OXFORD
UNIVERSITY PRESS

Archives of Clinical Neuropsychology 37 (2022) 91–103

Archives of CLINICAL NEUROPSYCHOLOGY

Validation of the Coin in Hand-Extended Version Among Older Adults With and Without Dementia

Inês Ferreira[1], Sandra Fernandes[1,2,*], Luís Querido[1], Rute Pires[1,2], Julia C. Daugherty[3,4], Natalia Hidalgo-Ruzzante[3,5], Miguel Pérez-García[3,4]

[1]*Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, Lisboa, Portugal*
[2]*CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, Lisboa, Portugal*
[3]*The Mind, Brain and Behaviour Research Centre (CIMCYC), University of Granada, Granada, Spain*
[4]*Department of Personality, Evaluation, and Psychological Treatment, University of Granada, Granada, Spain*
[5]*Department of Evolutionary and School Psychology, Universidad de Granada, Granada, Spain*

*Corresponding author at: Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, Lisboa 1649-013, Portugal. *E-mail address:* sfernandes@psicologia.ulisboa.pt (S. Fernandes).

Accepted 20 March 2021

Abstract

Objective: The current study aimed to validate the Performance Validity Test Coin in Hand-Extended Version (CIH-EV) in groups of healthy older adults and older adults with dementia.
Method: Using an analog simulation paradigm, the healthy control group and the clinical group were instructed to perform to the best of their ability, whereas the feigning older adults were instructed to simulate a memory deficit to obtain allowance, financial aid, or early retirement.
Results: Results showed that the control and clinical groups performed more optimally than the feigning group, although the clinical group had superior response times. The CIH-EV was insensitive to sociodemographic variables and neurocognitive functioning in all groups, demonstrated good convergent validity with other performance validity measures, and showed a reduced rate of false positives.
Conclusions: This study corroborates the CIH-EV's effectiveness in detecting the simulation of cognitive deficits in healthy older adults and older adults with dementia.

Keywords: Performance validity; Simulation; CIH-EV; Older adults; Dementia

Although symptoms simulation has been a recognized problem in all societies (Wessely, 2003), its research and clinical consideration is relatively recent, with approximately two decades of intensive work (Heilbronner et al., 2009). The growing interest in this subject derives from the recognition that test results may be affected by variables such as motivation, effort, and cooperation at the time of the assessment (Larrabee, 2005) and not just by a cognitive deficit or an organic brain injury (Simões et al., 2010). Research shows that clinicians who rely solely on their clinical intuition are unable to identify invalid performance (e.g., Faust, Hart, & Guilmette, 1988), thus specific measures were developed to maximize confidence in test results and diagnoses (Bush et al., 2005). However, many clinicians still do not include such measures in their assessments, which leads to incorrect predictions in about 24% of the cases (Dandachi-FitzGerald, Merckelbach, & Ponds, 2017). In fact, symptom exaggeration or fabrication for personal gains, as well as symptom minimization or omission to avoid undesirable consequences, is increasingly common in forensic contexts (Larrabee, 2005). Given that legal decisions have clinical, financial, and social impacts that limit the resources available to individuals with real deficits, it is critical to systematically assess the performance credibility whenever there is a possibility of external gain (Lippa, 2018). In addition, it is currently recommended (e.g., Bush et al., 2005) that performance validity assessment be extended to everyday clinical practice.

https://doi.org/10.1093/arclin/acab019 Advance Access publication 15 April 2021

Several definitions of simulation have been proposed. The most consensual one considers simulation as the exaggeration or intentional fabrication of cognitive impairment in order to achieve substantial material gain or to avoid or escape from formal duty or responsibility (Slick, Sherman, & Iverson, 1999, p. 552). Traditionally, simulation is seen as a diagnosis with specific criteria, as Slick and colleagues (1999) proposed for Malingering Neurocognitive Dysfunction (MND). However, more recently, some authors argued that simulation is not an all-or-nothing phenomenon (e.g., Fergunson, 2004). Feigning behaviors have been proposed as being dimensional, varying between different levels of simulation (Bush et al., 2005), which is reflected either in the existence of feigning subgroups or in different performance patterns on specific tests (Larrabee, 2012).

Simulation detection initially relied on the analysis of a series of criteria that included inconsistent response patterns (Slick et al., 1999; Slick & Sherman, 2013) on what are called embedded cognitive Performance Validity Tests (PVTs; Larrabee, 2005). Among these, the Wechsler Adult Intelligence Scale (WAIS-III) Digit Span subtest, namely the Reliable Digit Span (RDS) indicator (Greiffenstein, Baker, & Gola, 1994), is one of the most studied and empirically supported measures (Jasinski, Berry, Shandera, & Clark, 2011). More recently, another type of PVT has been developed, namely stand-alone or free-standing test. Similar to embedded PVTs, stand-alone PVTs also assess the cognitive performance validity, only they can be administered on their own outside of a larger neuropsychological battery as opposed to within an already existing measure.

The PVTs are based on the ceiling effect, that is, although seemingly difficult, they are so simple that people with real neurological deficits can perform them easily. Based on this idea, clinicians can infer that individuals who perform poorly or significantly worse than real patients may have an intentional non-credible behavior. Some of the PVTs are also based on a forced choice paradigm with two response options. According to the binominal theorem, it is possible to calculate the below-chance response patterns (Frederick & Speed, 2007). Due to the fact that the probability of randomly hitting one of the two options is 50%, scores significantly below the level of chance provide evidence on intentional simulation (Vickery, Berry, Inman, Harris, & Orey, 2001).

In order to enhance reliability, the current recommendation (Heilbronner et al., 2009) is to use multiple PVTs. Rey-15 Item Memory Test (Rey-15IMT; Rey, 1964) and Test of Memory Malingering (TOMM; Tombaugh, 1996) are the most used PVTs. Although the TOMM is considered a gold standard test (Slick et al., 1999; Slick & Sherman, 2013), the Rey-15IMT seems to be sensitive to gender, education, culture (e.g., Strutt, Scott, Shrestha, & York, 2011), age (e.g., Green et al., 2014), memory deficits, and dementia (e.g., Fazio, Faris, & Yamout, 2019). In addition to the fact that there are few PVTs, thus posing a greater risk for the successful coaching and familiarization with tests (Suhr & Gunstad, 2007), some of them have limitations, which further impedes compliance with the original recommendation of applying multiple PVTs.

Recently, Coin in Hand-Extended Version (CIH-EV; Daugherty et al., 2019), a computerized adaptation of Kapur's (1994) original Coin in Hand Test, was validated on young healthy adults from Portugal, Spain, Colombia, and the United States (Daugherty et al., 2019). The CIH-EV is a public domain and computerized test, which allows a wider and more standardized use, with a variety of indicators (i.e., response time, response pattern across three difficulty levels, and total correct response). The CIH-EV revealed high convergent validity with other simulation measures, suggesting a preliminary adequacy for performance credibility assessment among non-clinical individuals. In addition, there was a cultural equivalence between samples because the cutoff score was the same for all countries ($\geq$27), with sensitivity and specificity values of 62% and 95%, respectively, and a minimal false positive rate of 5%. Although CIH-EV proved to be effective in simulator/non-simulator discrimination, these values could be inflated given that all the participants were healthy young adults with high education. Therefore, it was suggested (Daugherty et al. 2019) that further research should include participants with real cognitive impairment to verify the potential need for cutoff adjustments to improve the test's diagnostic accuracy. Patients with dementia may be a relevant population for this purpose, with around 50 million people currently diagnosed and nearly 10 million new cases every year, approximately 60% of whom live in low- and middle-income countries. Despite the high prevalence, PVTs research in samples with this diagnosis is still scarce (Dean, Victor, Boone, Philpott, & Hess, 2009).

