# The Development of Performance Validity Tests (PVTs) for Indonesia: A Literature Review and Recommendations

Widhi Adhiatma[*1,2], Marc P. H. Hendriks[2,3], Magdalena S. Halim[1], Roy P. C. Kessels[2,4]

[1]*Faculty of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia*

[2]*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

[3]*Academic Centre of Epileptology, Kempenhaeghe, Heeze, The Netherlands* [4]*Vincent van Gogh Institute for Psychiatry, Venray, The Netherlands*

**Abstract.** Neuropsychological tests are proliferating in Indonesia. So far, unfortunately, it is not accompanied by the development of a performance validity test (PVT). According to international neuropsychological standards, using PVTs is essential to determine the validity of the neuropsychological test results. To date, there is no single standardized PVT available in Indonesia. In this article, we describe (1) the concept of performance validity testing, (2) the detection strategy, and (3) the procedure of PVT. Furthermore, several factors which affect PVT failure and its implication on the interpretation of a neuropsychological assessment are discussed. Considering the importance of PVTs incorporation and that it has become a standard by some neuropsychological associations, it is crucial to develop PVTs in Indonesia. Finally, the utilization of PVT should be performed cautiously, as many factors might influence the PVT result.

*Keywords:* neuropsychological tests; malingering; motivation; deception; cognition disorders

## Introduction

A neuropsychological test as part of a thorough neuropsychological test battery aims to give insights into the level of specific cognitive functions or dysfunctions (as the consequences of cerebral damage) (Kessels & Hendriks, 2022). Furthermore, this cognitive profile gives insights into the kind of rehabilitation or intervention that could be provided to patients. The first neuropsychological tests were developed in the 1940s to identify patients' cerebral dysfunction attributable to an 'organic' disease process (Kolb & Whishaw, 2015). Unfortunately, until 2008, only a few neuropsychological tests were available in Indonesia (Rambe, 2008). Since then, psychologists and neurologists in Indonesia have started developing neuropsychological tests. For example, the Indonesian version of the Wechsler Adult Intelligence Scale – Fourth Edition (Suwartono et al., 2014), the Indonesian version of the Montreal Cognitive Assessment (Husein et al., 2010), Screening Test for Luria Nebraska Neuropsychological Battery (Syahroni et al., 2011), Consortium to Establish a Registry for Alzheimer's Disease (Indrajaya et al., 2013), Status Mental Neurologi Strub dan Black (Widyarini et al., 2013),

---

[*]Address for correspondence: widhi.adhiatma@atmajaya.ac.id

Modified Mini-Mental State-Test (Khairunnisa et al., 2014), Indonesian Boston Naming Test (Sulastri et al., 2019), Rey Auditory Verbal Learning Test (Utami et al., 2019), and the HIV Dementia Scale (Dewiyana et al., 2019) have been developed and validated for Indonesian samples. The availability of these neuropsychological tests has enhanced the quality of neuropsychological assessments in Indonesia.

Clinical neuropsychology experts conclude that assurance of performance validity should accompany neuropsychological tests interpretation, that is, the examinee's performance in completing the tests is consistent with his or her actual cognitive ability (Bush et al., 2005; Heilbronner et al., 2009; Larrabee, 2012). Performance validity assessment is imperative because neuropsychological tests do not directly measure the intended abilities (e.g., memory). Neuropsychological tests measure behaviors from which inferences regarding these abilities are made (e.g., repeating some numbers mentioned by the tester), whereas these behaviors are expected to result from a valid performance (Bigler, 2012). Thus, a valid performance will reflect the examinee's actual cognitive ability so that the results of the neuropsychological tests can be interpreted validly (Laurent et al., 2019).

In the past, clinicians used to employ clinical judgment to determine the test results' validity (e.g., by comparing the consistency of test results and reported symptoms based on clinical observation). However, clinical judgment accuracy in detecting performance validity was only 53.3%, indicating poor accuracy (Dandachi-FitzGerald et al., 2017). The inaccuracy of clinical judgment is associated with the lack of objective standards, and accurate performance validity determination cannot be reached merely by analyzing the examinee's neuropsychological test results (Bigler, 2012; Greher & Wodushek, 2017; Guilmette, 2012; McGuire et al., 2019).

Therefore, clinical neuropsychologists should base their judgment regarding performance validity more on a structured and objective methodology, namely a validity test. With such a method, clinicians could directly measure the probability of invalid performance when the examinee obtains above or below a determined cut-off score; in another case, the clinicians could determine invalid performance based on a particular profile of the validity test. At first, the validity test was embedded in one of the most popular personality tests, the Minnesota Multiphasic Personality Inventory (Butcher, 2005), which aimed to detect invalid psychiatric symptoms. Later, this type of validity testing was called the symptom validity test (Larrabee, 2012). The assessment of validity in the field of neuropsychology (referred to as performance validity) emerged later (Rogers & Correa, 2008). To date, in Indonesia, validity testing is merely limited to SVT, such as the validity scales in MMPI-2-RF (Adhiatma & Halim, 2019), while the validity testing for the neuropsychological or cognitive test, which is referred to as performance validity test (PVT), is not available.

PVT allows clinical neuropsychologists to detect an invalid performance during neuropsychological testing (Critchfield et al., 2019; Richey & Doninger, 2020). PVTs resemble neuropsychological domain measures, such as memory or processing speed. However, PVTs are constructed to be considerably easier to complete than the "actual" neuropsychological tests that measure cognitive performance so that even most patients with cognitive impairments can pass PVTs (Larrabee, 2015).

Research on PVTs revealed some remarkable findings regarding performance invalidity rates in several settings. For instance, the failure rate is higher when the examinee perceives potential external incentives (e.g., disability compensation or avoiding a lawsuit) by demonstrating cognitive impairments. When external incentives were present, the invalid performance reached 40-50% (Larrabee et al., 2009). In contrast, invalid performance rate in clinical settings (i.e., psychiatric clinics, ADHD evaluations, neuropsychological evaluation for patients with mild traumatic brain injury [mTBI] history, and patients evaluated for dementia) when external incentives were absent was lower with an approximation of 5-20% (Martin & Schroeder, 2020). These facts show that invalid performance during neuropsychological evaluation is present in many settings, regardless of the availability of external incentives.

With the proliferation of neuropsychological tests in Indonesia, including the development of an online platform and dynamic database of neuropsychological tests (Wahyuningrum et al., 2021), PVT will become necessary for clinical practice. The utilization of PVTs in clinical practice is also echoed by some associations, such as the National Academy of Neuropsychology (NAN) and the American Academy of Clinical Neuropsychology (AACN) (Bush et al., 2005; Heilbronner et al., 2009). As no PVTs are available in Indonesia, in this paper, we discussed PVT by elaborating on its history, detection strategy, and procedure of PVT. Also, we described factors that contribute to PVT failure and its implication in interpreting the results of the neuropsychological tests.

