REVIEW

Performance validity test failure in clinical populations—a systematic review

Laura McWhirter , Craig W Ritchie, Jon Stone , Alan Carson

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jnnp-2020-323776).

Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK

Correspondence to
Dr Laura McWhirter, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH8 9YL, UK; laura.mcwhirter@ed.ac.uk

Received 8 May 2020
Revised 3 June 2020
Accepted 7 June 2020
Published Online First 10 July 2020

To cite: McWhirter L, Ritchie CW, Stone J, *et al*. *J Neurol Neurosurg Psychiatry* 2020;**91**:945–952.

ABSTRACT

Performance validity tests (PVTs) 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. 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. 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%. PVT 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 PVT 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 (PVTs), 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 test 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: that is, 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 PVT failure rates in clearly defined (by diagnosis) non-litigating, non-forensic, non-military, non-military veteran, **clinical** populations. Second, we aimed to consider the implications of our findings in terms of the uses of PVTs in clinical practice.

METHODS

Search strategy and selection criteria

We systematically searched the published peer-reviewed English language literature in MEDLINE, EMBASE and PsycINFO databases from inception to 5 July 2019 (figure 1). 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 PVTs (not symptom validity questionnaires) in one or more individuals with a recorded clinical diagnosis of a

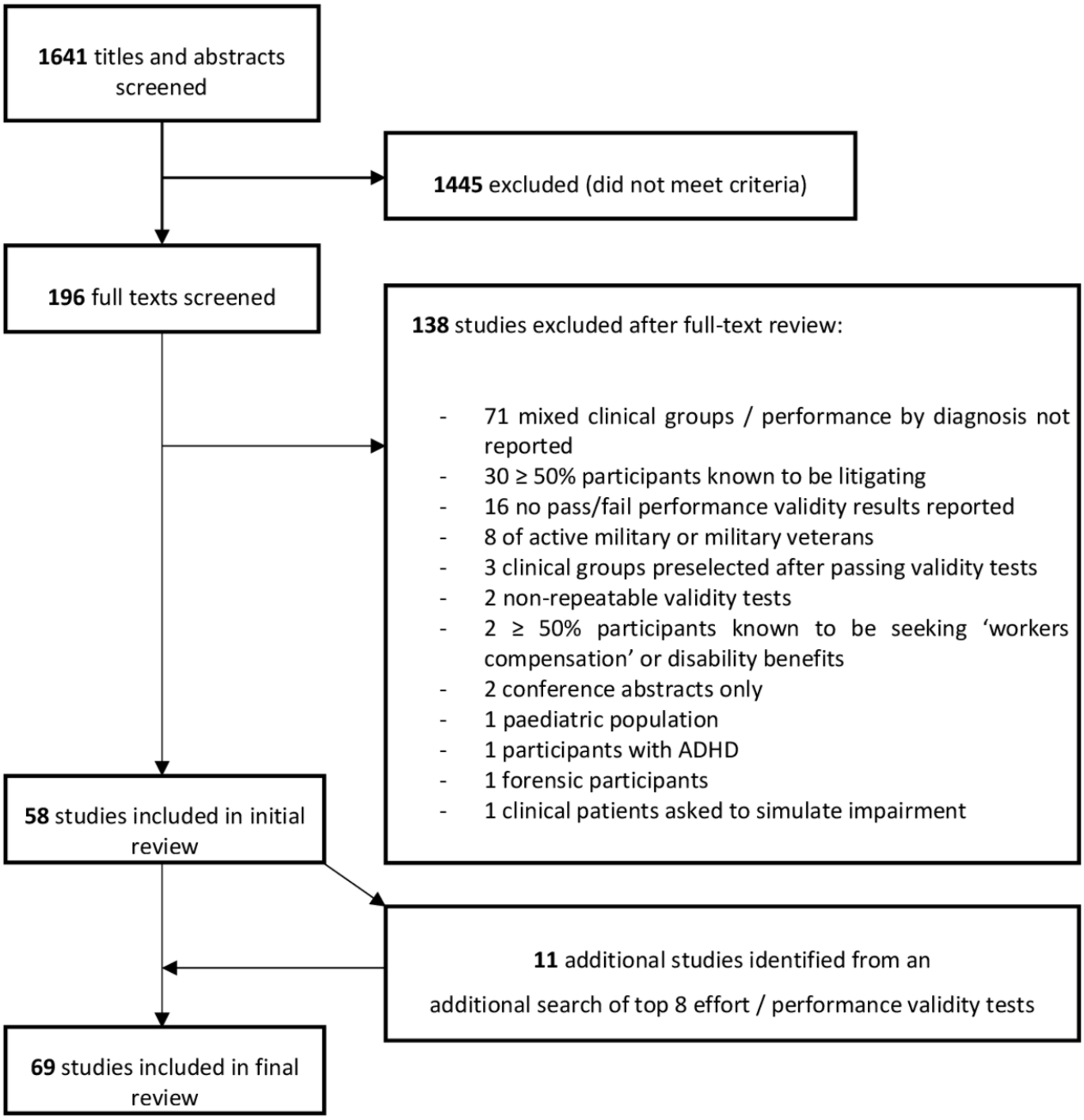

Figure 1 Selection of included studies.

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. 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 15 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 one author (LM) using Excel version 16, 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.

RESULTS

Search results and screening

Forty-five different PVTs or indices were identified (online 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% CIs), 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 (RDS, 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 medico-legal) 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 online 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

Three studies described PVT performance in adults with intellectual disability. In Goldberg and Miller, 6/16 (38%) adults with mean IQ (Intelligence Quotient) 63.9 failed (<9) the Rey 15-item test.[7] In the largest study included, 6 (2%) of 276 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 MSVT (criterion A) and 11 (5%) of 223 failed the WMT.[6] In the same study, 2 (14%) of 14 individuals in the same circumstances but with FSIQ (Full Scale Intelligence Quotient) ≤70 failed the WMT and 0 of 17 failed the MSVT.[6] (online supplementary table 2)

Mild cognitive impairment (MCI)

