Jurnal Psikologi | E-ISSN: 2302-1098
2025, Vol. 24, No. 2, 197-207
https://doi.org/10.14710/jp.24.2.197-207

Evaluating PVTs diagnostic accuracy through simulation study in Indonesian older adults

Alfo Yanuar[1], Widhi Adhiatma[1*], Marc P. H. Hendriks[2], and Magdalena S. Halim[1]
[1]Department of Psychology, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
[2]Department of Neuropsychology and Rehabilitation Psychology, Faculty of Social Sciences, Radboud University, Gelderland and North Brabant, The Netherlands

ABSTRACT
Background: The accurate assessment of cognitive ability requires valid data, particularly among older adults. Non-credible performance, or symptom exaggeration, threatens the validity of neuropsychological assessments in Indonesia.
Purpose: This study aims to examine the diagnostic accuracy of four Performance Validity Tests (PVTs) by comparing healthy older adults and individuals simulating cognitive impairment.
Method: An experimental two-group design was used (N=25). Participants passed rigorous cognitive screening (MoCA > 26) and were randomly assigned to a control (n=12) or experimental (n=13) group. PVTs included NV-MSVT, TMJPI, CIHT, and embedded TMT measures.
Findings: Mann-Whitney U tests showed significant differences ($p < .05$) between groups, with controls outperforming simulators. ROC analysis yielded optimal cut-off scores for all PVTs: NV-MSVT A1 (≤ 83), NV-MSVT A2 (≤ 79), TMJPI (≤ 87), CIHT (≤ 9), TMT-A (≥ 131), and TMT-B (≥ 200). AUC values ranged from .713 to .907, indicating good to excellent accuracy in distinguishing the groups.
Implication: These findings establish a robust empirical foundation for more ethical and valid neuropsychological research in Indonesia, providing culturally relevant cut-offs to mitigate misdiagnosis in clinical practice, advance cross-cultural neuropsychology in Asia, and strengthen administrative and forensic decision-making.

KEYWORDS
PVT; performance validity; neuropsychology; older adults; diagnostic accuracy

ARTICLE HISTORY
Received 28 August 2024
Revised 31 December 2024
Accepted 17 May 2025

Introduction

Cognitive decline is a common issue in the elderly, typically beginning around age 60 (Indonesia Statistical Bureau, 2022). Dementia has far-reaching consequences for families and caregivers, causing emotional stress, social isolation, and disruptions in daily life (Lindeza et al., 2020; Merrilees, 2016; Shin & Kim, 2022; Tran et al., 2025). Accurate assessment, diagnosis, and intervention are required to address these issues. Comprehensive medical histories, physical examinations, and cognitive assessments, such as the Montreal Cognitive Assessment-MoCA and Mini-Mental State Examination-MMSE (Nasreddine et al., 2005), are commonly used to screen for cognitive impairments (Alzheimer's Association, n.d.; National Institute on Aging, 2022; National Health Service, 2023).

Neuropsychological tests are crucial in this diagnostic pathway because they provide objective insights into cognitive function and behavior (Galioto et al., 2023; Pesau & Luijtelaar, 2021; Prasetya & Susanto, 2024; Sulastri et al., 2018; Wahyuningrum et al., 2021; Wahyuningrum et al., 2022; Widhianingtanti et al., 2022). Despite their usefulness, these assessments often face challenges in distinguishing valid from invalid performance, such as intentional deception or poor effort, which can result in diagnostic inaccuracies (Bass & Wade, 2019; Dandachi-Fitzgerald et al., 2017; Guilmette et al., 2020). To mitigate these risks, the performance validity test (PVT) has emerged as an important strategy.

CONTACT Widhi Adhiatma widhi.adhiatma@atmajaya.ac.id Jl. Jend. Sudirman, Jakarta, Indonesia

The scientific foundation of PVTs is grounded in effort and response bias (McWhirter et al., 2020). These measures are specifically constructed to leverage the expected performance disparity between individuals exhibiting genuine impairment and those engaging in non-credible performance. Fundamentally, PVTs present tasks that are sufficiently simple to be passed by individuals with genuine cognitive deficits, provided they exert full effort. Conversely, examinees attempting to simulate impairment often fail these items. This pattern, known as the malingering bias (McWhirter et al., 2020), results from overestimating the degree of impairment that would affect performance on basic tasks. Therefore, PVTs function as essential gatekeepers, ensuring the validity and interpretability of subsequent data within the neuropsychological assessment battery (Schroeder & Martin, 2022). Applying and validating this theoretical framework in diverse populations, such as in the Indonesian context, is critical to maintaining assessment standards globally (Adhiatma et al., 2025).