## Discussion

*History of PVT: From Case Study to Detection Strategy*

The introduction of PVTs cannot be separated from the advancement of detection strategy in symptom validity assessment. At first, symptom validity assessment was explicitly aimed at detecting feigning. Then, in the 19th century, feigning, or more specifically malingering, received more attention from clinicians in psychiatric settings (Rogers & Correa, 2008). According to DSM-5, malingering is the intentional production of false or exaggerated physical or psychological symptoms, motivated by external incentives, such as avoiding military duty or work, obtaining financial compensation, eluding criminal prosecution, or obtaining drugs (Association, 2013). At that period, malingering was determined case-by-case based on a range of indicators, such as patients' behaviors during the clinical interview (e.g., eye contact), feigned presentations (e.g., symptoms increase when under observation), areas of intact functioning (e.g., there is no disturbance in appetite or sleep), and atypical symptoms (e.g., absurd thoughts). However, these methods were criticized due to a lack of standard and independent criteria, i.e., salient characteristics of invalid performers (Rogers & Correa, 2008). A more objective approach was developed along with the construction of the MMPI. The MMPI included two test-taking approaches: underreporting (or faking good) and overreporting (or faking bad). This approach relies on score differences between groups (i.e., the non-clinical group who were asked to underreport or overreport vs. the genuine psychotic group) on several validity scales; for instance, score difference on the F scale (Hathaway & McKinley, 1940) (a higher score on this scale indicates a

tendency to endorse uncommon symptoms).

In the subsequent development, the detection strategy for validity assessment was done with a standardized method. This strategy is empirically validated and derives from a concept of how the examinee approaches the test. Researchers also considered that feigned psychiatric disorders and feigned cognitive impairments should be addressed independently. As previously mentioned, researchers termed the symptom validity test (SVT), which aimed to detect the accuracy of symptomatic complaints on self-report measures, and differentiated it from PVT, which focused on feigned cognitive impairments. Compared to self-report symptom questionnaires, such as MMPI, cognitive tests require an examinee to complete a performance test; therefore, a different strategy should be used in PVT (Larrabee, 2012; Rogers & Correa, 2008). However, in this paper, we elaborated exclusively on PVT.

Before the PVT terminology was coined, these tests were mentioned as "malingering tests" or "effort tests." Currently, this terminology is not used anymore due to conceptual reasons. As mentioned above, malingering requires an "intentional" element for its determination; therefore, if the term "malingering test" was used, it should be able to determine that the test failure was caused by "intention to fail." However, "intent" is complicated to prove scientifically and in clinical practice. Furthermore, clinical neuropsychologists considered the negative impact when an examinee was labeled as "malingering" or even when it is concluded that the examinee was showing "poor effort" during the testing. Currently, clinical neuropsychologists consider the term "PVT" an accurate description because this term avoids potential damage to the examinee while preventing a presumptuous argument when an examinee fails on PVTs (Greher & Wodushek, 2017).

PVTs were initially applied in a forensic setting where external incentives were highly suspected. It was indicated that 50% of other neuropsychological test performance might be explained by PVT in this setting (Critchfield et al., 2019). However, PVTs are now also commonly incorporated in neuropsychological evaluations in clinical settings with neurological or psychiatric patients, as performance invalidity is also found in these settings, even when the external incentive is absent (Bodner et al., 2019; Loring & Goldstein, 2019). Therefore, the inclusion of PVTs in a neuropsychological assessment is imperative, regardless of the presence of external incentives.

*PVT Detection Strategies*

Rogers categorizes PVTs into two detection strategies: Amplified Detection Strategies (ADS) and Unlikely Detection Strategies (Rogers, 2018a).

*Amplified Detection Strategies (ADS)*

ADS depends on the degree and intensity of the reported characteristics. For instance, if an examinee reports that he or she has a deficit in memory ability, the detection strategy will rely on the memory complaints degree and intensity rather than determining the presence of the actual impairment itself (Rogers & Correa, 2008). Two strategies are included in the ADS: a floor effect and a significantly below-chance performance (Rogers & Correa, 2008).

The floor effect strategy assumes that invalid performers do not recognize that patients with actual cognitive impairments are able to complete an elementary cognitive test. So, for instance, remembering a two-digit number in a sequence (e.g., 8-4) can be repeated correctly by severely impaired patients due to psychiatric disorders or neurological diseases like Alzheimer's dementia. Unfortunately, when invalid performers receive coaching (e.g., "Just succeed on this test"), it will become effortless to pass this test, primarily when the test is used as a standalone measure. The floor effect strategy has been adapted to dozens of feigning measures because it is relatively easy to adapt (Rogers, 2018a).

One of the most popular PVTs that utilizes this approach is the Test of Memory Malingering (Tombaugh, 1996). The TOMM employs a forced-choice format using 50 pictures of common objects. In the first trial, these stimuli are presented individually for three seconds. Then, each stimulus is presented with a distractor stimulus. The examinee is asked to choose the correct original stimulus. Feedback is given to indicate the response's correctness. In the second trial, the 50 original stimuli are shown again, but now in a different order, followed by immediate forced-choice recognition. A Retention Trial is administered after a 15-minute delay after the second trial, in which the examinee is given the same forced-choice recognition task. A score below the cut-off (< 45) in Trial 1 or Trial 2 is associated with possible invalid performance (Martin et al., 2020).

Another example of a PVT that utilizes a floor effect is the Amsterdam Short-Term Memory (ASTM). The ASTM consists of 30 items, each with three conditions. In the first condition, the examinee will face five words from a common semantic category, for instance, clothes (e.g., pants, skirt, shirt, sweater, coat). The examinee is instructed to read the words aloud and memorize them. In the second condition, the examinee has to answer a simple addition or subtraction task as a distractor (e.g., 18 + 7 = ?). In the third condition, the examinee will see five words from the same semantic category (i.e., clothes) as before (e.g., suit, pants, skirt, sweater, pajamas). The examinee must mention the three words presented in the first condition. Feedback is given on the number of correct words. This task is considerably easy because the two new words in the third condition have a lower level of familiarity in the respected semantic category (Schagen et al., 1997).

In validating floor effect strategy PVTs, the tests should be evaluated so that most patients with severe cognitive impairments, such as dementia, can still perform above this floor level (Rogers, 2018b). To overcome this, the ASTM uses different semantic categories in each item, so there is no build-up of memory interference between items. This strategy proved effective, as no significant differences have been found between closed-head injury patients and a cognitively unimpaired control group (Schagen et al., 1997).

The significantly below-chance performance (SBCP) strategy uses a forced-choice paradigm to calculate the probability of a below-chance performance. The most impaired examinee can succeed approximately 50% of the chance levels when the test items contain two equiprobable choices. Therefore, a performance below this chance level is considered highly unlikely, thus a strong indication of an invalid performance (Rogers, 2018b). In other words, the examinee intends to perform poorly on the test (Bigler, 2012; Binder et al., 2014). Furthermore, based on the proposed Neurocognitive

Malingering diagnostic criteria, the evidence of failure in one SBCP PVT shows invalid presentation indicative of feigning or exaggeration (Sherman et al., 2020). Nevertheless, SBCP is easily susceptible to coaching, which is the main limitation of the SBCP approach. When an examinee tries to feign severe cognitive problems but not extreme ones, this can result in a PVT score slightly above the probability of chance performance cut-off, which leads to a false-negative (the invalid performer's score is categorized as valid) result. Therefore, this strategy can only be successful in less than 25% of feigned cases (Rogers, 2018b).