Dementia is a syndrome in which there is cognitive function deterioration beyond what might be expected from normal aging, resulting from a variety of diseases and injuries that primarily or secondarily affect the brain (World Health Organization, 2020). Older adults with dementia are often excluded from studies validating PVTs because the cutoff scores are not adequate for these samples, producing high false positive rates (Dean et al., 2009). In addition, in tests based on the ceiling effect, older adults with dementia have lower performance, making it difficult to distinguish their scores from those who are feigning impairment (van Gorp & Hassenstab, 2009). Although some PVTs have been validated for populations with mild cognitive impairment, studies conducted on patients with dementia show that higher severity of cognitive decline was related to diminished diagnostic accuracy on the PVTs (McGuire, Crawford, & Evans, 2019).

Age is known to be the biggest risk factor for dementia (Alzheimer's Society, 2016). In Portugal (National Statistics Institute [INE], 2018), ranked fourth among the countries with the highest number of dementia cases, older adults represent 21.5% of the total population (OECD, 2018). Although this scenario might allow for cognitive deficit

simulation in older adults who pretend to obtain a secondary gain, such as early retirement, incapacity pensions, or health insurance (Simões et al., 2010), inaccurate classification may also discredit real patients' complaints.

The current study aimed to validate the CIH-EV Test (Daugherty et al., 2019) with older Portuguese adults aged 65 years or above, following an analog paradigm. A clinical group of older individuals with dementia was tested for performances' comparison. The study aimed to: analyze the relationship between demographic variables and CIH-EV performance in three groups (healthy control, healthy feigning, and older adults with dementia); do a comparative study of CIH-EV's three difficulty levels and total performance, both with regard to response accuracy and response time between the three groups; establish the CIH-EV cutoff scores, as well as their sensitivity and specificity values, for both healthy and demented older adults; and determine the CIH-EV convergent validity with the Portuguese version of the TOMM (Fernandes et al., 2009), Rey 15-IMT (Simões et al., 2010), and RDS measure (Pinho, 2012).

Several hypotheses were considered. Given that healthy older adults were randomly assigned to feigning and control groups, we hypothesized that groups would not differ in sociodemographic variables and performance on cognitive screening tests (H1). Because the clinical group was composed of institutionalized older adults with dementia, we hypothesized that the only sociodemographic variables in which the clinical and feigning groups would differ were cognitive performance level and residence (H2). Following the evidence of studies with healthy and demented older adults (e.g., Schroeder, Peck, Buddin, Heinrichs, & Baade, 2012), we hypothesized that the CIH-EV would be insensitive to all sociodemographic variables and neurocognitive functioning (H3). Based on the assumption of PVTs' ceiling effect, we expected that control and clinical groups would perform better on the CIH-EV than the feigning group (H4). Given that previous studies (e.g., Ferreira, Gomes, Moreira, Silva, & Cavaco, 2017) pointed to CIH insensitivity to age, we expected that CIH-EV's cutoff score for the healthy older adult group would be similar to the young adults' cutoff score (Daugherty et al., 2019; H5). Because previous studies (e.g., McGuire et al., 2019) suggested a decrease in cutoff score specificity as a function of dementia severity, we hypothesized that the CIH-EV's cutoff score for the clinical group would be lower than that of the healthy control group (H6). Finally, in line with previous studies (e.g., Daugherty et al., 2019), we hypothesized that the CIH-EV would have good convergent validity with the other performance validity measures (TOMM, Rey-15IMT, and RDS). However, according to previous evidence in samples of older adults with dementia (e.g., Yeh et al. 2019), it was expected that CIH-EV would show a lower false positive percentage than the remaining PVTs (TOMM, Rey-15IMT, and RDS; H7).

Method

Participants

The sample included 76 participants (64 women) with an age range between 65 and 94 ($M = 75.78$; standard deviation [SD] = 0.92). Following the analog simulation design of previous studies on validity tests (e.g., Simões et al., 2010), participants were randomly assigned to one of two groups: the control or feigning group. Control participants were instructed to perform to the best of their abilities, whereas feigning participants were asked to perform as if they had a memory impairment typical of dementia to obtain pension, allowance, or early retirement. The first group ($n = 30$, 25 women) had an age range between 65 and 94 ($M = 75.20$; $SD = 8.90$) and a mean education of 8.43 years ($SD = 3.94$). The second group ($n = 29$, 25 women) had an age range between 65 and 87 ($M = 73.31$; $SD = 6.40$) and a mean education of 9.07 years ($SD = 3.36$). To enhance the generalization and comparison of results (Rogers, 2008), a clinical group with a dementia diagnosis ($n = 17$, 14 women) was also included. In this group, age ranged from 68 to 92 years old ($M = 81.00$; $SD = 6.82$) and the mean education was 8.35 years ($SD = 3.66$). The sample size of at least 15 participants per group was calculated using the G*power program by setting the alpha at <0.05, the power at 0.90, and using the mean ($d = 2.11$) of the effect size for the CIH-EV total hits ($d = 3.3$) and total time ($d = 0.93$), which was found in the preliminary validation of the CIH-EV measure (Daugherty et al., 2019).

Inclusion criteria required participants to be 65 years or older and to dominate the Portuguese language. Exclusion criteria included chronic medical problems that interfered with cognitive functioning, history of traumatic brain injury or substance abuse, uncorrected visual or auditory problems, and oral comprehension difficulties. The presence of cognitive impairment was an exclusion criterion for the control and feigning groups and an inclusion criterion for the clinical group, as it is associated with a medical diagnosis of dementia. None of the participants were involved in forensic proceedings.

Instruments

A total of seven instruments were administered: one sociodemographic questionnaire, two cognitive screening tests (Mini-Mental State Examination [MMSE] and Montreal Cognitive Assessment [MoCA]), one measure for simulation assessment (WAIS-III Digit Span subtest), and three PVTs (Rey-15IMT, TOMM, and CIH-EV).

Sociodemographic questionnaire. To collect sociodemographic data, participants were asked about their age, gender, occupation, marital status, number of children, residence type, income, educational level, and professional activity.

Mini-Mental State Examination. The MMSE (Folstein et al., 1975; Guerreiro et al., 1994) is a brief global cognitive screening test that assesses temporal orientation, repetition, attention and calculation, memory, language, and constructive ability. A maximum score of 30 points indicates no cognitive impairment. The cutoff scores for cognitive impairment in the Portuguese population (Guerreiro et al., 1994) are scores ≤22 for individuals with 1–11 years of education and scores ≤27 for individuals with more than 11 years of education. For the clinical group, a cutoff score lower than 26 was used (e.g., Freitas, Simões, Alves, & Santana, 2013). The MMSE was included in the assessment protocol because it is the most used, validated, and referenced cognitive screening test in the literature (Simões, Santana, & Cerebral Aging and Dementia's Studies Group, 2015).

Montreal Cognitive Assessment. The MoCA (Freitas, Simões, Alves, & Santana, 2011; Nasreddine et al., 2005) is a brief cognitive screening test that assesses milder forms of cognitive decline to distinguish between the normative and pathological cognitive changes. It assesses the executive functioning, visuospatial ability, short-term memory, working memory, attention and concentration, language, and spatiotemporal orientation. According to the Portuguese norms (Freitas et al., 2011), the present study considered cutoff scores below 1.5 *SD* for individuals older than 65 years, depending on their education level. Thus, cognitive impairment was admitted for scores lower than 16 for 1–4 years of education; lower than 20 for 5–9 years of education; lower than 22 for 10–12 years of education; and lower than 24 for more than 12 years of education. For the clinical group, a cutoff score lower than 17 was used (Freitas et al., 2013). The MoCA was included in the assessment protocol because, in contrast to the MMSE, it also assesses executive functions and is more sensitive to the early stages of dementia (Freitas et al., 2011; Nasreddine et al., 2005).