Nine studies reported PVT performance in 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 153 (42%) of 365 individuals with amnestic MCI in Loring *et al*; 29 (36%) of 80 with minor neurocognitive disorder failed the Rey 15-item test (cut-off <20) in Fazio *et al*; 1462 (27%) of 5414 with MCI failed the logical memory test (cut-off <14) and 1354 (25%) of 5414 failed semantic word generation (cut-off <13) in Davis, and 13 (22%) of 60 individuals with 'probable MCI' in Green *et al*.[8–11] Of note, 11 of the 13 MCI individuals in Green *et al* who failed criterion A of the WMT did not meet criterion B (easy–hard difference <30) and so had a possible dementia profile.[11] Pooled failure rates for RDS 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] (online supplementary table 3)

Functional disorders

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), 13 (10%) of 132 failed the TOMM.[15–17] In two other studies of PNES, 25 (44%) of 57 met criterion A (therefore failed) on the standard WMT.[18 19] In two studies of individuals with chronic fatigue syndrome, 374 (25%) of 1526 failed the Amsterdam Short Term Memory Test (scoring <86/100).[20 21](online supplementary table 4)

Failure rates higher than 25% were reported by Tyson *et al* in 33 individuals with PNES on RDS (cut-off ≤7), vocabulary–digit span (≥3), forced choice recall on the California Verbal Learning Test (CVLT) (≤15) and the Boston Naming Test.[17]

Epilepsy

Eleven studies reported PVT performance in people with epilepsy. In five studies including 246 people with epilepsy, 31 (13%) 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 WMT.[19 24 25] Two studies reported RDS 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. (online supplementary table 5)

Notably, Tyson *et al* reported higher failure rates in epilepsy compared with a group with PNES (see online supplementary table 4) in six of eight tests included (TOMM, RDS, 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 WMT in epilepsy and PNES (31% and 29% respectively).

Acquired brain injury

The studies included in online 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 (mTBI). Results in this group as a whole were highly variable, suggesting between-group differences. Most studies in mTBI reported low PVT failure rates (<20%). In contrast, however, Novitski *et al* reported failure rate of 52% (13/25) on Reversible Battery for the Assessment of Neuropsychological Status (RBANS) digit span (cut-off <9) in 25 individuals who had sustained a mTBI more than 6 months previously, and Erdodi and Roth 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 mTBI failed on criterion A of the WMT: 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 WMT results after moderate and severe brain injury, resulting in a pooled failure rate of 28% (63 of 228; 95% CI 22% to 34%).[25 27 28] Results of other tests studied in moderate and severe brain injury were heterogeneous. Macciocchi *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 MSVT 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 Wechsler Adult Intelligence Scale (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 WMT.[33 34] Oudman *et al* reported that 2 (10%) of 20 individuals with Korsakoff Amnesia failed the second trial of the TOMM.[35]

Neurodegenerative disease

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 (online supplementary table 7, figure 2).