The Portland Digit Recognition Test (PDRT) is one example of a test that applies the SBCP strategy. The PDRT consists of 72 forced-choice recognition memory items. These items are classified as easy with short delays (first 36 items) and difficult with longer delays (last 36 items). A five-digit number is presented for each trial, followed by a distractor period. During the distractor period, the examinee is required to count backward for 5 seconds (first 18 items), 15 seconds (second 18 items), and 30 seconds (last 36 items). Next, two five-digit numbers are presented, and the examinee's task is to identify which five-digit number appeared initially. The distractor period's increasing time is considered an "increasing level of item difficulty" (Binder & Willis, 1991). The cut-off score of PDRT shows moderate sensitivity for detecting invalid performance and has 100% specificity, which means some of the invalid performers can be identified correctly, and none of the valid performers are misclassified as invalid (Binder, 2003).

*Unlikely Detection Strategy (UDS)*

The UDS is considered more sophisticated than ADS (Rogers, 2018a). The principles underlying UDS are well developed and more difficult to recognize by invalid performers who intend to feign, even if they have received coaching. UDS utilizes the reported presence of highly unusual or bogus response patterns, which are not characteristic of genuine patients with cognitive impairments responses. There are three strategies: (1) the magnitude of error strategy, (2) the performance curve strategy, and (3) the violation of the learning principle (Rogers & Correa, 2008).

1. The 'magnitude of error strategy' hypothesizes that invalid performers do not consider their incorrect responses typical in actual patients who perform optimally (i.e., neurological or psychiatric patients); these mistakes are too far or too close to the correct responses. The invalid performers are frequently detectable by selecting very wrong answers and sometimes even bizarre in content; both are unlikely among actual patients who perform optimally. For instance, patients with dementia are more likely to choose answers which resemble the correct answer. The main strength of the strategy is that it is less transparent among underperformers and less vulnerable to coaching (Rogers & Correa, 2008). One example of this type of PVT is the 'b-Test.' This test is a letter-recognition task, where the examinee is asked to detect the letter 'b' by circling them. Meanwhile, some distractors (e.g., 'p', 'q', or 'd') are also interspersed with the b's. (see Figure 1). This test consists of a 15-page stimulus booklet. In each successive presentation, the array becomes smaller to impress that the task appears more difficult, although the actual difficulty is trivial. The recognition task by detecting 'b' is considered an over learned skill. This recognition ability is

relatively resistant to acquired brain injury, and the invalid performers may not be aware of this preservation (Strauss et al., 2006).

**Figure 1**
*Sample Item of The b Test*
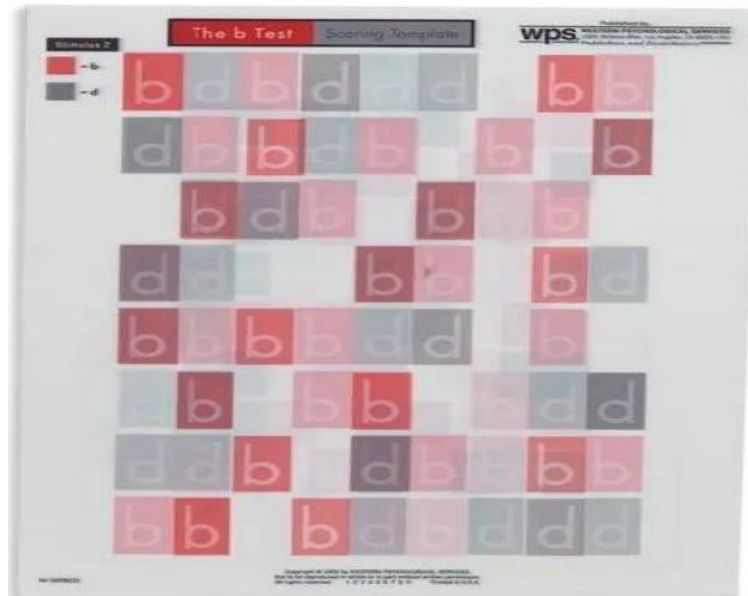


*Note* This figure is obtained from the (Western Psychological Services, 2021)

1. The performance curve strategy is based on the idea that patients with actual cognitive impairments will generate a predictable pattern of score based on item difficulty. They perform better on easy items and produce more errors on difficult items. When the performance is analyzed, results characterize a meaningful "performance curve." The invalid performers are unaware of this pattern; they typically produce less discrimination between easy and difficult items. The performance curve strategy shares similar strength as the magnitude of error strategy, in which both are resistant to coaching. Nevertheless, this strategy is quite challenging to implement in a neuropsychological test battery because it requires a wide range of item difficulties (Rogers, 2018b). Frederick's Validity Indicator Profile (VIP) is an example of a PVT that adopted this detection strategy (Rogers, 2018b). The VIP utilizes an analyzing tool which is called a Performance Curve. It is a graphical representation of the test taker's average performance compared to the item difficulty of the test. For the easiest items, it is expected that the average proportion of correct answers to be 1.0, which means all examinees can answer correctly. As the item difficulty increases, there should be a declining performance. When the examinee has reached a maximum ability, the performance should not differ significantly from the chance level (0.5). In this case, the Performance Curve can be used to examine the examinees' response style to infer the performance validity (Pearson, 2021).

2. The violation of the learning principles strategy differs from the two previous UDS based on its conceptual complexity. This strategy involves well-established learning concepts, such as performance differences between recognition vs. recall. Recognition tasks are considered to be easier compared to free recall tasks. However, invalid performers may not realize these difficulty differences because the task conditions come in one test. Invalid performers often obtain comparable scores on these conditions, although recognition tasks are easier than recall tasks. In line with the complexity of this strategy, it becomes less susceptible to coaching. In addition to its complexity, unfortunately, this strategy produces a modest difference between a simulator group and patients with a clinical diagnosis. Therefore, it is crucial to anticipate false-positive (the valid performers, usually clinical patients, are categorized as invalid due to their actual cognitive impairments) results (Rogers, 2018b).

One example of this strategy is the Non-Verbal Medical Symptom Validity Test (NV-MSVT) by Paul Green (Green, 2008). The NV-MSVT consists of seven subtests, namely: Immediate Recognition (IR), Delayed Recognition (DR), Consistency (CNS), Delayed Recognition Archetypes (DRA), Delayed Recognition Variations (DRV), Paired Associate (PA), and Free Recall (FR). The first five subtests are relatively easy, as they are recognition tasks so that perfect answers can be made. The last two subtests, PA and FR, require more memory ability to complete; therefore, they are more difficult than the first five subtests. The typical profile of healthy participants will exhibit a high score on the first five subtests, then a slight decline on the last two. People with dementia, for instance, only score slightly lower on the first five subtests than healthy participants and significantly decline on the last two subtests. A simulator group (i.e., healthy research participants asked to simulate cognitive impairments) scores significantly lower on the first five subtests compared to healthy participants; at a glance, this profile indicates severe cognitive impairments. However, they score higher than the dementia patients on the PA subtest, although this subtest is considered more difficult by the healthy participants and people with dementia. As this profile is implausible, invalid performance should be concluded. One advantage of this test is that all the stimuli are non-verbal (a universally recognizable pair of artist-drawn colored images), making this test available to examinees with low reading levels.

*Performance Validity Testing Procedures in Clinical Practice*

In clinical practice, there are two PVT procedures available, namely: (1) a PVT as a standalone test and (2) PVT indicators that are embedded in neuropsychological tests (sometimes also termed as embedded PVT).

1. A standalone PVT is an independent test designed and administered specifically to determine the examinees performance validity. It resembles a neuropsychological test and can be a paper and pencil or computer-based administration. However, not all these tests measure actual cognitive abilities (e.g., ASTM); therefore, it is expected to be insensitive to actual cognitive impairments. The previously mentioned TOMM, ASTM, and NV-MSVT are examples of standalone PVTs.