PVTs are cognitive tests used to determine the validity of cognitive performance, distinguishing between valid and invalid responses (Larrabee, 2012; McWhirter et al., 2020; Soble, 2021). These tests are essential in neuropsychological evaluations to ensure the accuracy of test results, which is critical for a reliable diagnosis (Lippa, 2018; Ord et al., 2021). However, PVT scores may be influenced by various factors, including cognitive impairments due to neurological disorders, psychiatric disorders, or a patient's condition (Greher & Wodushek, 2017; McWhirter et al., 2020).

Unfortunately, the utilization of PVTs in Indonesia remains limited. Research on PVTs has only recently emerged, and the number of tests that have been systematically studied is still small (Adhiatma, 2024). This issue highlights a critical gap in improving diagnostic accuracy, particularly among the rapidly growing older population (Indonesia Statistical Bureau, 2022). As Indonesia's demographics shift toward an aging population, there is a pressing need to validate PVTs specifically for this group. While PVTs are primarily designed to distinguish between valid and invalid cognitive performance, their use in neuropsychological evaluations can enhance diagnostic accuracy and support neuropsychological intervention by reducing the risk of misdiagnosis in healthy older adults.

Adhiatma (2024) has made significant advances in customizing PVTs for the Indonesian setting. His study presents culturally appropriate tests such as the *Tes Memori Jangka Pendek Indonesia* (TMJPI), the Non-Verbal Symptom Validity Test (NV-MSVT), and the embedded measures of Digit Span, demonstrating their specificity in clinical settings (Adhiatma, 2024).

Furthermore, additional research is necessary to validate these various PVTs across diverse populations, including healthy older adults, establish normative data, and optimize the clinical utility of PVTs. In this study, we examined the diagnostic accuracy of multiple PVTs, namely the NV-MSVT, TMJPI, Coin-in-the-Hand Test (CIHT), and the embedded validity measures of the Trail Making Test (TMT), in a sample of healthy older adults. We hypothesized that these PVTs would accurately distinguish between genuine participants and those who were simulating cognitive impairment.

Method

This study employed an experimental two-group design (Creswell & Creswell, 2018), comparing PVT scores between healthy older adults and those simulating dementia. Participants were randomly assigned to either a control group, who were instructed to perform the tests to their best abilities, or an experimental group, who were instructed to simulate dementia in completing the PVTs.

Participants

Initially, we recruited 80 older adult participants (aged 60 years or older). As we only involved healthy older adults (MoCA score above 26), we excluded 55 participants whose MoCA score indicated cognitive impairments (MoCA score below 26) (Nasreddine et al., 2005). Therefore, the total number of participants in this study was 25 (M = 70.4, SD = 7.4, range = 60–85 years old). Participants were recruited from several nursing homes and individual sources, including Panti Sosial Tresna Werdha Budi Mulia 2 Cengkareng, Panti Sosial Tresna Werdha Budi Mulia 2 Jelambar, and Panti Werdha Wisma Mulia.

Instruments

The study employed several PVTs. First, the Non-Verbal Medical Symptom Validity Test (NV-MSVT; α = .917) uses computerized, non-verbal colored image stimuli. This PVT uses two criteria to determine performance validity: criteria A1 and A2 (Green, 2008). Second, the Coin-in-the-Hand Test (CIHT; α = .686) (Kapur, 1994) consists of ten trials where participants must identify which of the test administrator's hands (i.e., left hand or right hand) holds a coin after a brief distraction. Third, the Indonesian Short-Term Memory Test (*Tes Memori Jangka Pendek Indonesia* - TMJPI; α = .943) (Adhiatma, 2024), a PVT that was developed based on the Amsterdam Short-Term Memory Test (Schagen et al., 1997). This PVT consists of 30 items where participants are required to recognize words after performing a simple arithmetic task as a distractor. Fourth, the embedded validity measures of the Trail Making Test (TMT). The TMT (α = .779) contains two tasks, namely TMT-A and TMT-B. In both tasks, the participants should connect letters for the TMT-A and letters and numbers for the TMT-B, and then the time completion (in seconds) is recorded (Bowie & Harvey, 2006).