WAIS-III Digit Span subtest. Digit Span is a subtest of the WAIS-III (Wechsler, 2008) that assesses memory, attention, and concentration abilities. The participant is asked to repeat the same number sequences, which progressively increase in length and thus difficulty. The subtest has two conditions: forward repetition, with a cutoff score ≤5, and backward repetition, with a cutoff score ≤2. Further, it allows for the derivation of the RDS indicator with a cutoff score ≤6 (Pinho, 2012). This subtest was included in the assessment protocol as literature (e.g., Jasinski et al., 2011) shows its relevance for the detection of simulated neurocognitive deficits.

Rey 15-Item Memory Test. The Rey-15IMT (Boone, Salazar, Lu, Warner-Chacon, & Razani, 2002; Rey, 1964; Simões et al., 2010) is a PVT that aims to assess the effort and/or memory impairment simulation. It consists of a card with 15 items (letters, geometric shapes, and numbers) that are arranged in three columns by five rows. The card is presented for 10 s and then the participants are asked to draw the items from memory. In this Free Recall Trial, each correctly reproduced item is scored with 1 point regardless of its spatial location. The following Recognition Trial consists of a page with 15 original items and 15 new similar items. Participants must then select the original ones. In this test, the number of correctly recognized items, the number of false positives (i.e., incorrectly identified items), and a Combined Result (number of correctly recalled items + [number of correctly recognized items − number of false positives]) are calculated. A cutoff score lower than 9 for the Free Recall Trial and lower than 20 for the Combined Result were determined (Simões et al., 2010). The Rey-15IMT was considered in the assessment protocol, as it is a widely used PVT and is validated for older Portuguese adults (Simões et al., 2010). As such, the test is also useful for the analysis of convergent validity with the CIH-EV (Daugherty et al., 2019).

Test of Memory Malingering. The TOMM (Fernandes, 2009; Tombaugh, 1996) is a visual recognition PVT. It consists of two learning trials in which 50 images of the common objects are presented for 3 s each at a 1-s interval. Immediately after each trial, the same images are shown paired up with a new distractor image. Because it is a forced-choice task, for each pair, the participant must choose which image was previously presented. Both trials differ in the image presentation order and the distractor images used. Feedback is provided for all responses and 1 point is given to each correct response, with a maximum score of 50 points per trial. A cutoff score lower than 45 for the second trial was used (e.g., Fernandes, 2009). The TOMM is a PVT of reference in research and clinical practice and is validated for older Portuguese adults (e.g., Fernandes, 2009), thus making it a relevant measure for the convergent validity analysis with the CIH-EV.

Coin in Hand–Extended Version. The CIH (Daugherty et al., 2019; Kapur, 1994) was originally developed to distinguish between patients with neurocognitive disorders and patients who exaggerate or simulate memory complaints (Kapur, 1994). In this forced-choice test, the participant is initially informed that the test's purpose is to assess how memory can resist distraction. First, he/she sees a pair of hands with a coin either in the left or right hand. Then the hands close into a fist and disappear. A countdown occurs and the closed fists appear again. At this time, the participant must select the hand in which the coin is held. Feedback is provided for all responses. An extensive computerized version (CIH-EV; Daugherty et al., 2019) was used in the present study. This version includes three difficulty levels, with 10 trials for each (i.e., the coin randomly appears five times in the left hand and five times in the right hand), for a total of 30 trials. As such, the total score ranges from 0 to 30. The level of difficulty is determined by the duration of the countdown: At the first level, the participant counts down from 10 for 10 s and then makes a choice; from 99 for 15 s on the second level; and from 999 for 20 s on the third level. The participant is informed about the increased difficulty at the second and third levels. The CIH-EV can be accessed on the instrument's webpage (https://projectbelieve.info/en/professionals/), or by downloading it from the app store. The current version requires Internet connection, although future versions of the instrument will be provided for use offline. Prior to using the CIH-EV, professionals must be granted access by developers at the University of Granada (Spain) via the aforementioned web page.

Procedure

Participants in the control and feigning groups were tested at three senior universities, non-profit facilities that offer social, cultural, and educational opportunities in an informal learning context for seniors in the community, and were randomly distributed into groups. Neuropsychological assessment of the clinical group participants took place at two residential care facilities, and dementia diagnosis was previously determined by the medical team of both residences. All the institutions were from Lisbon, Portugal. All participants signed an informed consent about the purpose of the study and the type and duration of tasks. After their consent, each participant was assigned an alphanumeric code to ensure the anonymity of the data. The project was approved by the Faculty of Psychology—University of Lisbon's ethics committee before testing.

The neuropsychological protocol had a duration of approximately 50 min and it was individually administered. For the clinical group, the protocol was divided into two sessions of approximately 30 min each. The session started by asking the participant about the exclusion criteria, followed by the sociodemographic questionnaire, except for in the case of the clinical group, whose information was provided by the residential facilities. After the cognitive screening with the MMSE and MoCA, the feigning group participants were instructed to perform the following tests (Digit Span subtest, TOMM, REY-15IMT, and CIH-EV) simulating a memory impairment typical of dementia to obtain pension, allowance, or early retirement. Test administration order was randomized, and the examiner was aware of who was in the control and feigning group. The TOMM and CIH-EV were administrated using a computer with a 13.3″ screen. Specifically, in the case of the CIH-EV, all participants were initially informed about the three difficulty levels in accordance with Daugherty and colleagues (2019).

Statistical analyses were performed using IBM SPSS Statistics software.

Results

Analysis of Differences Between Groups in Sociodemographic Variables and Cognitive Performance

Differences between groups for categorical sociodemographic variables were compared using a Chi-square test. There were no significant differences between the three groups (see Table 1). Differences between control and feigning groups for age, MMSE, and MoCA results were compared using *t*-tests. For feigning and clinical groups' comparison, the Mann–Whitney *U* non-parametric test was used. Significant differences were found for age ($U = 103.50$, $p < .001$) and cognitive screening test results (MMSE: $U = 37.00$, $p < .001$; MoCA: $U = 3.00$, $p < .001$; see Table 1).

Analysis of the Relationship Between Sociodemographic variables, Cognitive Performance, and CIH-EV Performance

Pearson's correlation coefficient was used for the analysis of the relationship between sociodemographic variables, cognitive performance, and CIH-EV results for each group. There were no significant correlations in any group (see Table 2).