2. In contrast, an embedded PVT is a performance validity indicator derived from an existing

neuropsychological test. For instance, three embedded validity indicators can be obtained from the Digit Span subtest of the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV), namely the Reliable Digit Span (Greiffenstein et al., 1994; Zenisek et al., 2016), Longest Digit Forward – 1 Trial (LDF-1), and Longest Digit Forward – 2 Trials (Babikian et al., 2006; Kiewel et al., 2012). Thus, in addition to measuring working memory, the Digit Span can indicate performance validity using particular calculations (Greher & Wodushek, 2017).

The embedded PVT is considered cost-effective because clinical neuropsychologists are not required to administer additional tests into the neuropsychological test battery. Also, the examinee might not realize the use of PVT because this embedded PVT does not appear as a separate test. However, a problem with this procedure is that some scores that approximate the threshold indicative of cognitive impairments can be deemed invalid, a paradox called "invalid-before-impaired" (Erdodi & Lichtenstein, 2017). Unfortunately, the test manual does not provide a solution for this problem because the embedded indicators mainly developed later. As embedded PVTs measure performance validity while at the same time measuring cognitive ability, they may render a high rate of false-positive scores, especially in patients with cognitive impairments. Furthermore, this PVT procedure is also sensitive to age, which is not expected for performance validity testing. In contrast, the issue of age sensitivity does not appear in standalone PVTs, because they are constructed to measure performance validity exclusively and are not interpreted in the context of cognitive ability (Erdodi & Lichtenstein, 2017).

Research suggests that standalone PVT has superior sensitivity and specificity to embedded PVT. Sensitivity is the proportion of participants who have the condition in question (i.e., invalid performers) and score positively on the test, while specificity is the proportion of participants who do not have the condition in questions (i.e., valid performers, either healthy or clinical participants) and score negatively on the test (Dandachi-FitzGerald & Martin, 2022). A study on a mixed-clinical sample of veterans indicated that standalone PVTs discriminated against the valid and invalid performers with significant effect sizes (Critchfield et al., 2019). Furthermore, a study on aphasia patients found that a standalone PVT (TOMM) had the lowest false positive rate (6.7%) compared to three embedded PVTs, (Bodner et al., 2019). There are two possible explanations why standalone PVT is more sensitive and specific to performance validity than the embedded one. First, although the psychometric quality of an embedded indicator is quite good, its validation is frequently performed in a more restricted population compared with the standalone PVT (Greher & Wodushek, 2017). Second, the patients cognitive impairments yield a lower score on embedded PVTs since they derive from standard neuropsychological tests (Bodner et al., 2019). Table 1 summarizes the comparison between standalone and embedded PVT.

**Table 1**

*Comparison Between Embedded and Standalone PVT*

| Embedded PVT | Standalone PVT |
|---|---|
| More cost-effective because the neuropsychologist does not need to administer additional tests. | The clinical neuropsychologist should administer additional test(s), so the assessment could be longer. |
| The test form varies, including memory, attention, and verbal fluency. | Most standalone PVTs take the form of "memory tests". |
| Performance validity can be examined simultaneously with the administration of neuropsychological tests. | Clinical neuropsychologists must decide when they administer the PVTs, whether at the beginning, middle, or end of the neuropsychological testing. |
| The cut-off score lacks universally standardized and is more varied. | The cut-off score is well-established but can be adjusted in certain clinical groups. |
| Higher false positive rate in clinical samples. | Lower false positive rate in clinical samples. |
| More sensitive to age. | Less sensitive to age. |
| Example: Reliable Digit Span (RDS), Rey 15-Item Test (RFIT), Failure to Maintain Set from Wisconsin Card Sorting Test (WCST). | Example: Test of Memory Malingering (TOMM), Amsterdam Short Term Memory (ASTM), Non-Verbal Medical Symptom Validity Test (NV-MSVT). |

*Factors Affecting PVT Failure*

As previously discussed, a PVT was once termed a "malingering test", indicating that malingering is the main factor contributing to PVT failure (Greher & Wodushek, 2017). There are two main elements in malingering definition based on DSM-5 (Association, 2013): (1) symptoms exaggeration (either physical or psychological) and (2) the presence of external incentives (e.g., financial compensation from an insurance company or avoiding lawsuit), which represent that malingering is a distinct behavior. Therefore, malingering could prompt examinees to fail PVTs unless clinical neuropsychologists can verify those two elements. The complexity lies in determining that the intent of the symptom exaggeration is "caused" by those external incentives. As research and theoretical discussion developed, many factors that affect PVTs failure were identified, including when the external incentives were absent.

Although failure on PVTs is expected not to be influenced by clinical conditions, some studies indicated that some actual patients with cognitive impairments could fail PVTs, even when the intention to perform invalidly is not suspected. Loring and Goldstein (2019) report a case study of a patient with multiple sclerosis (MS) suggested that the deterioration in processing speed, which led to cognitive inefficiency and working memory impairments, contributed to the patient's PVTs failure. A study in MS patients using a bigger sample size supported the previous finding, whereas the number of PVTs failures in this group hit 14.4-21.2%. Because this study used retrospective data, the authors considered that the disability status of MS might account for this failure. Interestingly, patients who failed the PVTs also had a higher score on Beck Depression Inventory (BDI) and Beck Anxiety Inventory

(BAI), suggesting that failure on PVTs was related to overreporting of complaints (Galioto et al., 2020). More recent findings also supported that PVT failure in MS patients was not necessarily caused by intentionality. Nauta et al. (2022) found that higher disability levels, lower educational attainment, lower cognitive functioning, and being male were related to lower PVTs scores. Moreover, the authors also hypothesized that emotional or behavioral aspects, such as an expression of disease burden or feeling of distress, might induce invalid performance.

A recent meta-analysis of PVT in clinical populations revealed that patients with moderate and severe traumatic brain injury (TBI) and both mild and moderate-to-severe dementia were more likely to fail PVT compared to other neurological disorders (McWhirter et al., 2020). The PVTs failures in neurological patients were associated with the patients' cognitive or somatic symptoms, such as memory impairment, apathy, fatigue, or attention deficits, which were unintentionally measured by PVTs. Based on these findings, McWhirter et al. (2020) suggested that PVTs not only measured performance validity but also slightly captured patients' cognitive abilities or even somatic symptoms experienced by the patients. Nevertheless, they contended that true invalid performance could occur in clinical patients, either intentionally or unintentionally.

In addition, some mental states or disorders also might affect PVTs score, including severe emotional distress, depression, affective disturbance, cogniphobia (cognitive task avoidance to prevent headache pain), psychogenic nonepileptic seizures, conversion and somatoform disorder, and motivation to be in a "sick role" (Greher & Wodushek, 2017; McGuire et al., 2019; Richey & Doninger, 2020). A meta-analysis in psychotic patients indicated that the PVTs failure rates in this group were 18% (Ruiz et al., 2020). While previous studies presumed this failure was related to the psychotic patients' negative symptoms, i.e., apathy and avolition (Foussias et al., 2015; Gorissen et al., 2005; Morra et al., 2015). Ruiz et al. (2020) argued that PVTs failures in psychotic patients were associated with actual cognitive deficits mediated by low IQ. Finally, they viewed that PVTs utilization for the psychotic patient should be conducted and interpreted cautiously (Ruiz et al., 2020).