The WMT and MSVT were most frequently described. Green *et al* reported high failure rates in clinically defined 'probable, mild and moderate' dementia on the WMT (71% of 42) and MSVT (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 MSVT in 13 with mild dementia, all of whom met the 'dementia profile', and 15 (83%) of 18 with advanced dementia met the 'dementia profile'.[36] 18 (90%) of 20 mild Alzheimer's dementia examinees in Merten *et al*'s study failed the delayed recall component of the WMT, 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 MSVT.[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 (10%) of 47 failed the MSVT 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

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 WMT.[42] In contrast, Schroeder and Marshall'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 study of 60 individuals with schizophrenia or schizoaffective disorder reported that 28% failed RDS ≤6% and 36% failed finger-tapping.[44] (online supplementary table 8)

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 Autistic Spectrum Disorder (ASD) and 14% with ADHD.[47] Price *et al* reported no failures on the TOMM in 71 individuals with methamphetamine dependence.[48]

Other conditions

Heintz *et al* reported 23% of 13 individuals with Gilles de la Tourette syndrome failed the Amsterdam Short-term Memory Test (ASMT).[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 RDS with a ≤7 cut-off.[52] In Rossetti *et al*, 2 of 10 deep brain stimulation

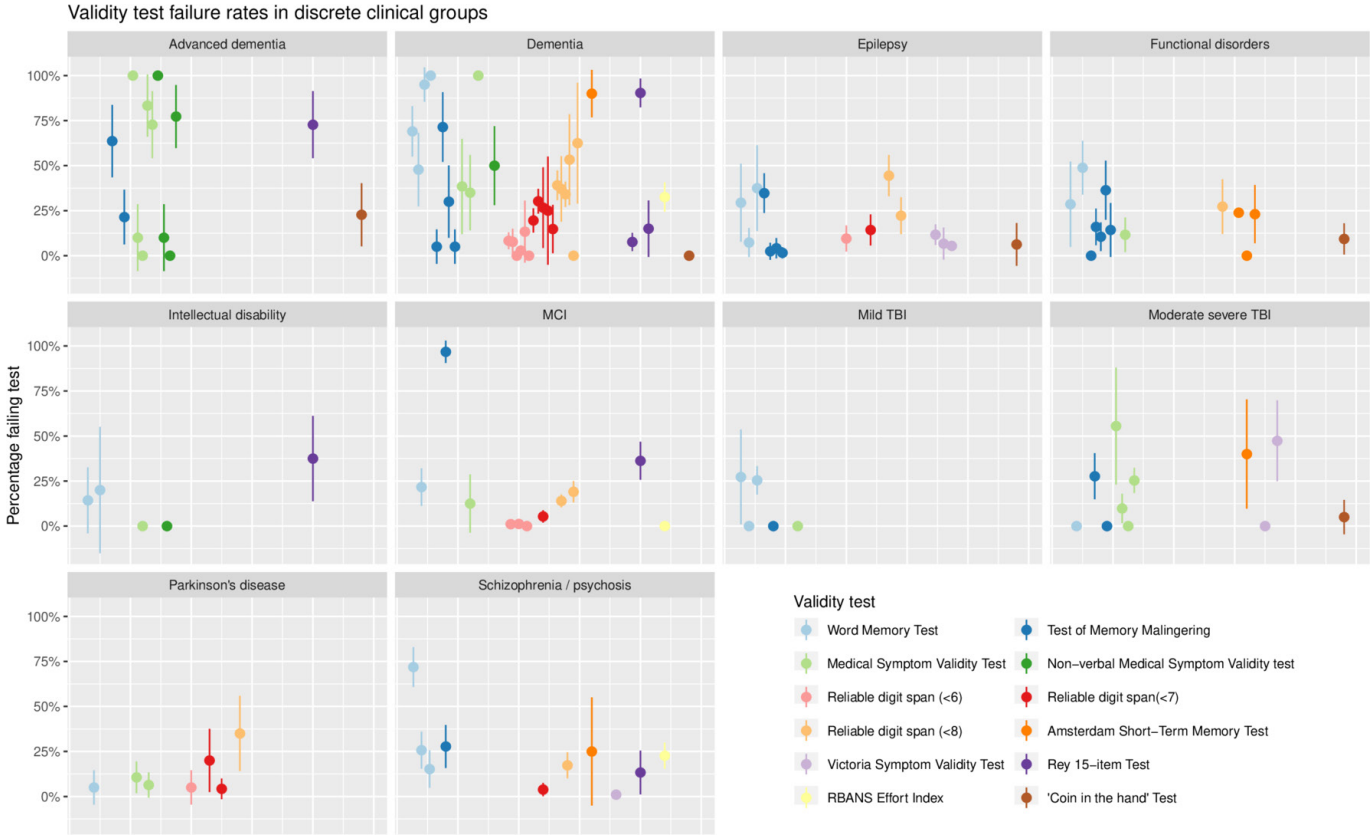

Figure 2 Failure rates in the 12 most frequently reported tests by diagnosis. Each point represents reported failure rate, in a particular test (indicated by colour), as reported by an individual study. Points are grouped along the X axis in the same test (colour) order in each plot, so as to allow visual comparison between plots. Vertical lines indicate the asymptotic 95% CI for each reported failure rate. MCI, mild cognitive impairment; TBI, traumatic brain injury;

candidates with essential tremor failed the WMT.[41] (online supplementary table 9)

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.

DISCUSSION

Our review suggests that failure of PVTs 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 TBI, and both mild and moderate-to-severe dementia (where the 'severe impairment' profile on the WMT often applies). Of note, while some individuals with functional disorders fail PVTs, failure rates are no higher than in a range of other diverse conditions, including epilepsy, and MCI.

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 TBI 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 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 WMT and the MSVT—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 PVTs.

The poor quality of the PVT 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 'non-credible' 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 'non-credible': (1) motive to feign symptoms (litigation or seeking disability compensation), (2) failure on two independent PVTs and (3) 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. First, 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 '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 first 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.

Twitter Laura McWhirter @lauramcw, Craig W Ritchie @Craig_ritchie68 and Jon Stone @jonstoneneuro

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