Research Procedures

Firstly, participants underwent cognitive screening using the MoCA to exclude those with cognitive impairments. Those who met the inclusion criteria were then randomly assigned to either the experimental group (n = 13), instructed to feign symptoms, or the control group (n = 12), instructed to perform to their best ability. The experimental group participants were instructed to feign symptoms during test administration, while the control group participants were instructed to complete the tests to their best ability. Subsequently, the TMJPI, NV-MSVT, TMT, and CIHT were administered. Following the assessments, participants underwent a debriefing session to provide feedback on their experience. Finally, the collected data were organized and analyzed. This study has gained approval from the Atma Jaya Catholic University of Indonesia Research Ethics Commission (Number: 0011L/III/PPPE.PM.10.05/12/2023, issued on December 18, 2023).

Data Analysis

The data were analyzed using Jamovi version 2.5 (The Jamovi Project, 2024). Statistical analyses included descriptive statistics, inferential statistics (Mann-Whitney U Test, due to non-normally distributed scores [p<.05] in all PVTs, except for the TMT-B), and receiver operating characteristic (ROC) curve. The ROC curve resulted in several cut-off scores and their corresponding sensitivity and specificity (Šimundić, 2009). We also calculated the area under the curve (AUC), which indicates the overall accuracy or discriminative ability of the tests. An AUC value of .50 indicates no discrimination ability, .70 to .80 indicates acceptable, .81 to .90 indicates excellent, and >.90 is outstanding accuracy in discriminating true positives and true negatives (Daugherty et al., 2021; Ord et al., 2021; Unal, 2017; Šimundić, 2009).

Results

Based on the normality test, it is observed that the data distribution for all of NV-MSVT's criteria (A1 and A2), as well as for TMJPI, CIHT, and TMT-A, was found to be non-normal. However, the data distribution for TMT-B was normal (Shapiro-Wilk p = .375 > .05). Given these distribution characteristics**,** comparisons between the control and experimental groups were conducted using a non-parametric approach.

Table 1.

Comparison Between Control and Experiment Groups Across PVTs

Test/Index	Mean		Mann-Whitney U Test	
	Control Group	Experiment Group	Statistics	*p*
NV-MSVT A1	92.67	76.54	23	.001
NV-MSVT A2	63.25	52	21	.001
TMJPI	88.67	77.23	15	<.001
CIHT	9.92	7.62	14.5	<.001
TMT – A	61.42	109.08	40	.982
TMT – B	118.58	168.77	45	.966

Note: NV-MSVT = Non-Verbal Medical Symptom Validity Test; TMJPI = Tes Memori Jangka Pendek Indonesia; CIHT = Coin in the Hand Test; TMT = Trail Making Test.

Table 1 showed that the control group reached the maximum score in several PVTs, namely NV-MSVT, TMJPI, and CIHT, indicating performance near the ceiling level. In addition, the Mann-Whitney U tests indicated that the control group scored significantly higher than the experimental group in all PVTs (p < .05), except for the TMT-A and TMT-B, where no significant differences were found**.**

Table 2.

Level of Accuracy of Tests with AUC Values

Test	AUC Score	Level of Accuracy
TMJPI	.904	Outstanding
CIHT	.907	Outstanding
TMT – A	.744	Acceptable
TMT – B	.713	Acceptable
NV-MSVT A1	.853	Excellent
NV-MSVT A2	.865	Excellent

Note: NV-MSVT = Non-Verbal Medical Symptom Validity Test; TMJPI = Tes Memori Jangka Pendek Indonesia; CIHT = Coin in the Hand Test; TMT = Trail Making Test.

The AUC values (see Table 2) demonstrated that the evaluated PVTs had diagnostic accuracy ranging from acceptable (TMT) to outstanding (TMJPI and CIHT), with a range of .713–.907**,** indicating their adequate performance in discriminating between valid and invalid performances.