Certain personality characteristics are also related to PVT failure. A study in the military sample using the MMPI-2-RF and three PVTs (TOMM, VSVT, and WMT) demonstrated that participants who failed all PVTs were more likely to overreport on MMPI-2-RF compared to participants who passed all PVTs. This overreporting profile led to significant elevation scores on scales related to emotional and somatic/cognitive dysfunction but not on scales related to behavioral and thought dysfunction. These elevations on emotional and somatic/cognitive dysfunction-related scales should be considered in how invalid performers express their psychological conditions (Jones et al., 2012). A similar study with the MMPI-2-RF was conducted in a veteran sample but only included two PVTs (i.e., TOMM and WMT). Participants who failed both PVTs also demonstrated an overreporting profile on the MMPI-2-RF. After including a scale that detects overreporting psychopathological symptoms (i.e., the F scale) as a covariate, the invalid performers had a higher score on the Neurological Complaints-scale (a scale that represents neurological symptom items) while they scored lower on the Behavioral / Externalizing Dysfunction, Antisocial Behavior, and the Disconstraint-Revised scale. This profile suggested that the invalid performers were characterized by behavioral constraint and inhibition, more

deliberative, exercised caution before acting, and conformed to rules and norms. These personality characteristics were hypothesized as an attempt to generate a more "psychologically healthy" profile, and this impression management would make their feigned cognitive dysfunction appear genuine (Patrick & Horner, 2014).

On the contrary, PVTs failures were also encountered in people who underreported their psychological complaints. Richey and Doninger (2020) found in their disability claimant participants that participants who failed PVTs were more likely to endorse virtuous personality characteristics from the MMPI-2. Consistent with Patrick and Horner (2014) findings, Richey and Doninger (2020) agreed that the tendency to manage impressions by discrediting possible psychopathology and antisocial behavior might explain this result. Furthermore, they also concluded that this effort might be a decision-making strategy to increase the likelihood of obtaining a higher-value reward in the future (Richey & Doninger, 2020). Merckelbach et al. (2019) found that people with a high level of negative affectivity, including anxiety, were more likely to report symptoms than people with a low negative affectivity level, making them prone to experience an attentional bias. Finally, highly anxious people became more sensitive to physical or psychological (including cognitive) symptoms Merckelbach et al. (2019), a state that could induce invalid performance.

Merckelbach et al. (2019) explain that test administration order might affect the PVT score, mainly when a standalone PVT is administered as part of a battery test (e.g., PVT was either administered in the beginning, middle, or end of the test battery). The initial test may lower the threshold affirmation to a later test. Two possible reasons explaining this situation: the examinee wants to display themselves consistently instead of showing their best effort, or the earlier test engenders a bias when the examinee attempts to clarify vague items in the later test. Furthermore, symptom misinformation about the medical condition or diagnosis can lead the examinee to experience more subjective symptoms, affecting the PVTs score. (Merckelbach et al., 2019). Although it looks trivial, clinical neuropsychologists should attend to these external situations to minimize PVT failure occurrence.

Older people usually have poor medical or physical conditions, such as visual impairments, speech difficulties, or chronic health conditions that cause fatigue, making this age group susceptible to PVTs failures (McGuire et al., 2019). In addition, older people who showed a lack of interest and comprehension of the neuropsychological tests' utility might also fail PVTs (Barker et al., 2010). Davis and Millis (2014) found that lower educational attainment was associated with a greater likelihood of PVTs failures. However, it remained unclear whether people with lower educational attainment were prone to be falsely positive or invalidly performed in an obvious manner. In addition, the authors discovered that low-level functioning, including activities of daily living (e.g., dressing, bathing, etc.), medication management, financial management, driving, and occupation, was related to the higher likelihood of PVTs failure.

Other factors contributing to PVTs failures were sleep disturbance, pain, fatigue, medication side effects, sensory problems (e.g., visual disturbance), and language barrier (Greher & Wodushek, 2017). Lastly, cultural factors also play an imperative role in PVTs results. Nijdam-Jones et al. (2019) argued

that poor familiarity with the PVTs procedure in low educational attainment examinees made them prone to PVTs failures. Other cross-cultural neuropsychology experts (Ardila, 2005; Strutt & Stinson, 2022; van de Vijver & Tanzer, 2004) argued that educational level attainment itself does not fully explain the examinees' performance of neuropsychological tests, but the culture inside the educational system that induced examinees' familiarity in addressing the tests. Consequently, the cultural effect made the PVTs cut-off score should be adjusted from their original cut-off score derived from Western studies (Ryan et al., 2019).

It is important to note that not every examinee where these conditions are met would fail the PVTs. Clinical neuropsychologists must examine the examinee's history and clinical condition before interpreting the PVTs result. PVT with a performance curve strategy, such as the NV-MSVT, is preferred as this strategy uses profile analysis that is more resistant to coaching to differentiate the genuine and feigned memory impairment profiles (Laurent et al., 2019). In addition, PVTs only determine the examinee's performance validity without explicitly explaining the reasons for performance validity or invalidity (e.g., inadequate effort, deliberate exaggeration, clinical conditions, or malingering). Accordingly, it is unwise for clinicians to straightforwardly judge the cause of PVTs failures except when there is incontrovertible evidence (Schroeder & Martin, 2022). Finally, performance validity should not be considered as a binary variable with a single dimension as it is a mixture of cognitive and emotional processes (Guilmette et al., 2020; Loring & Goldstein, 2019; McWhirter et al., 2020; Ruiz et al., 2020). Interpreting PVTs failures dichotomously (valid or invalid performance) can lead to false-negatives or false-positives. Clinicians should consider the available PVTs cut-off score as a guideline for clinical decision-making and not rigidly treat the score (Bigler, 2012).

*PVT Utilization in Clinical Practice*

Clinical neuropsychologists are highly recommended to use multiple PVTs in their clinical practice. There are two explanations for why clinicians should use multiple PVTs. First, the sensitivity and specificity of PVTs vary, depending on the sample characteristics when the PVTs are psychometrically evaluated. Incorporating multiple PVTs could generate better sensitivity and specificity, finally increasing the ability to detect invalid performance while minimizing the false-positive rate. Second, examinees' performance can vary during the assessment. Therefore, the availability of multiple PVTs makes the clinicians distribute different PVT throughout the evaluation so the examinees' performance validity can be thoroughly captured (Bush et al., 2005; Heilbronner et al., 2009; Sherman et al., 2020).

It is a debatable issue what to do when clinicians face a PVT failure. When the failure is detected early during an evaluation, the clinicians should decide to continue or terminate the evaluation. Completing the evaluation when PVT failure is evident may give additional evidence of the examinee's consistency and minimum ability levels, but on the other side, more time and resources should be invested. If the clinicians decide to terminate the evaluation early, the potential problem is that the examinee may realize which test is used to determine performance validity, and it can reduce the PVT's utility for future evaluation. Thus, in determining whether to continue or terminate the evaluation, the

clinicians should consider the goal and context of the evaluation (Carone et al., 2010).