Accuracy and response time analysis of the three groups for difficulty levels and total CIH-EV results. Table 2 displays the three groups' performance on the CIH-EV by difficulty level for accuracy and response time. To analyze the differences between the three difficulty levels, two 3 × 3 ANOVAs were performed, with difficulty level (Levels 1–3) and group (control, feigning,

Table 1. Mean, *SD* and percentage for sociodemographic variables and cognitive screening of the three groups

	Control group (n = 30)	Range	Feigning group (n = 29)	Range	t[a]	Clinical group (n = 17)	Range	U[b]	X^2[c]
Sex									0.50
Women	25 (83.33%)		25 (86.21%)			14 (82.35%)			
Men	5 (16.67%)		4 (13.79%)			3 (17.65%)			
Age	75.20 ± 8.90		73.31 ± 6.40		0.93	81.00 ± 6.82		**103.50***	
Years of education									3.08
0–4	9 (30.00%)		4 (13.79%)			3 (17.65%)			
5–9	10 (33.33%)		13 (44.83%)			7 (41.18%)			
10–12	8 (26.67%)		10 (34.48%)			5 (29.41%)			
>12	3 (10.00%)		2 (6.90%)			2 (11.76%)			
Mean	8.43 ± 3.94	4–17	9.07 ± 3.36	4–15		8.35 ± 3.66	4–14		
Marital status									9.51
Single	3 (10.00%)		3 (9.69%)			4 (23.53%)			
Married	13 (43.33%)		11 (37.93%)			3 (17.65%)			
Widower	5 (16.67%)		10 (34.48%)			8 (47.06%)			
Divorced	9 (30.00%)		5 (17.24%)			2 (11.76%)			
Children	1.73 ± 1.14		1.42 ± 0.85			1.60 ± 1.35			13.81
Habitation									
Alone	14 (46.67%)		18 (58.06%)			0 (0.00%)			2.41
Partner	12 (40.00%)		8 (25.81%)			0 (0.00%)			
Relatives	3 (10.00%)		5 (16.13%)			0 (0.00%)			
Residential care	1 (3.33%)		0 (0.00%)			17 (100.00%)			
Income									4.18
≤1200€	18 (60.00%)		21 (67.74%)			—			
1200€ to <1800€	6 (20.00%)		9 (29.03%)			—			
≥1800€	6 (20.00%)		1 (3.23%)			—			
Longest profession									3.48
Low	12 (40.00%)		14 (48.28%)			11 (64.71%)			
Medium	12 (40.00%)		12 (41.38%)			4 (23.53%)			
High	6 (20.00%)		3 (10.34%)			2 (11.76%)			
Mean	Medium (M = 1.80 ± 0.76)		Medium (M = 1.62 ± 0.68)			Low (M = 1.47 ± 1.71)			
Last profession									
Low	12 (40.00%)		13 (44.83%)			11 (64.71%)			3.36
Medium	13 (43.33%)		13 (44.83%)			4 (23.53%)			
High	5 (16.67%)		3 (10.34%)			2 (11.76%)			
Mean	Medium (M = 1.77 ± 0.73)		Medium (M = 1.66 ± 0.67)			Low (M = 1.47 ± 0.72)			
MMSE	29.07 ± 0.17	27–30	28.86 ± 0.18	27–30	0.85	19.59 ± 1.36	10–30	**37.00***	
MoCA	24.67 ± 0.49	19–29	23.69 ± 0.41	17–28	1.53	11.41 ± 1.15	4–20	**3.00***	

Note: MMSE = Mini-Mental State Examination; MoCA = Montreal Cognitive Assessment. Bold corresponds to the sociodemographic variables where the groups statistically differ.

[a]t-test values.

[b]Mann–Whitney U values.

[c]Chi-square values.

*$p < .05$.

and clinical) as independent variables and with accuracy and response time as dependent variables. An additional ANOVA was also performed with the same groups for total accuracy (i.e., the sum of hits at the three difficulty levels) and the total response time between the three difficulty levels.

Regarding accuracy, the ANOVA did not reveal a significant difficulty level effect; $F(2, 146) = 1.31$, $p > .10$; however, there was a significant group effect; $F(2, 73) = 312.78$, $p < .001$, $\eta_p^2 = 0.89$. Moreover, the interaction between difficulty level and group was not significant; $F(4, 146) = 1.04$, $p > .10$. According to the post hoc analysis, the control group had a better performance (M = 9.08) than the clinical group (M = 8.33), and both performed better than feigning group (M = 4.30).

Regarding response time, the ANOVA did not reveal a significant effect for difficulty level; $F(2, 146) = 0.058$, $p > .10$; however, a significant group effect was shown; $F(2, 73) = 567.43$, $p > .001$, $\eta_p^2 = 0.94$. Additionally, there was not a significative interaction between the difficulty level and group; $F(4, 146) = 0.101$, $p > .10$. The post hoc analysis revealed a better response

Table 2. Mean and *SD* of accuracy and response time for each Coin in Hand-Extended Version's (CIH-EV) difficulty level for the three groups

	Control group			Feigning group			Clinical group		
	L1 *M* (*SD*)	L2 *M* (*SD*)	L3 *M* (*SD*)	L1 *M* (*SD*)	L2 *M* (*SD*)	L3 *M* (*SD*)	L1 *M* (*SD*)	L2 *M* (*SD*)	L3 *M* (*SD*)
Accuracy	9.77 (0.57)	9.83 (0.46)	9.80 (0.48)	4.14 (1.85)	4.72 (1.41)	4.03 (1.43)	8.359 (1.00)	8.35 (1.46)	8.29 (1.53)
RT	2,501.40 (1,371.64)	2,386.15 (822.01)	2,410.40 (732.52)	3,108.47 (1,332.62)	3,118.69 (1,259.09)	3,199.12 (1,558.57)	4,331.24 (2,123.76)	4,440.94 (2,123.76)	4,498.18 (2,299.87)

Note: Levels of difficulty; RT = response time in milliseconds.

Table 3. Mean and *SD* of total accuracy and response time in the Coin in Hand-Extended Version (CIH-EV) for the three groups

	Control group		Feigning group		Clinical group		*F*
	M (*SD*)	Range	*M* (*SD*)	Range	*M* (*SD*)	Range	
Accuracy	29.40 (0.19)	26–30	12.90 (0.58)	7–18	25.00 (0.81)	17–29	312.78*
RT	2,373.38 (135.03)	1,291.00–4,607.00	2,999.55 (182.41)	1,808.00–5,293.00	4,205.62 (424.48)	1,915.00–8,874.50	14.48*

Note: RT = response time in milliseconds.
*$p < .01$.

time for the control group (M = 2,432.65) over the feigning group (M = 3,142.09), and both groups gave faster responses than clinical group (M = 4,423.45).

Difficulty levels were then grouped, and total accuracy and response time were analyzed according to the group. For accuracy, there was a group effect, $F(2, 75) = 312.78$, $p > .001$, $\eta_p^2 = .89$, with the feigning group performing with lower scores (M = 12.90) than clinical group (M = 25.00) and with both performing with lower scores than the control group (M = 29.40; see Table 3).

For response time, a group effect was also observed; $F(2, 75) = 14.48$, $p < .001$, $\eta_p^2 = 0.28$. Thus, the control group had a better response time (M = 2373.38) than the feigning group (M = 2999.55) and both were faster than the clinical group (see Table 3).

Analysis of CIH-EV cutoff points for healthy older adults. The cutoff scores for healthy older adults were determined using receiver operating characteristic (ROC) analysis. For healthy control and feigning groups' comparison, the ROC curve analysis revealed that the response time was not discriminative between the two groups, with a reduced area under the curve (AUC = 0.30). However, the CIH-EV performance accuracy comparison between the two groups revealed a good discriminative ability (Difficulty Level 1: AUC = 0.98; Difficulty Levels 2 and 3: AUC = 1.00). The most appropriate cutoff point was the same for all difficulty levels (≤8), with sensitivity between 97% and 100% and specificity between 93% and 97% (see Table 4).

The ROC curve analysis comparing control and feigning groups for the total number of hits revealed an excellent AUC (1.00). Therefore, cutoff points were determined as a function of this variable. A cutoff score ≤26 was selected for the healthy older adults, with 100% sensitivity and 97% specificity (see Table 4).