Table 3.
Cut-off Scores, Sensitivity, and Specificity of the PVTs

Cut-off Scores	Sensitivity	Specificity
NV-MSVT A1 < 84	.69	.91
NV-MSVT A2 < 80	.69	.91
TMJPI < 88	.84	.91
CIHT < 10	.84	.91
TMT – A > 132	.38	.91
TMT – B > 201	.38	.91

Note: NV-MSVT = Non-Verbal Medical Symptom Validity Test; TMJPI = Tes Memori Jangka Pendek Indonesia; CIHT = Coin in the Hand Test; TMT = Trail Making Test.

Table 3 showed that the specificity of the NV-MSVT (Criteria A1 and A2), TMJPI, CIHT, and TMT was higher than .90. However, the sensitivity of the embedded validity measures within the TMT was the lowest, indicating that their cut-off scores produced the highest false-negative rates among the four PVTs.

Discussion

This study aimed to investigate the diagnostic accuracy of multiple PVTs in healthy elderly by examining their AUC to determine the overall diagnostic accuracy and determine the optimal cut-off score for each PVT. Before the diagnostic accuracy evaluation, we compared the PVT score between the control group (consisting of healthy elderly performing at their best ability) and the experimental group (consisting of healthy elderly simulating dementia symptoms).

Our findings showed a significant difference between the control and experimental groups, where the control group scored higher than the experimental group. However, these results did not apply to the TMTs. The insignificant difference in the TMTs indicates that both participants in the control and experimental group performed 'slow.' This finding is supported by Widhianingtanti et al. (2022) findings, where the elderly tend to have slower performance times compared to younger individuals in the TMT-B. Furthermore, slower completion time in the TMT-B among the elderly is also documented in other studies (MacPherson et al., 2017). Meanwhile, in our study, the control group performed slowly due to the need to perform at their best; therefore, they worked very carefully. On the other hand, the experimental group, as part of their comprehension of the instructions, needed to work slowly and make mistakes.

Our findings in the NV-MSVT were consistent with a large body of literature and meta-analyses, which show that stand-alone PVTs maintain satisfactory specificity rates across comparisons involving healthy individuals, those with genuine cognitive impairments, and simulators (Roor et al., 2023; Adhiatma, 2024). This means that the true-negative rates in healthy elderly could be maintained at a low level. However, it should be noted that our findings indicated the NV-MSVT had lower sensitivity rates compared to those of other studies.

This discrepancy can be explained by several factors. First, stand-alone PVT tends to have lower sensitivity and higher specificity to minimize misclassification of valid and invalid responses (Roor et al., 2023). Second, most studies use clinical samples to determine true positives, which results in more robust diagnostic accuracy (Erdodi, 2023; McGuire et al., 2019; Schroeder & Martin, 2022). Lastly, the NV-MSVT effectively discriminates between clinical and simulator participants but is not as effective in distinguishing between healthy and simulator participants (Green, 2008). Despite these issues, the NV-MSVT remains a valuable PVT.

Our findings showed that the TMJPI had satisfactory specificity rates while also maintaining its sensitivity at relatively high rates. In other words, the TMJPI can identify true positives and true negatives optimally. A recent study also supported these findings, indicating that TMJPI had high sensitivity and specificity (Adhiatma, 2024). However, we observed that some of our participants felt bored after several questions due to the lack of challenge or interest. This boredom may have contributed to a higher failure rate by lowering participants' attention levels, potentially leading to misclassification as invalid performers. Our TMJPI cut-off score was higher than the (Schagen et al., 1997) ASTM (< 85) and the TMJPI in Adhiatma's study (2024; <87), to achieve higher sensitivity without compromising specificity.