PVTs failures are associated with more subjective cognitive complaints and worse neuropsychological test results (Votruba et al., 2020). If the neuropsychological test results indicate that the examinee's cognitive abilities lie on an average (or higher) level while simultaneously failing the PVT, these results should be considered his or her basal level of cognitive ability. In other words, the test results underestimate his or her actual cognitive functioning. Clinicians should be cautious regarding the accuracy of the patient's symptoms when PVTs failures are identified. As previously mentioned, a malingering diagnosis can be made only when irrefutable evidence indicates that the PVTs failures are caused by external incentives (Greher & Wodushek, 2017).

After the PVTs interpretation is concluded, the next step is communicating the finding to the examinee in a feedback session. This step can be challenging as it may create a significant interpersonal conflict with the examinee, making the examinee and clinician uncomfortable. Clinicians may ask the examinee to think about his or her test performance so there is an acknowledgment of the examinee's perspective. Then, the clinicians may tell that all people have strengths and weaknesses; therefore, getting some low scores are relatively common. The clinicians can tell which test(s) the examinee obtains a relatively high score (if it applies) and low scores. The clinician might associate this low score with problems in "staying motivated" or "cognitive effort appear to vary a lot during the testing." It should be emphasized that the conclusion is reached based on objective data, without informing the specific tests or methods used, so the examinee understands the conclusion is not based on subjective impression. Next, the clinicians inform that non-neurological factors interfere with the examinee's performance and do not reflect his or her actual ability. Finally, the clinician can convey the "good news" that poor test scores can be improved when addressing the non-neurological factors (Carone et al., 2010; Martin & Schroeder, 2022).

*Research in PVT*

Four research designs are commonly used in PVT research: (1) simulation research (asking the healthy participants to mimic patients with cognitive impairments), (2) known-groups comparison (using validated multiple PVTs to accurately classify known groups for validating a new PVT), (3) differential prevalence design (comparing validated PVTs in a group with assumed incentives [e.g., litigation] and a group without incentives), and (4) partial criterion design (validating a new PVT in a similar way as the known-groups comparison, except for the lower accuracy classification of the known groups (Rogers, 2018b). As no validated PVT is available, we argue that simulation research is the most appropriate to be conducted in Indonesia because the latter three research designs require validated and available PVTs.

Although many PVTs are available outside Indonesia, those tests cannot be used immediately in Indonesia as they use cut-off scores derived from normative datasets of a particular culture. Therefore, the PVTs cut-off score should be adjusted based on the Indonesian context by performing several studies, such as simulation research and PVTs administration to clinical patients (i.e., neurological and psychiatric patients). The cut-off score determination should balance sensitivity and specificity, as

increasing one will decrease the other. However, in PVT studies, researchers weigh more on specificity because making a false-positive decision leads to a more detrimental effect (i.e., a genuine patient is concluded to perform invalidly and finally does not receive the treatment). It is recommended that specificity of .90 is a stringent point (Davis & Millis, 2014; Larrabee et al., 2019). Cut-off scores from a simulation study should be evaluated in clinical patients because some exhibit small to moderate failure rates. Therefore, such studies can develop cut-off scores specific to clinical patients.

## Conclusion

Various neuropsychological tests have been developed and adapted in the last decade in Indonesia (Suwartono et al., 2014; Wahyuningrum et al., 2021). Unfortunately, it has not been accompanied yet by the development of PVTs. As suggested by the NAN and AACN, PVTs are required to ensure the validity of neuropsychological test results, so interpretation can be made accurately, which leads to appropriate treatments and recommendations in clinical practice. More importantly, multiple PVTs should be utilized to minimize the risk of false-positives; on the other hand, it also increases sensitivity (Bush et al., 2005; Heilbronner et al., 2009; Sherman et al., 2020). However, in Indonesia, there is a lack of PVTs that are validated psychometrically. Therefore, the development of multiple PVTs in Indonesia is necessary.

In our project, we plan to address this urge by adapting multiple PVTs in Indonesia: two standalone PVTs (i.e., Indonesian version of ASTM and NV-MSVT) and two embedded PVTs derived from Digit Span subtest of WAIS-IV-ID (i.e., RDS, LDF-1, and LDF-2). The adaptation will be conducted using a simulation research design and validating the PVTs in clinical patients. Regardless of our project, research using other designs would further validate these PVTs, or even validate other new PVTs.

*Recommendation*

When the PVTs are available in Indonesia, all clinicians who require PVT should be cautious when employing this test. The selection of PVTs should be evidence-based and acknowledge the patient's (differential) diagnosis and disorder severity. Failure to take these into account can increase the probability of false-positive detection (Greher & Wodushek, 2017; Laurent et al., 2019; Strutt & Stinson, 2022). The interpretation of PVTs failures should be made cautiously, as many factors can affect them. It is discouraged to interpret the PVT in a binary way (i.e., pass or not pass) without acknowledging the factors behind the failure. Also, clinical neuropsychologists should weigh the utilization and interpretation more on PVTs with higher specificity (lower false-positive rate) and good sensitivity when they discover inconsistent findings between PVTs (Bodner et al., 2019). Eventually, it will become the clinician's responsibility to make the PVTs resistant to coaching by maintaining the PVTs' confidentiality.

## Declarations

*Author Contribution*

The authors have contributed equally in the process of writing this article. Conceptualization: WA, MPPH, MSH, and RPCK; Methodology: WA and MPPH; Supervision: MPPH and MSH; Writing original draft preparation: WA; Writing, review, and editing: WA, MPPH, and RPCK.

*Conflict of Interest*

The authors declare no conflict of interest.

*Orcid ID*

Widhi Adhiatama ⓘ https://orcid.org/0000-0002-0052-0183
Marc P. H. Hendriks ⓘ https://orcid.org/0000-0002-5321-7970
Magdalena S Halim ⓘ https://orcid.org/0000-0003-2158-7587
Roy P. C. Kessels ⓘ https://orcid.org/0000-0001-9500-9793

## References

Adhiatma, W., & Halim, M. S. (2019). Exploratory factor analysis of the Indonesian version of MMPI-2-RF Restructured Clinical (RC) Scales. *HUMANITAS: Indonesian Psychological Journal*, *16*(1), 66–73. https://doi.org/10.26555/humanitas.v16i1.9420

Ardila, A. (2005). Cultural values underlying psychometric cognitive testing. *Neuropsychology Review*, *15*(4), 185–195. https://doi.org/10.1007/s11065-005-9180-y

Association, A. P. (2013). *Diagnostic and statistical manual of mental disorder: DSM-5*. American Psychiatric Publishing, Inc.

Babikian, T., Boone, K. B., Lu, P., & Arnold, G. (2006). Sensitivity and specificity of various digit span scores in the detection of suspect effort. *The Clinical Neuropsychologist*, *20*(1), 145–159. https://doi.org/10.1080/13854040590947362

Barker, M. D., Horner, M. D., & Bachman, D. L. (2010). Embedded indices of effort in the repeatable battery for the assessment of neuropsychological status (RBANS) in a geriatric sample. *The Clinical Neuropsychologist*, *24*(6), 1064–1077. https://doi.org/10.1080/13854046.2010.486009

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*, 632–642. https : / / doi . org / 10 . 1017 / S1355617712000252

Binder, L. M. (2003). The Portland Digit Recognition Test: A review of validation data and clinical use. *Journal of Forensic Neuropsychology*, 2(3-4), 27–41. https://doi.org/10.1300/J151v02n03_02

Binder, L. M., & Willis, S. C. (1991). Assessment of motivation after financially compensable minor head trauma. *Psychological Assessment*, *3*(2), 175–181. https://doi.org/10.1037/1040-3590.3.2.175

Binder, L. M., Larrabee, G. J., & Millis, S. R. (2014). Intent to fail: Significance testing of forced choice test results. *The Clinical Neuropsychologist*, *28*(8), 1366–1375. https://doi.org/10.1080/13854046.2014.978383

Bodner, T., Merten, T., & Benke, T. (2019). Performance validity measures in clinical patients with aphasia. *Journal of Clinical and Experimental Neuropsychology*, *41*(5), 476–483. https://doi.org/10.1080/13803395.2019.1579783

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C. R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy Planning Committee. *Archives of Clinical Neuropsychology*, *20*(4), 419–426. https://doi.org/10.1016/j.acn.2005.02.002

Butcher, J. N. (2005). *A beginner's guide to the MMPI-2* (2nd ed.). American Psychological Association.