Analysis of CIH-EV cutoff points for older adults with dementia. The ROC curve analyses were performed to determine appropriate cutoff points for older adults with dementia. Comparison between feigning and clinical groups showed that the difficulty levels were discriminative (AUC = 0.95 for Level 1; AUC = 0.96 for Levels 2 and 3), thus the most appropriate cutoff points were selected. For Difficulty Levels 1 and 3, a cutoff point ≤5 represented 94% sensitivity and a specificity of 79% and 83%, respectively. For Level 2, a cutoff score ≤6 revealed 88% specificity and 97% sensitivity (see Table 5). Due to the fact that these results represent a clinical non-litigating sample, the negative predictive power (NPP) and positive predictive power (PPP) were calculated using the sensitivity and specificity of our sample as well as the prevalence of our sample. In addition, NPP and PPP were calculated using the prevalence rate of a non-litigating clinical sample of patients with memory complaints (Zenisek, Millis, Banks, & Miller, 2016).

For the comparison between the feigning and clinical groups regarding performance accuracy (i.e., the total number of hits), the ROC curve analysis showed an excellent AUC (0.99). A cutoff score ≤17 for older adults with dementia was selected, with 94% of sensitivity and 97% of specificity (see Table 5).

Table 4. Sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP) of Coin in Hand-Extended Version (CIH-EV) cutoff points for healthy older adults compared to the feigning group

Difficulty level	Hits	Sensitivity	Specificity	PPP	NPP
1	<6	0.93	1	1	0.94
	<7	0.97	1	1	0.97
	≤8	**0.97**	**0.93**	**0.93**	**0.97**
	≤9	0.97	0.83	0.85	1
2	<6	0.97	1	1	0.97
	<7	1	1	1	1
	≤8	**1**	**0.97**	**0.97**	**1**
	≤9	1	0.87	0.88	1
	<6	0.83	1	1	1
3	<7	1	1	1	1
	≤8	**1**	**0.97**	**0.97**	**1**
	≤9	1	0.83	0.85	1
Total	<25	1	1	1	1
	≤26	**1**	**0.97**	**0.97**	**1**
	≤27	1	0.90	0.91	1
	≤28	1	0.87	0.88	1

Note: Bold corresponds to the values associated with the selected cutoff points.

Table 5. Sensitivity, specificity, positive predictive power (PPP) and negative predictive power (NPP) of Coin in Hand-Extended Version (CIH-EV) cutoff points for older adults with dementia compared to the feigning group

Difficulty level	Hits	Sensitivity	Specificity	Without prevalence rate		With prevalence rate, our sample		With prevalence rate of non-litigating dementia patients	
				PPP	NPP	PPP	NPP	PPP	NPP
1	≤6	0.94	0.93	0.96	0.88	0.46	0.49	0.38	0.41
	≤7	**0.94**	**0.97**	**0.97**	**0.94**	**0.66**	**0.50**	**0.58**	**0.42**
	≤8	0.53	0.66	0.78	0.89	0.09	0.08	0.07	0.06
2	≤5	0.94	0.66	0.95	0.62	0.15	0.41	0.11	0.33
	≤6	**0.88**	**0.97**	**0.93**	**0.94**	**0.65**	**0.34**	**0.57**	**0.27**
	≤7	0.77	1	0.88	1	1	0.21	1	0.16
3	≤5	0.94	0.83	0.96	0.76	0.26	0.46	0.20	0.38
	≤6	**0.94**	**1**	**0.97**	**0.94**	**1**	**0.51**	**1**	**0.43**
	≤7	**0.77**	**1**	**0.88**	**1**	**1**	**0.21**	**1**	**0.16**
Total	≤16	1	0.83	1	0.77	0.27	1	0.21	1
	≤17	**0.94**	**0.97**	**0.97**	**0.94**	**0.66**	**0.50**	**0.58**	**0.42**
	≤18	0.88	1	0.94	1	1	0.34	1	0.27

Note: Bold corresponds to the values associated with the selected cutoff points.

Convergent validity analysis. The CIH-EV convergent validity with other performance validity measures was analyzed using a Pearson's correlation coefficient. Regarding the Rey-15IMT, there was a low positive correlation between CIH-EV and Free Recall Trial (r = .24, p < .05) and a moderate positive correlation with Combined Result (r = .62, p < .001). With RDS, the CIH-EV showed a strong positive correlation (r = .73, p < .01), and finally with TOMM, very strong positive correlations were found for both Trial 1 (r = .92, p <. 01) and Trial 2 (r = .93, p < .01). Thus, the CIH-EV presents good convergent validity with the measures used to assess performance validity.

Additionally, the percentages of true negatives (i.e., participants correctly identified as controls) and false positives (i.e., incorrectly identified as simulators) were analyzed. Percentages were compared using the cutoff scores proposed for older Portuguese adults for the TOMM (e.g., Fernandes, 2009), Rey-15IMT (Simões et al., 2010), and RDS (Pinho, 2012) and the cutoff point recommended in the present study for the CIH-EV. The CIH-EV and Rey-15IMT's Combined Result were found to be the instruments with the lowest percentage of false positives (5.9% and 0.0%, respectively). However, comparing the two measures, the CIH-EV correctly detected more simulators (96.6%) and controls (100.0%; see Table 6).

Table 6. Percentages of true negatives and false positives comparison between the tests

Group	Cutoff point	Real condition	
		Control	Feigning
Control			
CIH-EV_total	≤26	100.0%	0.0%
TOMM1	≤33	100.0%	0.0%
TOMM2	≤45	100.0%	0.0%
Rey_FR	5.70	48.4%	51.6%
Rey_CR	11.70	76.7%	23.3%
RDS	≤6	96.7%	3.3%
Feigning			
CIH-EV_total	≤26	3.4%	96.6%
TOMM1	≤33	6.9%	93.1%
TOMM2	≤45	0.0%	100%
Rey_FR	5.70	96.6%	3.4%
Rey_CR	11.70	6.9%	93.1%
RDS	≤6	20.7%	79.3%
Clinical			
CIH-EV_total	≤17	94.1%	5.9%
TOMM1	≤33	70.6%	29.4%
TOMM2	≤45	35.3%	64.7%
Rey_FR	1.90	70.6%	29.4%
Rey_CR	2	100.0%	0.0%
RDS	≤6	64.7%	35.3%

Note: CIH-EV_total = CIH-EV total number of hits; Rey_FR = REY-15IMT Free Recall Trial; Rey_CR = REY-15IMT Combined Result; TOMM1 = TOMM's first trial; TOMM2 = TOMM's second trial. CIH-EV = Coin in Hand-Extended Version.

Discussion and Conclusions

The present study aimed to validate a PVT, the CIH-EV (Daugherty et al., 2019) with healthy older adults and older adults diagnosed with dementia, using a simulation paradigm (Rogers, 2008). Particularly, it aimed to: analyze the relationship between the sociodemographic variables and CIH-EV performance of the three groups (i.e., healthy control, healthy feigning, and clinical group); run a comparative study of the CIH-EV's difficulty levels and total performance regarding accuracy and response time between the three groups; analyze the CIH-EV cutoff points for healthy older adults and those diagnosed with dementia, determining their sensitivity and specificity; and conduct a convergent validity study of the CIH-EV with the Portuguese versions of the TOMM (Fernandes et al., 2009), Rey-15-IMT (Simões et al., 2010), and RDS (Pinho, 2012).

In order to achieve these goals, healthy controls and participants with dementia were instructed to perform as well as possible, whereas healthy simulators were instructed to feign memory impairment typical of dementia to obtain some form of pension, allowance, or early retirement. All participants were assessed with a neuropsychological protocol consisting of a sociodemographic questionnaire, the MMSE (Folstein et al., 1975; Guerreiro et al., 1994), MoCA (Freitas et al., 2011; Nasreddine et al., 2005), Rey-15IMT (Rey, 1964; Simões et al., 2010), TOMM (Fernandes, 2009; Tombaugh, 1996), WAIS-III Digit Span subtest (Wechsler, 2008), and CIH-EV (Daugherty et al., 2019).