Our study found that the CIHT also had satisfactory specificity rates while maintaining relatively high sensitivity. In other words, the CIHT can identify true positives and true negatives optimally. Our CIHT cut-off score (<10) was higher than Kapur's (1994) study (<5) and Schroeder's (2012) study (<9) to maintain the specificity rates at an acceptable level. Additionally, we observed that simulators often answered the test items incorrectly in an overly obvious manner and displayed uncertainty when choosing their responses. Moreover, studies by Kapur (1994) and Schroeder et al. (2012) have demonstrated that patients with genuine neurocognitive impairments perform differently on the test than those suspected of malingering or feigning memory problems. Specifically, individuals who are malingering or exaggerating their symptoms tend to make more errors on the test, as noted by Kapur (1994). Additionally, research by Daugherty et al. (2021) suggests that participants displaying feigned symptoms may exhibit slower responses on an extended version of CIHT. Given these consistent findings across multiple studies, the CIHT emerges as a valuable tool for clinicians and researchers because its administration, scoring, and interpretation are easy, straightforward, and time-efficient, while also keeping its diagnostic accuracy at a satisfactory level. Additionally, its simplicity and lack of licensing or significant expenses make it particularly suitable for Indonesia, where access to PVTs is limited.

Compared to Green (2008), our NV-MSVT A1 (<84) and A2 (<80) cut-off scores were higher. This made the cut-off more specific without reducing sensitivity. Compared to Adhiatma's (2024) study, our NV-MSVT A1 cut-off score achieved higher sensitivity without compromising the specificity, while the A2 cut-off score maintains the same level of sensitivity and specificity. Our TMT-A (>132) cut-off score was higher than Peak's study (>61); however, our TMT-B cut-off score (>201) was lower (>219; Peak, 2023). As the embedded validity measures of TMT-A and TMT-B had low sensitivity rates, we considered the NV-MSVT, TMJPI, and CIHT to be more appropriate for use in elderly settings.

The failure rate in the control group using our cut-off scores (ranging from 8% to 33%) was attributed to several factors. As previously mentioned, the lack of challenge might lead to boredom and a subsequent loss of attention. Our debriefing sessions revealed that some participants reported losing attention due to insufficient stimulation while completing the TMJPI. Furthermore, a minority of participants exhibited behaviors indicative of excessive effort to remember, which ironically led to anxiety and restlessness. This heightened effort appeared to reduce their attentiveness during the task, ultimately impacting their performance on the NV-MSVT.

Similarly, during the TMT, participants reported distractions and were diverted due to their emotional state. For example, some participants reported that recent personal problems affected their mood and hindered their ability to concentrate, particularly when drawing lines to connect numbers. This issue was even more pronounced in the TMT-B form, where alternating between numbers and letters presented additional challenges. These findings align with current literature, which identifies various factors contributing to failure, such as cognitive issues (lack of understanding, inattention), physical factors (fatigue, disturbance), psychological factors (disinterest, lack of cooperation), or other conditions that divert attention (McWhirter et al., 2020; Ord et al., 2021).

All PVTs in this study demonstrated their utility, with acceptable to outstanding diagnostic accuracy. It is important to note that clinicians cannot solely rely on PVT results for determining performance validity, as performance validity can be inferred from other information, such as behavioral presentation during the examination and discrepancies between clinical information and patients' self-reports (Sherman et al., 2020).

Based on our study, we found that Indonesian elderly perform better with simple tasks, a directive approach, and short-duration assessments to maintain their attention and prevent boredom. Instruments that are simple but lengthy and minimally engaging tend to cause disengagement, while complex tasks may lead to confusion. Given these findings, the type of instrument used can be a crucial factor influencing performance. Tests with lengthy or complex instructions may not be suitable for this population. Therefore, modifying instructions to be simpler and more familiar can enhance comprehension and ensure that information is effectively processed. For the development of Indonesian neuropsychological tests, it is recommended to design assessments that are engaging, culturally relevant, and tailored to the cognitive processing style of the elderly. Future research should explore how specific modifications, such as interactive elements, visual aids, or structured breaks, can further improve test performance.

Our findings extend beyond merely statistical accuracy, establishing a robust empirical foundation for more ethical and valid neuropsychological research throughout Indonesia. We have successfully determined locally optimized cut-off scores for Performance Validity Tests (PVTs) within the older Indonesian demographic, which has three significant, interconnected implications.

Clinically, the primary concern we tackle is the alleviation of misdiagnosis and the ensuing ethical ramifications (Roor et al., 2023). Without locally validated PVTs, clinicians may incorrectly attribute inadequate effort to authentic cognitive decline. This attribution raises ethical concerns and frequently leads to the misallocation of resources towards ineffective and unsuitable treatment. Our research refutes this by establishing elevated specificity cut-off values, offering a crucial scientific instrument to validate that atypical scores genuinely indicate cognitive dysfunction, thus enhancing the ethical standards of clinical practice in Indonesian neuropsychology.