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

Critchfield, E., Soble, J. R., Marceaux, J. C., Bain, K. M., Bailey, K. C., Webber, T. A., Alverson, W. A., Messerly, J., González, D. A., & O'Rourke, J. J. (2019). Cognitive impairment does not cause invalid performance: Analyzing performance patterns among cognitively unimpaired, impaired, and noncredible participants across six performance validity tests. *The Clinical Neuropsychologist*, *33*(6), 1083–1101. https://doi.org/10.1080/13854046.2018.1508615

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

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

Davis, J. J., & Millis, S. R. (2014). Examination of performance validity test failure in relation to number of tests administered. *The Clinical Neuropsychologist*, *28*(2), 199–214. https://doi.org/10.1080/13854046.2014.884633

Dewiyana, Satiti, S., & Gofir, A. (2019). Penggunaan HIV dementia scale dan international HIV dementia scale pada skrining demensia [Use of the HIV dementia scale and the international HIV dementia scale in dementia screening]. *Berkala Neurosains*, *18*(2), 84–89.

Erdodi, L. A., & Lichtenstein, J. D. (2017). Invalid before impaired: an emerging paradox of embedded validity indicators. *The Clinical Neuropsychologist*, *31*(6-7), 1029–1046. https://doi.org/10.1080/13854046.2017.1323119

Foussias, G., Siddiqui, I., Fervaha, G., Mann, S., McDonald, K., Agid, O., Zakzanis, K. K., & Remington, G. (2015). Motivated to do well: An examination of the relationships between motivation, effort, and cognitive performance in schizophrenia. *Schizophrenia Research*, *166*(1-3), 276–282. https://doi.org/10.1016/j.schres.2015.05.019

Galioto, R., Dhima, K., Berenholz, O., & Busch, R. (2020). Performance validity testing in multiple sclerosis. *Journal of the International Neuropsychological Society*, *26*(10), 1028–1035. https://doi.org/10.1017/S1355617720000466

Gorissen, M., Sanz, J. C., & Schmand, B. (2005). Effort and cognition in schizophrenia patients. *Schizophrenia Research*, *78*(2-3), 199–208. https://doi.org/10.1016/j.schres.2005.02.016

Green, P. (2008). *Green's Non-Verbal Medical Symptom Validity Test (NV-MSVT): User's manual for Microsoft Windows*. Edmonton, Green's Publishing Inc.

Greher, M. R., & Wodushek, T. R. (2017). Performance validity testing in neuropsychology: Scientific basis and clinical application - A brief review. *Journal of Psychiatric Practice*, *23*(2), 134–140. https://doi.org/10.1097/PRA.0000000000000218

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures with a large clinical sample. *Psychological Assessment*, *6*(3), 218–224. https://doi.org/10.1037//1040-3590.6.3.218

Guilmette, T. J. (2012). The role of clinical judgment in symptom validity assessment. In D. A. Carone & S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 31–43). New York, Springer Publishing Company.

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

Hathaway, S. R., & McKinley, J. C. (1940). A multiphasic personality schedule (Minnesota): I. Construction of the schedule. *Journal of Personality*, *10*, 249–254.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist*, *23*(7), 1093–1129. https://doi.org/10.1080/13854040903155063

Husein, N., Lumempouw, S. F., Ramli, Y., & Herqutanto. (2010). Uji validitas dan reliabilitas Montreal Cognitive Assessment versi Indonesia (MoCA-Ina) untuk skrining gangguan kognitif. *Neurona*, *27*(4).

Indrajaya, A. W., Lumempouw, S. F., Ramli, Y., & Prihartono, J. (2013). Nilai normal pemeriksaan neuropsikologi CERAD di Jakarta [Normal values for CERAD neuropsychological examination in Jakarta]. *Neurona*, *30*(3).

Jones, A., Ingram, M. V., & Ben-Porath, Y. S. (2012). Scores on the MMPI-2-RF scales as a function of increasing levels of failure on cognitive symptom validity tests in a military sample. *The Clinical Neuropsychologist*, *26*(5), 790–815. https://doi.org/10.1080/13854046.2012.693202

Kessels, R. P. C., & Hendriks, M. P. H. (2022). Neuropsychological assessment. In H. S. Friedman (Ed.), *Encyclopedia of mental health*. Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-323-91497-0.00017-5

Khairunnisa, G., Putri, P., Cheerson, F., Junita, F., Suwartono, C., & Halim, M. S. (2014). Uji validitas konstruk The Modified Mini Mental State-Test (3MS) [Modified Mini Mental State-Test (3MS) construct validity test]. *Jurnal Pengukuran Psikologi dan Pendidikan Indonesia*, *3*(4), 329–359.

Kiewel, N. A., Wisdom, N. M., Bradshaw, M. R., Pastorek, N. J., & Strutt, A. M. (2012). A retrospective review of Digit span-related effort indicators in probable Alzheimer's disease patients. *The Clinical Neuropsychologist*, *26*(6), 965–974. https://doi.org/10.1080/13854046.2012.694478

Kolb, B., & Whishaw, I. Q. (2015). *Fundamentals of human neuropsychology*. Worth Publishers.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 625–631. https://doi.org/10.1017/S1355617712000240

Larrabee, G. J. (2015). The multiple validities of neuropsychological assessment. *American Psychologist*, *70*(8), 779–788. https://doi.org/10.1037/a0039835

Larrabee, G. J., Millis, S. R., & Meyers, J. E. (2009). 40 plus or minus 10, a new magical number: Reply to Russell. *The Clinical Neuropsychologist*, *23*(5), 841–849. https://doi.org/10.1080/13854040902796735

Larrabee, G. J., Rohling, M. L., & Meyers, J. E. (2019). Use of multiple performance and symptom validity measures: Determining the optimal per test cutoff for determination of invalidity, analysis of skew, and inter-test correlations in valid and invalid performance groups. *The Clinical Neuropsychologist*, *33*(8), 1354–1372. https://doi.org/10.1080/13854046.2019.1614227

Laurent, R. L., Whiteside, D. M., & Basso, M. R. (2019). Performance validity testing in an older adult population: Considerations for clinical practice. In L. D. Ravdin & H. L. Katzen (Eds.), *Handbook on the neuropsychology of aging and dementia* (2nd ed., pp. 77–88). Cham, Springer Nature Switzerland AG. https://doi.org/10.1007/978-1-4614-3106-0