Regarding the first goal of the study, we found that the feigning and control groups did not differ in cognitive performance (MMSE and MoCA). This result was expected (H1) because participants were randomly assigned to the conditions. By contrast, expected differences (H2) were found between feigning and clinical groups because the latter had a dementia diagnosis. The two groups differed in residence type, as clinical participants were institutionalized. Although feigning and clinical groups should not differ in the remaining demographic variables, the clinical group was significantly older. This difference was also reported by Technier and Wagner (2004) with a clinical sample of older adults with dementia. Indeed, in older and institutionalized (Matthews & Dening, 2002) older adults, there is a higher prevalence of dementia, with a higher incidence in women (Alzheimer's Society, 2016). Thus, this may explain the higher mean age of the group with dementia, which was also mostly women and institutionalized. However, this difference provides more robustness to the results because despite having a dementia diagnosis and being significantly older, the clinical group performed better than the healthy feigning group. This result will be later discussed with more detail.

Additionally, as hypothesized (H3), the CIH-EV showed insensitivity to all sociodemographic variables and neurocognitive functioning, both in healthy older adults and those with dementia. Schroeder and colleagues (2012) also found this using the

original version of the CIH (Kapur, 1994) with healthy older adults. These results provide solid evidence to consider the CIH-EV as a valid instrument for feigning detection.

For the second goal, a comparative study of the CIH-EV's three difficulty levels and total performance between the three groups was performed. As expected (H4), control and clinical groups performed better than the feigning group, both on CIH-EV difficulty levels and total hits. The fact that there were no differences in the performance accuracy between difficulty levels suggests that for older adults it is possible to administer just one of the three trials. On the one hand, the CIH-EV had already demonstrated a good discriminative ability between the control and feigning groups of college students (Daugherty et al., 2019). In this study, feigning participants also showed significantly less accurate performance between the difficulty levels and in total performance. On the other hand, using the original version of CIH (Kapur, 1994) with healthy older adults (controls and simulators) and older adults with dementia, Yeh and colleagues (2019) already reported the superior performance of control and clinical groups over simulators. Therefore, our study replicates the pattern of results obtained with the original version of the CIH in healthy older adults and those with dementia, using the extended computerized version of the test (CIH-EV; Daugherty et al., 2019). Results also corroborate the ceiling effect assumption of this test, as older adults with dementia showed more accurate results than healthy simulators. The inference of intentional feigning behaviors associated with significantly lower performance (Vickery et al., 2001) was also corroborated. These aspects reinforce the suitability of the CIH-EV as an effective PVT. However, the clinical group had significantly slower response times than the healthy groups. This may be due to the fact that cognitive functions such as processing and execution speed are significantly more impaired in this sample and thus may influence response time (e.g., Gainotti, Marra, & Villa, 2001).

Regarding the third goal of this study, the cutoff points for healthy older adults were, as expected (H5), the same or closer to those determined for young adults in the discrimination between feigning and control groups (Daugherty et al., 2019). For difficulty levels, the most appropriate cutoff point was the same for the three levels ($\leq$8), with sensitivity between 97% and 100% and specificity between 93% and 97%. In fact, this was the same cutoff point suggested for college students (Daugherty et al., 2019), with sensitivities between 84% and 95% and 95% of specificity. Using the original version of CIH (Kapur, 1994), Yeh and colleagues (2019) also determined a cutoff point $\leq$8 for healthy older adults. Regarding the total number of hits, the cutoff point suggested by Daugherty and colleagues (2019) for young adults ($\leq$27) revealed 95% of sensitivity and specificity. Using this cutoff for healthy older adults in our study is also valid, with 100% of sensitivity and 90% of specificity. However, by choosing a lower, but close cutoff point ($\leq$26), it is possible to increase specificity to 97% and maintain sensitivity at 100%. Thus, although the same cutoff point can be chosen, a more conservative cutoff of $\leq$26 is proposed, giving a greater precedence to specificity compared to sensitivity as has been suggested in the literature (e.g., Reznek, 2005). Of note is that the similarity between the cutoff points for both healthy samples (young adults and older adults) corroborates the evidence that the CIH (Kapur, 1994) is insensitive to variables such as age (e.g., Ferreira et al., 2017; Schroeder et al., 2012), a result also found in this study for the extended version of the instrument (CIH-EV; Daugherty et al., 2019).

Cutoff points were also determined for older adults diagnosed with dementia. As expected (H6), they were lower than those suggested for healthy controls. For the first difficulty level, the most appropriate cutoff point was $\leq$7, representing 94% sensitivity and 97% specificity. For the second and third difficulty levels, the selected point ($\leq$6) represented 88% sensitivity and 97% specificity and 94% sensitivity and 100% specificity, respectively. For total accuracy in older adults with dementia, a lower cutoff point ($\leq$17) was selected compared to the healthy older adults ($\leq$26), showing 94% sensitivity and 97% specificity. In fact, the need for an adjusted cutoff point for the clinical group reflects the inverse relationship between diagnostic accuracy and the degree of cognitive impairment reported in the literature (e.g., McGuire et al., 2019). However, for several tests, the cutoff adjustment becomes impracticable as it represents an extremely reduced sensitivity (Dean et al., 2009). This is not the case for the CIH-EV because it showed sensitivity and specificity values that largely meet the criteria proposed by Sugarman and Axelrod (2015) of at least 50% sensitivity and 90% specificity.

Finally, according to the last goal of the study, the CIH-EV (Daugherty et al., 2019) revealed good convergent validity with the other performance validity measures (Rey-15IMT, TOMM, and RDS) as we hypothesized (H7). However, a low correlation with the Rey-15IMT was found. This result is not surprising, considering the Rey-15IMT has low sensitivity values, which limits its use in feigning detection (e.g., Reznek, 2005). This is also the reason why this test has a low correlation with other PVTs (e.g., TOMM; Bailey, Soble, & O'Rourke, 2018). Higher correlations were instead found between the CIH-EV and the TOMM. This finding is similar to those reported by Yeh and colleagues (2019) in healthy older adults and those with dementia, which is likely due to the proven diagnostic accuracy of the TOMM. In addition, moderate correlations between the CIH-EV and the RDS may suggest that this embedded performance validity measure has good feigning detection ability, as has been reported by Miele, Gunner, Lynch, and McCaffrey (2012). Daugherty and colleagues (2019) also found good convergent validity of the CIH-EV with these instruments. However, previous studies suggest that the TOMM (e.g., Rudman, Oyebode, Jones, & Bentham, 2011), RDS (e.g., Dean et al., 2009), and Rey-15IMT (e.g., Fazio et al., 2019) show a reduced diagnostic accuracy when applied to groups with dementia, leading to an unacceptably high percentage of false positives (Rudman et al.,

2011). In this study, the CIH-EV was expected (H7) to have a lower percentage of false positives than the remaining tests. In fact, in healthy older adults, the CIH-EV correctly detected 100% of controls and 96.6% of simulators, as well as the TOMM, both with very high specificities and sensitivities. However, for older adults with dementia, the TOMM was found to have an unacceptably high percentage of false positives for both trials (29.4% and 64.7%, respectively, when up to 10% is recommended; Sugarman & Axelrod, 2015). Similarly, although the Rey-15IMT's Combined Result showed 0% false positives for the clinical group, it revealed a high percentage (23.3%) for the healthy older groups (23.3%). Also, analyses showed the inadequacy of the free recall trial among healthy older adults, as it did not detect feigners in the feigning group (3.4%) and incorrectly detected feigners in the control group (51.6%). The same happened for older adults with dementia, with 29.4% of false positives. Regarding the RDS, although for healthy older adults, it showed appropriate values of specificity (96.7%) and sensitivity (79.3%), for those with dementia, it revealed an unacceptably high percentage of false positives (35.3%). Thus, the TOMM and RDS proved to be effective for feigning detection in healthy older adults but not in older adults with dementia. For the clinical group, the Rey-15IMT's Combined Result was adequate but not for healthy older adults. These results support the clinician's need to know the type of groups in which PVTs are appropriate and those in which they should not be administered (Dean et al., 2009) so that the feigning detection can be valid and accurate. Also, results demonstrated that the CIH-EV was the only test showing high specificity and sensitivity values as well as low percentages of false positives in both healthy and demented older adults. Together, these results suggest that this test has a good diagnostic accuracy for feigning detection in healthy young adults (Daugherty et al., 2019), healthy older adults, and older adults with dementia.