From an academic standpoint, our work significantly advances the field of cross-cultural neuropsychology in Asia. Given that our ideal cut-offs deviate from conventional Western standards, it is clear that renormalization and local validation are necessary conditions for equitable PVT use (Ali et al., 2022). Other developing countries in Southeast Asia should use this study as a model for their methodology. Importantly, it emphasizes that PVTs require more than just translation; they must be locally recognized to gauge effort, not just differences in education or cultural barriers (Franzen et al., 2021).

Lastly, the findings provide crucial assistance for forensic and administrative decision-making. Locally validated PVT data is crucial because invalid performance can be driven by external motivation in high-stakes situations like court cases or disability claims (Sherman et al., 2020). Our unbiased, scientifically supported evidence reduces the possibility of systemic abuse. This study directly contributes to safeguarding the integrity of Indonesia's social justice and health systems by ensuring that public resources are only allocated to those with confirmed cognitive impairment.

Future research efforts should take into account a number of limitations, even though our study offered insightful information about the diagnostic accuracy of multiple PVTs among healthy elderly people in Indonesia. First off, our study's small sample size may have limited the findings' applicability to more general attributes of older adults, particularly in Indonesia. A bigger sample size and a wider range of demographic traits should be used in future studies. Furthermore, the intricacies and subtleties present in actual instances of feigning or malingering may not be accurately replicated by the experimental use of simulated dementia. When used in clinical settings, this simulation design may overestimate the PVTs' actual diagnostic accuracy, indicating the need for validation in real patient populations. Last but not least, the study's geographically limited focus on a single area restricts the findings' generalizability across Indonesia's enormous linguistic and educational diversity. Future studies should concentrate on three main areas in light of these constraints. In order to verify the cut-off scores' practical sensitivity and specificity, it is imperative that they be verified in real clinical samples, which are patients who exhibit documented cognitive impairments. Second, in order to guarantee cultural and linguistic neutrality, research should broaden the demographic reach to include senior citizens from various Indonesian provinces. Third, in order to create the most reliable and effective protocol for use in clinical practice in low-resource Indonesia, researchers should investigate the best use of composite PVT batteries, which combine standalone and embedded measures.

Conclusion

This study effectively assessed the diagnostic accuracy of various Performance Validity Tests (PVTs) in older Indonesian adults, resulting in the recommendation of refined local cut-off scores based on acceptable specificity rates. The CIHT exhibited the highest diagnostic accuracy among the assessed measures, succeeded by the TMJPI, NV MSVT, and the embedded TMT validity measures, the latter of which displayed the lowest sensitivity. These results give Indonesian doctors the first scientifically based standards they need to question false performance and make sure that neuropsychological diagnoses are ethical and accurate. Ultimately, the successful validation of these PVTs provides essential scientific foundations for enhancing the standard of care throughout Indonesia and contributes critical methodological insights to the global domain of cross-cultural neuropsychology.

Credit/Acknowledgement

The authors wish to thank Atma Jaya Catholic University of Indonesia and the ethics committee for their support in carrying out this research. We are also deeply grateful to the older adult volunteers, including residents from nursing homes and individual volunteers recruited from the community, for their generous time and effort.

References

Adhiatma, W. (2024). The development of performance validity tests (PVTs) for use in neuropsychological assessment in Indonesia. *Donders Series*. https://hdl.handle.net/2066/302608

Adhiatma, W., Hendriks, M. P. H., Halim, M. S., & Kessels, R. P. C. (2025). The validation of multiple performance validity tests (PVTs) for Indonesia. *Psychological Neuroscience.* https://doi.org/10.1037/pne0000340

Ali, S., Crisan, I., Abeare, C. A., & Erdodi, L. A. (2022). Cross-cultural performance validity testing: Managing false positives in examinees with limited English proficiency. *Developmental Neuropsychology*, *47*(6), 273–294. https://doi.org/10.1080/87565641.2022.2105847

Alzheimer's Association (n.d.). *Medical tests for diagnosing Alzheimer's.* https://www.alz.org/alzheimers-dementia/diagnosis/medical_tests

Indonesia Statistical Bureau. (2022). *Statistik penduduk lanjut usia* 2022. BPS.