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests 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 JS is a Chief Scientists Office NHS Research Scotland Career Researcher. CWR declares no conflicts of interest.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Laura McWhirter http://orcid.org/0000-0001-9839-6549
Jon Stone http://orcid.org/0000-0001-9829-8092

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, testing Svalidity. 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:e1000097.
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, *et al*. 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, *et al*. 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, *et al*. Cognitive functioning in fibromyalgia: the central role of effort. *J Psychosom Res* 2016;87:30–6.
15 Cragar DE, Berry DTR, Fakhoury TA, *et al*. Performance of patients with epilepsy or psychogenic non-epileptic seizures on four measures of effort. *Clin Neuropsychol* 2006;20:552–66.
16 Hill SK, Ryan LM, Kennedy CH, *et al*. 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 BT, 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, *et al*. 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, *et al*. 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, *et al*. Feedback on underperformance in patients with chronic fatigue syndrome: the impact on subsequent neuropsychological test performance. *Appl Neuropsychol Adult* 2020;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, *et al*. 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, *et al*. The repeatable battery for the assessment of neuropsychological status effort scale. *Arch Clin Neuropsychol* 2012;27:190–5.
27 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.
28 TC W, Allen MD, Goodrich-Hunsaker NJ, *et al*. Functional neuroimaging of symptom validity testing in traumatic brain injury. *Psychol Inj Law* 2010;3:50–62.
29 Macciocchi SN, Seel RT, Alderson A, *et al*. 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, *et al*. 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, *et al*. Effects of Korsakoff amnesia on performance and symptom validity testing. *Appl Neuropsychol Adult* 2019:1–9.
36 Howe LLS, Anderson AM, Kaufman DAS, *et al*. 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, *et al*. 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, Ashaye K, *et al*. 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* 2020;27:9–21.
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 Wheatly KM, Allen DN, Lee BG, *et al*. 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 KB, Lesser I, *et al*. 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 MJV, *et al*. Cognitive underperformance and symptom over-reporting in a mixed psychiatric sample. *Clin Neuropsychol* 2011;25:812–28.

48 Price KL, DeSantis SM, Simpson AN, *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, *et al*. 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, *et al*. Effort and neuropsychological performance in HIV-infected individuals on stable combination antiretroviral therapy. *J Neurovirol* 2017;23:725–33.
52 Dorociak KE, Schulze ET, Piper LE, *et al*. 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, *et al*. Functional cognitive disorders: a systematic review. *Lancet Psychiatry* 2020;7:191–207.
55 Slick DJ, Sherman EM, 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, *et al*. 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.