Although in an analog study on the CIH-EV, the test proved high cultural sensitivity, it must be noted that the PVT performance may vary across the cultures (Nijdam-Jones, Rivera, Rosenfeld, & Arango-Lasprilla, 2017). Further, PVT performance may vary by the clinical group and level of cognitive impairment. As such, we highly caution against using the cutoff scores suggested in this study for populations that are unrepresentative of this sample in terms of culture, health status, and involvement in judicial proceedings.

The present study has some limitations. First, the clinical sample is small and mostly women, which could have implications for the generalization of results. Second, the sample did not have a clinical group that met Slick and Sherman's criterion (2013), that is, individuals with suspected simulation, external incentive, or secondary gain. In fact, the feigning group was instructed to feign, and some authors (e.g., Rogers, 2008) claim that because they do not have a secondary gain to motivate them, participants do not feign exactly, which makes the generalization of results difficult. However, authors like Haines and Norris (1995) added a financial incentive to the feigning group and found that there were no differences between incentivized and instructed simulators. Nevertheless, caution should be taken as research shows that the existence and level of financial incentives are not significant (Elhai et al., 2007). This may be due to the fact that real simulators are exposed to other variables in real-life contexts that participants instructed to simulate do not experience, such as distress or fear of being caught. Finally, predictive values (PPP and NPP), despite being extremely useful when making clinical decisions, have the limitation of depending on how the frequent simulation behaviors are in the population under study. However, to our knowledge, there are no data regarding the prevalence rates of insufficient effort/simulation in Portuguese population and therefore we could not address this question.

For future research, it would be interesting to investigate whether there are differences in test performance as a function of the type of dementia, including clinical groups with different levels of cognitive impairment, in order to understand at what deterioration level the cutoff is no longer appropriate. Moreover, including a fourth group of clinical feigners, as proposed by Rogers (2008), in future studies will increase the generalization of results.

Conclusion

In summary, the CIH-EV validation with healthy older adults and those diagnosed with dementia was pertinent, allowing for the existence of an additional PVT in Portugal. Specifically, the CIH-EV was insensitive to sociodemographic variables and neurocognitive functioning in healthy older adults and those with dementia; control and clinical groups were found to perform more accurately on the CIH-EV than the feigning group, with the clinical group showing longer response times; cutoff points were determined for healthy older adults ($\leq$26) and for older adults diagnosed with dementia ($\leq$17), both with high sensitivity and specificity values; good convergent validity was found with other performance validity measures (TOMM and RDS); and finally, it proved to be the instrument with the best diagnostic accuracy in the three groups tested.

Conflict of Interest

None declared.

References

Alzheimer's Society. (2016). *Risk factors for dementia (Factsheet 450LP)*. Author. United Kingdom.

Bailey, K. C., Soble, J. R., & O'Rourke, J. J. (2018). Clinical utility of the Rey 15-item test, recognition trial, and error scores for detecting noncredible neuropsychological performance in a mixed clinical sample of veterans. *The Clinical Neuropsychologist*, *32*(*1*), 119–131. doi: 10.1080/13854046.2017.1333151.

Boone, K. B., Salazar, X., Lu, P., Warner-Chacon, K., & Razani, J. (2002). The Rey 15-item recognition trial: A technique to enhance sensitivity of the Rey 15-item memorization test. *Journal of Clinical and Experimental Neuropsychology*, *24*(*5*), 561–573. doi: 10.1076/jcen.24.5.561.1004.

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., et al. (2005). Symptom validity assessment: Practice issues and medical necessity NAN policy & planning committee. *Archives of Clinical Neuropsychology*, *20*(*4*), 419–426. doi: 10.1016/j.acn.2005.02.002.

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. doi: 10.1080/13803395.2016.1223278.

Daugherty, J., Querido, L., Quiroz, N., Wang, D., Hidalgo-Ruzzante, N., Fernandes, S., et al. (2019). The coin in hand-extended version: Development and validation of a multicultural performance. *Assessment*, *28*(*1*), 186–198. doi: 10.1177/1073191119864652.

Dean, A. C., Victor, T. L., Boone, K. B., Philpott, L. M., & Hess, R. A. (2009). Dementia and effort test performance. *The Clinical Neuropsychologist*, *23*(*1*), 133–152. doi: 10.1080/13854040701819050.

Elhai, J. D., Butcher, J. J., Reeves, A. N., Baugher, S. N., Gray, M. J., Jacobs, G. A., et al. (2007). Varying cautionary instructions, monetary incentives, and comorbid diagnostic training in malingered psychopathology research. *Journal of Personality Assessment*, *88*(*3*), 328–337. doi: 10.1080/00223890701332136.

Faust, D., Hart, K., & Guilmette, T. J. (1988). Pediatric malingering: The capacity of children to fake believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *56*, 578–582.

Fazio, R. L., Faris, A. N., & Yamout, K. Z. (2019). Use of the Rey 15-item test as a performance validity test in an elderly population. *Applied Neuropsychology: Adult*, *26*(*1*), 28–35. doi: 10.1080/23279095.2017.1353994.

Ferguson, K. E. (2004). Detecting malingering in forensic neuropsychological evaluations in litigants with mild traumatic brain injury. In *Handbook of forensic psychology*, (pp. 301–314). Academic Press. doi: 10.1016/B978-012524196-0/50015-2.

Fernandes, S. P. (2009). *Test of memory malingering (TOMM): Estudos de validação em adultos idosos com declínio cognitivo ligeiro*. https://estudogeral.uc.pt/handle/10316/15764. Estudo Geral. Universidade de Coimbra, Coimbra.

Ferreira, I., Gomes, F., Moreira, I., Silva, A., & Cavaco, S. (2017). Measuring the validity of neuropsychological assessment in patients with multiple sclerosis or systemic lupus erythematosus. *Revista Evidências*, *2*, 27–38.

Folstein, M., Folstein, M., & McHugh, P. (1975). Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(*3*), 189–198.

Frederick, R. I., & Speed, F. M. (2007). On the interpretation of below-chance responding in forced-choice tests. *Assessment*, *14*(*1*), 3–11. doi: 10.1177/1073191106292009.

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2011). Montreal Cognitive Assessment (MoCA): Normative study for the portuguese population. *Journal of Clinical and Experimental Neuropsychology*, *33*(*9*), 989–996. doi: 10.1080/13803395.2011.589374.

Freitas, S., Simões, M. R., Alves, L., & Santana, I. (2013). Montreal Cognitive Assessment (MocA): Validation study for mild cognitive impairment and Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, *27*(*1*), 37–43. doi: 10.1097/WAD.0b013e3182420bfe.