Bass, C., & Wade, D. T. (2019). Malingering and factitious disorder. *Practical Neurology, 19*(2), 96–105. https://doi.org/10.1136/practneurol-2018-001950

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the trail making test. *Nature Protocols, 1*(5), 2277–2281. https://doi.org/10.1038/nprot.2006.390

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

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

Daugherty, J. C., Querido, L., Quiroz, N., Wang, D., Hidalgo-Ruzzante, N., Fernandes, S., Pérez-García, M., De los Reyes-Aragon, C. J., Pires, R., & Valera, E. (2021). The coin in hand-extended version: Development and validation in a multicultural performance validity test. *Assessment, 28*(1), 186–198. https://doi.org/10.1177/1073191119864652

Erdodi, L. A. (2023). Multivariate models of performance validity: The Erdodi Index captures the dual nature of non-credible responding (continuous and categorical). *Assessment, 30*(5), 1467–1485. https://doi.org/10.1177/10731911221101910

Franzen, S., Watermeyer, T. J., Pomati, S., Papma, J. M., Nielsen, T. R., Narmel, P., Mukadam, N., Lozano-Ruiz, Á., Ibanez-Casas, I., Goudsmit, M., Fasfous, A., Daugherty, J. C., Canevelli, M., Calia, C., van den Berg, E., & Bekkhus-Wetterberg, P. (2021). Cross-cultural neuropsychological assessment in Europe: Position statement of the European Consortium on cross-cultural neuropsychology (ECCroN). *The Clinical Neuropsychologist*, *36*(3), 546–557. https://doi.org/10.1080/13854046.2021.1981456

Galioto, R., Grezmak, T., Swetlik, C., Abbatemarco, J. R., Titulaer, M. J., Finke, C., & Kunchok, A. (2023). Neuropsychological testing in autoimmune encephalitis: A scoping review. *Neurology Neuroimmunology & Neuroinflammation, 11*(1), e200179. https://doi.org/10.1212/NXI.0000000000200179

Green, P. (2008). Non-verbal medical symptom validity test for windows: User's manual and program. Green's Publishing.

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

Guilmette, T. J., et al. (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

Kapur, N. (1994). The coin-in-the-hand test: A new "bed-side" test for the detection of malingering in patients with suspected memory disorder. *Journal of Neurology, Neurosurgery, and Psychiatry, 57*(3), 385–386. https://doi.org/10.1136/jnnp.57.3.385

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

Lindeza, P., Rodrigues, M., Costa, J., Guerriro, M., & Rosa, M.M. (2020). Impact of dementia on informal care: A systematic review of family caregivers' perceptions. *BMJ Supportive & Palliative Care*, *0*, 1-12. http://dx.doi.org/10.1136/bmjspcare-2020-002242

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Macpherson, S.E., Cox, S.R., Dickie, D.A., Karama, S., Starr, J.M., Evans, A.C., Bastin, M.E., Wardlaw, J.M., & Deary, I.J. (2017). Processing speed and the relationship between Trail Making Test-B performance, cortical thinking, and white matter microstructure in older adults. *Cortex*, *95*(0), 92-103. https://doi.org/10.1016/j.cortex.2017.07.021

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, 34(1), 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*(9), 945–952. https://doi.org/10.1136/jnnp-2020-323776

Merrilees, J. (2016). The impact of dementia on family caregivers: What is research teaching us? *Current Neurology and Neuroscience Reports*, *16*(88). https://doi.org/10.1007/s11910-016-0692-z

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). *The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment.*

National Health Services (2023). *Tests for diagnosing dementia.* https://www.nhs.uk/conditions/dementia/symptoms-and-diagnosis/tests/

National Institute of Aging (2022). *What is dementia? Symptoms, types, and diagnosis.* https://www.nia.nih.gov/health/what-is-dementia#:~:text=The%20following%20procedures%20also%20may,Brain%20scans

Ord, A. S., Shura, R. D., Sansone, A. R., Martindale, S. L., Taber, K. H., & Rowland, J. A. (2021). Performance validity and symptom validity tests: Are they measuring different constructs? *Neuropsychology*, *35*(3), 241–251. https://doi.org/10.1037/neu0000722

