and Dot Counting Tests. Clin Neuropsychol 2000; 14: 303–8.

Loring D, Lee G, Meador K. Victoria Symptom Validity Test performance in non-litigating epilepsy surgery candidates. J Clin Exp Neuropsychol 2005; 27: 610–7.

Loring DW, Goldstein FC, Chen C, et al. False-Positive Error Rates for Reliable Digit Span and Auditory Verbal Learning Test Performance Validity Measures in Amnestic Mild Cognitive Impairment and Early Alzheimer Disease. Arch Clin Neuropsychol 2016; 31: 313–31.

Macciocchi SN, Seel RT, Alderson A, Godsall R. Victoria Symptom Validity Test performance in acute severe traumatic brain injury: implications for test interpretation. Arch Clin Neuropsychol 2006; 21: 395–404.

Macciocchi SN, Seel RT, Yi A, Small S. Medical Symptom Validity Test Performance Following Moderate-Severe Traumatic Brain Injury: Expectations Based on Orientation Log Classification. Arch Clin Neuropsychol 2017; 32: 339–48.

Maiman M, Del Bene VA, MacAllister WS, et al. Reliable digit span: Does it adequately measure suboptimal effort in an adult epilepsy population? Arch Clin Neuropsychol 2019; 34: 259–67.

Merten T, Bossink L, Schmand B. On the limits of effort testing: symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. J Clin Exp Neuropsychol 2007; 29: 308–18.

Meyer SRA, de Jonghe JFM, Schmand B, Ponds RWHM. The Visual Association Test-Extended: a cross-sectional study of the performance validity measures. Clin Neuropsychol 2017; 31: 798–813.

Moore R, Davine T, Harmell A, Cardenas V, Palmer B, Mausbach B. Using the repeatable battery for the assessment of neuropsychological status (RBANS) effort index to predict treatment group attendance in patients with schizophrenia. J Int Neuropsychol Soc 2013; 19: 198–205.

Morra L, Gold J, Sullivan S, Strauss G P. Predictors of neuropsychological effort test performance in schizophrenia. *Schizophr Res* 2015; **162**: 205–10.

Novitski J, Steele S, Karantzoulis S, Randolph C. The Repeatable Battery for the Assessment of Neuropsychological Status Effort scale. Arch Clin Neuropsychol 2012; 27: 190–5.

Oudman E, Krooshof E, van Oort R, Lloyd B, Wijnia JW, Postma A. Effects of korsakoff amnesia on performance and symptom validity testing. Appl Neuropsychol Adult 2019;: 1-9 https://doi.org/10.1080/23279095.2019.1576180

Paul R, Rhee G, Baker LM, Vaida F, Cooley SA, Ances BM. Effort and neuropsychological performance in HIV-infected individuals on stable combination antiretroviral therapy. J Neurovirol 2017; 23: 725–33.

Price K, Desantis S, Simpson A, et al. The impact of clinical and demographic variables on cognitive performance in methamphetamine-dependent individuals in rural South Carolina. Am J Addict 2011; 20: 447–55.

Rees L, Tombaugh T, Gansler D, Moczynski N. Five validation experiments of the Test of Memory Malingering (TOMM). Psychol Assess 1998; 10: 10–20.

Rees L, Tombaugh T, Boulay L. Depression and the Test of Memory Malingering. Arch Clin Neuropsychol 2001; 16: 501–6.

Roor JJ, Knoop H, Dandachi-FitzGerald B, Peters MJ V, Bleijenberg G, Ponds RWHM. Feedback on underperformance in patients with chronic fatigue syndrome: The impact on subsequent neuropsychological test performance. Appl Neuropsychol Adult 2018;27(2): 188-196

Rossetti MA, Collins RL, York MK. Performance Validity in Deep Brain Stimulation Candidates. Arch Clin Neuropsychol 2018; 33: 508–14.

Rudman N, Oyebode JR, Jones CA, Bentham P. An investigation into the validity of effort tests in a working age dementia population. Aging Ment Health 2011; 15: 47–57.

Ruocco A. Compliance on neuropsychological performance validity testing in patients with borderline personality disorder. Psychol Assess 2016; 28: 345–50.

Schroeder RW, Marshall PS. Evaluation of the appropriateness of multiple symptom validity indices in psychotic and non-psychotic psychiatric populations. Clin Neuropsychol 2011; 25: 437–53.

Sherer M, Davis LC, Sander AM, et al. Factors Associated with Word Memory Test Performance in Persons with Medically Documented Traumatic Brain Injury. Clin Neuropsychol 2015; 29: 522–41.

Sieck BC, Smith MM, Duff K, Paulsen JS, Beglinger LJ. Symptom validity test performance in the Huntington Disease Clinic. Arch Clin Neuropsychol 2013; 28: 135–43.

Singhal A, Green P, Shankar K, Gill D. High specificity of the Medical Symptom Validity Test in patients with very severe memory impairment. Arch Clin Neuropsychol 2009; 24: 721–8.

Stevens A, Schneider K, Liske B, Hermle L, Huber H, Hetzel G. Is subnormal cognitive performance in schizophrenia due to lack of effort or to cognitive impairment? Ger J Psychiatry 2014; 17: 1–9.

Strauss G, Morra L, Sullivan S, Gold J. The role of low cognitive effort and negative symptoms in neuropsychological impairment in schizophrenia. Neuropsychology 2015; 29: 282–91.

Teichner G, Wagner M. The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. Arch Clin Neuropsychol 2004; 19: 455–64.

Terry DP, Adams TE, Ferrara MS, Miller LS. fMRI hypoactivation during verbal learning and memory in former high school football players with multiple concussions. Arch Clin Neuropsychol 2015; 30: 341–55.

Tyson B, Baker S, Greenacre M, et al. Differentiating epilepsy from psychogenic nonepileptic seizures using neuropsychological test data. Epilepsy Behav 2018; 87: 39–45.

Van Der Werf S, Prins J, Jongen P, Van Der Meer J, Bleijenberg G. Abnormal neuropsychological findings are not necessarily a sign of cerebral impairment. A matched comparison between chronic fatigue syndrome and multiple sclerosis. Neuropsychiatry, Neuropsychol Behav Neurol 2000; 13: 199–203.

Walter J, Morris J, Swier-Vosnos A, Pliskin N. Effects of severity of dementia on a symptom validity measure. Clin Neuropsychol 2014; 28: 1197–208.

Whearty K, Allen C, Lee B, GP S. The evaluation of insufficient cognitive effort in schizophrenia in light of low IQ scores. J Psychiatr Res 2015; 68: 397–404.

Wodushek TR, Domen CH. Comparing two models of performance validity assessment in patients with parkinson's disease who are candidates for deep brain stimulation surgery. Appl Neuropsychol Adult 2018; 27(1): 9-21

Woods SP, Conover E, Weinborn M, et al. Base rate of Hiscock Digit Memory Test failure in HIV-associated neurocognitive disorders. Clin Neuropsychol 2003; 17: 383–9.

Wu TC, Allen MD, Goodrich-Hunsaker NJ, Hopkins RO, Bigler ED. Functional neuroimaging of symptom validity testing in traumatic brain injury. Psychol Inj Law 2010; 3: 50–62.

Zenisek R, Millis SR, Banks SJ, Miller JB. Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. Arch Clin Neuropsychol 2016; 31: 426–33.