Loring, D. W., & Goldstein, F. C. (2019). If invalid PVT scores are obtained, can valid neuropsychological profiles be believed? *Archives of Clinical Neuropsychology*, *34*(7), 1192–1202. https://doi.org/10.1093/arclin/acz028

Martin, P. K., & Schroeder, R. W. (2020). Base rates of invalid test performance across clinical non-forensic contexts and settings. *Archives of Clinical Neuropsychology*, *35*(6), 717–725. https://doi.org/10.1093/arclin/acaa017

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

Martin, P. K., Schroeder, R. W., Olsen, D. H., Maloy, H., Boettcher, A., Ernst, N., & Okut, H. (2020). A systematic review and meta-analysis of the Test of Memory Malingering in adults: Two decades of deception detection. *The Clinical Neuropsychologist*, *34*(1), 88–119. https://doi.org/10.1080/13854046.2019.1637027

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*, 114–131. https://doi.org/10.1093/arclin/acy012

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations - A systematic review. *Journal of Neurology, Neurosurgery and Psychiatry*, *91*, 945–952. https://doi.org/10.1136/jnnp-2020-323776

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

Morra, L. F., Gold, J. M., Sullivan, S. K., & Strauss, G. P. (2015). Predictors of neuropsychological effort test performance in schizophrenia. *Schizophrenia research*, *162*, 205–210. https://doi.org/10.1016/j.schres.2014.12.033

Nauta, I. M., Bertens, D., van Dam, M., Huiskamp, M., Driessen, S., Geurts, J. J., Uitdehaag, B. M., Fasotti, L., Hulst, H. E., de Jong, B. A., & Klein, M. (2022). Performance validity in outpatients with multiple sclerosis and cognitive complaints. *Multiple Sclerosis Journal*, *28*(4), 642–653. https://doi.org/10.1177/13524585211025780

Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2019). The effect of literacy and culture on cognitive effort test performance: An examination of the Test of Memory Malingering in Colombia. *Journal of Clinical and Experimental Neuropsychology*, *41*(10), 1015–1023. https://doi.org/10.1080/13803395.2019.1644294

Patrick, R. E., & Horner, M. D. (2014). Psychological characteristics of individuals who put forth inadequate cognitive effort in a secondary gain context. *Archives of Clinical Neuropsychology*, *29*(8), 754–766. https://doi.org/10.1093/arclin/acu054

Pearson. (2021). Validity Indicator Profile (VIP). Retrieved July 13, 2021, from https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Validity-Indicator-Profile/p/100000284.html?tab=faqs

Rambe, A. S. (2008). Neuropsychological evaluation against geriatric background. *Majalah Kedokteran Nusantara*, *41*(4).

Richey, L. N., & Doninger, N. A. (2020). A comparison of performance validity measures in predicting MMPI-2 lie scale results. *The Clinical Neuropsychologist*, *34*(2), 353–367. https://doi.org/10.1080/13854046.2019.1643922

Rogers, R. (2018a). Detection strategies for malingering and defensiveness. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (4th ed., pp. 18–41). New York, The Guilford Press.

Rogers, R. (2018b). An introduction to response styles. In R. Rogers & S. D. Bender (Eds.), *Clinical assessment of malingering and deception* (pp. 3–17). New York, The Guilford Press.

Rogers, R., & Correa, A. A. (2008). Determinations of malingering: Evolution from case-based methods to detection strategies. *Psychiatry, Psychology and Law*, *15*(2), 213–223. https://doi.org/10.1080/13218710802014501

Ruiz, I., Raugh, I. M., Bartolomeo, L. A., & Strauss, G. P. (2020). A Meta-analysis of neuropsychological effort test performance in psychotic disorders. *Neuropsychology Review*, *30*(3), 407–424. https://doi.org/10.1007/s11065-020-09448-2

Ryan, J. J., Yamaguchi, T., & Kreiner, D. S. (2019). Preliminary validation of the rey 15-item test and reliable digit span in native japanese samples. *Psychological Reports*, *122*(5), 1925–1945. https://doi.org/10.1177/0033294118792697

Schagen, S., Schmand, B., De Sterke, S., & Lindeboom, J. (1997). Amsterdam short-term memory test: A new procedure for the detection of feigned memory deficits. *Journal of Clinical and Experimental Neuropsychology*, *19*(1), 43–51. https://doi.org/10.1080/01688639708403835

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

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

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A Compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.

Strutt, A. M., & Stinson, J. M. (2022). Performance validity testing with culturally diverse individuals and non-native English speakers: The need for a cultural perspective in neuropsychological practice. In R. W. Schroeder & P. K. Martin (Eds.), *Validity assessment in clinical neuropsychological practice: Evaluating and managing noncredible performance* (pp. 211–232). New York, The Guilford Press.

Sulastri, A., Utami, M. S. S., Jongsma, M., Hendriks, M., & van Luijtelaar, G. (2019). The Indonesian Boston Naming Test: Normative data among healthy adults and effects of age and education on naming ability. *International Journal of Science and Research*, *8*(11), 134–139. https://doi.org/10.21275/ART20202307

Suwartono, C., Halim, M. S., Hidajat, L. L., Hendriks, M. P. H., & Kessels, R. P. C. (2014). Development and reliability of the Indonesian Wechsler Adult Intelligence Scale—Fourth Edition (WAIS-IV). *Psychology*, *05*(14), 1611–1619. https://doi.org/10.4236/psych.2014.514171

Syahroni, Lumempouw, S. F., Lastri, D. N., & Herqutanto. (2011). Validasi transkultural dan uji reliabilitas Screening Test for Luria Nebraska Neuropsychological Battery (ST-LNNB) untuk penapisan gangguan kognitif [Transcultural validation and reliability testing of the Screening Test for Luria Nebraska Neuropsychological Battery (ST-LNNB) for cognitive impairment screening]. *Neurona*, *28*(2).

Tombaugh, T. N. (1996). *Test of Memory Malingering (TOMM)*. Multi-Health System, Inc.

Utami, M. S. S., Sulastri, A., & Guritno, H. (2019). Rey auditory verbal learning test of university students. *Psikodimensia*, *18*(1), 37–48. https://doi.org/10.24167/psidim.v18i1.1740

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *Revue europeenne de psychologie appliquee*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Votruba, K. L., Rykulski, N., Dumitrescu, C., & Abeare, C. A. (2020). Handedness and performance validity test performance. *Psychology and Neuroscience*, *13*(2), 196–205. https://doi.org/10.1037/pne0000188

Wahyuningrum, S. E., van Luijtelaar, G., & Sulastri, A. (2021). An online platform and a dynamic database for neuropsychological assessment in Indonesia. *Applied Neuropsychology:Adult*, 1–10. https://doi.org/10.1080/23279095.2021.1943397

Western Psychological Services. (2021). The b Test. Retrieved June 15, 2021, from https://www.wpspublish.com/the-b-test

Widyarini, N., Lumempouw, S. F., Ramli, Y., & Herqutanto. (2013). Nilai normal pemeriksaan Status Mental Neurologi Strub dan Black versi Indonesia [The normal value of the Indonesian version of Strub and Black Mental Neurology Status examination]. *Neurona*, *30*(3).

Zenisek, R., Millis, S. R., Banks, S. J., & Miller, J. B. (2016). Prevalence of below-criterion Reliable Digit Span scores in a clinical sample of older adults. *Archives of Clinical Neuropsychology*, *31*(5), 426–433. https://doi.org/10.1093/arclin/acw025