Pesau, H. G., & Luijtelaar, G. V. (2021). Equivalence of traditional and internet-delivered testing of word fluency tasks. *Jurnal Psikologi*, *20*(1), 35–49. https://doi.org/10.14710/jp.20.1.35-49

Peak, A. M., Marceaux, J. C., Chicota-Carroll, C., & Soble, J. R. (2023). Cross-validation of the Trail Making Test as a non-memory-based embedded performance validity test among veterans with and without cognitive impairment. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/13803395.2023.2287784

Prasetya, V. G. H., & Susanto, H. (2024). Construction of Indonesian executive function self-report scale (IEF-SRS). *Psikostudia: Jurnal Psikologi*, *13*(1), 66. https://doi.org/10.30872/psikostudia.v13i1.12918

Roor, J. J., Peters, M. J. V., Dandachi-FitzGerald, B., & Ponds, R. W. H. M. (2023). Performance validity test failure in the clinical population: A systematic review and meta-analysis of prevalence rates. *Neuropsychology Review*. Springer. https://doi.org/10.1007/s11065-023-09582-7

Schagen, S., Schmand, B., & 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), 163-172. https://doi.org/10.1080/01688639708405063

Schroeder, R. W., Peck, C. P., Buddin, W. H., Heinrichs, R. J., & Baade, L. E. (2012). The coin-in-the-hand test and dementia: More evidence for a screening test for neurocognitive symptom exaggeration. *Cognitive Behavioral Neurology*, *25*(3), 139-143.

Schroeder, R. W., & Martin, P. K. (2022). Validity assessment within the memory disorders/dementia clinic. In R. W. Schroeder & P. K. Martin (Eds.), *Validity Assessment in Clinical Neuropsychological Practice: Evaluating and Managing Noncredible Performance* (pp. 254–279). The Guilford Press.

Šimundić, A.-M. (2009). Measures of diagnostic accuracy: Basic definitions. *EJIFCC*, *19*(4), 203–211. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4975285/

Sherman, E. M. S., 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

Shin, J. H., & Kim, J. H. (2022). Family caregivers of people with dementia associate with poor health-related quality of life: A nationwide population-based study. *International Journal of Environmental Research and Public Health*, *19*(23), 16252. https://doi.org/10.3390/ijerph192316252

Soble, J. R. (2021). Future directions in performance validity assessment to optimize detection of invalid neuropsychological test performance: Special issue introduction. *Psychological Injury and Law*, *14*(4), 227–231. https://doi.org/10.1007/s12207-021-09425-x

Sulastri, A., Utami, M. S. S., Jongsma, M., Hendriks, M., & Van Luijtelaar, G. (2018). 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*, https://doi.org/10.21275/ART20202307

The jamovi project (2024). *jamovi.* (Version 2.5) [Computer Software]. Retrieved from https://www.jamovi.org

Tran, T. A., Mattap, S. M., Warren, N., Teng, J. H. J., Duong, M. D., Hoang, V. M., Yasin, S., & Mohan, D. (2025). Caregiver burden among dementia caregivers in low-and middle-income countries in Asia: A systematic review. *Aging & Mental Health*, 29(4), 578–590. https://doi.org/10.1080/13607863.2025.2462110

Unal, I. (2017). Defining an optimal cut-point value in ROC analysis: An alternative approach. *Computational and Mathematical Methods in Medicine*, 2017, 14 pages. https://doi.org/10.1155/2017/3762651

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

Wahyuningrum, S. E., Sulastri, A., Hendriks, M., Consortium, I. N., & Van Luijtelaar, G. (2022). The Indonesian neuropsychological test battery (INTB): Psychometric properties, preliminary normative scores, the underlying cognitive constructs, and the effects of age and education. *Acta Neuropsychologica*, 20(4), 445–470. https://doi.org/10.5604/01.3001.0016.1339

Widhianingtanti, L. T., Luijtelaar, G. Van, Suryani, A. O., Hestyanti, Y. R., & Sulastri, A. (2022). Indonesian trail making test: Analysis of psychometric properties, effects of demographic variables, and norms for Javanese adults. *Jurnal Psikologi*, *49*(2), 104. https://doi.org/10.22146/jpsi.68953