Gainotti, G., Marra, C., & Villa, G. (2001). A double dissociation between accuracy and time of execution on attentional tasks in Alzheimer's disease and multi-infarct dementia. *Brain*, *124*(*4*), 731–738. doi: 10.1093/brain/124.4.731.

Green, C. M., Kirk, J. W., Connery, A. K., Baker, D. A., & Kirkwood, M. W.. (2014). The use of the Rey 15-Item Test and recognition trial to evaluate noncredible effort after pediatric mild traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology*, *36*(*3*), 261–267. doi: 10.1080/13803395.2013.879096.

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.

Guerreiro, M., Silva, A. P., Botelho, M. A., Leitão, O., Castro-Caldas, A., & Garcia, C. (1994). Adaptação à população Portuguesa da tradução do mini mental state examination (MMSE). *Revista Portuguesa de Neurologia*, *1*, 9–10.

Haines, M. E., & Norris, M. P. (1995). Detecting the malingering of cognitive deficits: An update. *Neuropsychology Review*, *5*(*2*), 125–148.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., Millis, S. R., & Conference Participants. (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. doi: 10.1080/13854040903155063.

Jasinski, L. J., Berry, D. T., Shandera, A. L., & Clark, J. A. (2011). Use of the Wechsler Adult Intelligence Scale Digit Span subtest for malingering detection: A meta-analytic review. *Journal of Clinical and Experimental Neuropsychology*, *33*(*3*), 300–314. doi: 10.1080/13803395.2010.516743.

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.

Larrabee, G. J. (Ed.) (2005). *Forensic neuropsychology: A scientific approach*. Oxford, England: Oxford University Press.

Larrabee, G. J. (2012). Performance validity and symptom validity in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(*4*), 625–630. doi: 10.1017/S1355617712000240.

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

Matthews, F. E., & Dening, T. (2002). Prevalence of dementia in institutional care. *The Lancet*, *360*(*9328*), 225–226. doi: 10.1016/S0140-6736(02)09461-8.

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A systematic review. *Archives of Clinical Neuropsychology*, *34*(*1*), 114–131. doi: 10.1093/arclin/acy012.

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are embedded validity indices equivalent to free-standing symptom validity tests? *Archives of Clinical Neuropsychology*, *27*(*1*), 10–22. doi: 10.1093/arclin/acr084.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(*4*), 695–699. doi: 10.1111/j.1532-5415.2005.53221.x.

National Statistics Institute [INE]. (2018). *Statistical yearbook of Portugal* (ISSN 0871-8741). Author. Portugal, Lisbon.

Nijdam-Jones, A., Rivera, D., Rosenfeld, B., & Arango-Lasprilla, J. C. (2017). A cross-cultural analysis of the test of memory malingering among Latin American Spanish-speaking adults. *Law and Human Behavior*, *41*(*5*), 422–428. doi: 10.1037/lhb0000250.

Organization for Economic Co-operation and Development [OECD]. (2018). *Care needed: Improving the lives of people with dementia*. Author. France, Paris.

Pinho, J. I. (2012). *Testes de Validade de Sintomas: Validação de um protocolo em amostras de adultos idosos*. https://estudogeral.uc.pt/bitstream/10316/83982/1/Ana_Beatriz_Lopes_teses__%20%281%29.pdfEstudo Geral. Universidade de Coimbra, Coimbra.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses universitaires de France.

Reznek, L. (2005). The Rey 15-item memory test for malingering: A meta-analysis. *Brain Injury*, *19*(*7*), 539–543. doi: 10.1080/02699050400005242.

Rogers, R. (2008). Detection strategies for malingering and defensiveness. In Rogers, R., & Bender, S. (Eds.), (Eds.)*Clinical assessment of malingering and deception* (pp. 18–41). New York, USA: The Guilford Press.

Rudman, N., Oyebode, J. R., Jones, C. A., & Bentham, P. (2011). An investigation into the validity of effort tests in a working age dementia population. *Aging & Mental Health*, *15*(*1*), 47–57. doi: 10.1080/13607863.2010.508770.

Schroeder, R. W., Peck, C. P., Buddin, W. H., Jr., 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 and Behavioral Neurology*, *25*(*3*), 139–143. doi: 10.1097/WNN.0b013e31826b71c1.

Simões, M. R., Santana, I., & Cerebral Aging and Dementia's Studies Group. (2015). *Scales and tests in dementia*. Lisboa: Novartis.

Simões, M. R., Sousa, L., Duarte, P., Firmino, H., Pinho, M. S., Gaspar, N., et al. (2010). Avaliação da simulação ou esforço insuficiente com o Rey 15-item memory test (15-IMT): Estudos de validação em grupos de adultos idosos. *Análise Psicológica*, *28*(*1*), 209–226.

Slick, D. J., Sherman, E. M., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *The Clinical Neuropsychologist*, *13*(*4*), 545–561. doi: 10.1080/13854046.2011.609839.

Slick, D. J., & Sherman, E. M. S. (2013). Differential diagnosis of malingering. In Carone, D. A., & Bush, S. S. (Eds.), (Eds.)*Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 57–72). New York, USA: Springer Publishing Company.

Strutt, A. M., Scott, B. M., Shrestha, S., & York, M. K. (2011). The Rey 15-item memory test and Spanish-speaking older adults. *The Clinical Neuropsychologist*, *25*(*7*), 1253–1265. doi: 10.1080/13854046.2011.609839.

Sugarman, M. A., & Axelrod, B. N. (2015). Embedded measures of performance validity using verbal fluency tests in a clinical sample. *Applied Neuropsychology: Adult*, *22*(*2*), 141–146. doi: 10.1080/23279095.2013.873439.

Suhr, J. A., & Gunstad, J. (2007). Coaching and malingering: A review. In Larrabee, G. J. (Ed.), (Ed.)*Assessment of malingered neuropsychological deficits* (pp. 287–311). Oxford, England: Oxford University Press.

Teichner, G., & Wagner, M. T. (2004). The Test of Memory Malingering (TOMM): Normative data from cognitively intact, cognitively impaired, and elderly patients with dementia. *Archives of Clinical Neuropsychology*, *19*(3), 455–464. doi: 10.1016/S0887-6177(03)00078-7.

Tombaugh, T. N. (1996). *Test of memory malingering: TOMM*. New York: Multi Health Systems.

Van Gorp, W. G., & Hassenstab, J.. (2009). Why questions regarding effort and malingering are always raised in forensic neuropsychological evaluations. In J. Morgan & J. Sweet (Eds.), *Neuropsychology of malingering casebook* (pp. 9–20). New York: Psychology Press.

Vickery, C. D., Berry, D. T., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*(*1*), 45–73. doi: 10.1093/arclin/16.1.45.

Wechsler, D. (2008). *Escala de Inteligência de Wechsler para Adultos – 3ª Edição (WAIS-III)*. Lisboa: CEGOC.

Wessely, S. (2003). Malingering: Historical perspectives. In Halligan, P. W., Bass, C., & Oakley, D. A. (Eds.), (Eds.)*Malingering and illness deception* (pp. 31–41). Oxford, England: Oxford University Press.

World Health Organization. (2020, September 21). Dementia. Retrieved from: https://www.who.int/news-room/fact-sheets/detail/dementia.

Yeh, T. C., Chou, Y. C., Weng, J. P., Yeh, H. W., Kao, Y. C., Chiang, W. S., et al. (2019). Detection of malingering in the memory of patients with dementia: A pilot study on coin-in-the-hand test in a Northern Taiwan Memory Clinic. *Journal of Medical Sciences*, *39*(*2*), 81–89. doi: 10.4103/jmedsci.jmedsci_100_18.

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. doi: 10.1093/arclin/acw025.