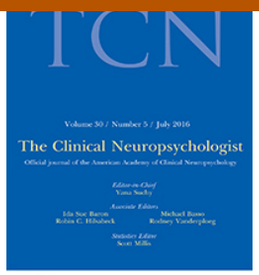

The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: http://www.tandfonline.com/loi/ntcn20

Discriminative capacity and construct validity of the Clock Drawing Test in Mild Cognitive Impairment and Alzheimer's disease

Diana Duro, Sandra Freitas, Miguel Tábuas-Pereira, Beatriz Santiago, Maria Amália Botelho & Isabel Santana

To cite this article: Diana Duro, Sandra Freitas, Miguel Tábuas-Pereira, Beatriz Santiago, Maria Amália Botelho & Isabel Santana (2018): Discriminative capacity and construct validity of the Clock Drawing Test in Mild Cognitive Impairment and Alzheimer's disease, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2018.1532022

To link to this article: https://doi.org/10.1080/13854046.2018.1532022

Published online: 13 Nov 2018.

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2018.1532022

Discriminative capacity and construct validity of the Clock Drawing Test in Mild Cognitive Impairment and Alzheimer's disease

Diana Duro[a,b,c], Sandra Freitas[c,d], Miguel Tábuas-Pereira[a,b], Beatriz Santiago[a], Maria Amália Botelho[e] and Isabel Santana[a,b]

[a]Neurology Department, Centro Hospitalar e Universitário de Coimbra, Coimbra, Portugal; [b]Faculty of Medicine, University of Coimbra, Coimbra, Portugal; [c]Center for Research in Neuropsychology and Cognitive and Behavioral Intervention, University of Coimbra, Coimbra, Portugal; [d]Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal; [e]Nova Medical School, Nova Lisbon University, Lisbon, Portugal

ABSTRACT

Objectives: The aim of this study was to analyze the psychometric and diagnostic properties of the Clock Drawing Test (CDT), scored according to the Babins, Rouleau, and Cahn scoring systems, for Mild Cognitive Impairment (MCI) and Alzheimer's disease (AD) screening, and develop corresponding cutoff scores. Additionally, we assessed the construct validity of the CDT through exploratory and confirmatory factor analysis.

Methods: We developed a cross-sectional study of ambulatory MCI and AD patients, divided in two clinical groups (450 MCI and 250 mild AD patients) and a normal control group ($N = 400$). All participants were assessed with the CDT, Mini-Mental State Examination (MMSE) and Montreal Cognitive Assessment (MoCA) for convergent validity.

Results: The selected scoring systems presented adequate validity and reliability values. The proposed cutoff scores showed 60 to 65% sensitivity and 58 to 62% specificity to identify MCI patients. The corresponding values for AD were 84 to 90% sensitivity and 76 to 78% specificity. Exploratory and confirmatory factor analysis revealed that the Babins scoring system had good construct validity and allowed us to propose a three-factor model for this system.

Conclusions: Our results confirmed the complexity of the CDT and support it as a cognitive screening instrument particularly sensitive to AD. The use of the CDT with MCI patients should be interpreted with more caution due to the lower sensitivity and specificity for milder forms of cognitive impairment.

ARTICLE HISTORY
Received 16 March 2018
Accepted 27 September 2018

KEYWORDS
Neuropsychological tests; cognitive screening; Mild Cognitive Impairment; Alzheimer's disease; exploratory factor analysis; confirmatory factor analysis

Introduction

The Clock Drawing Test (CDT) was created to assess visuospatial functions related to the parietal lobes (Battersby, Bender, Polack, & Kahn, 1956). Forthcoming studies

CONTACT Diana Duro diana.duro@gmail.com Neurology Department, Centro Hospitalar e Universitário de Coimbra, Praceta Mota Pinto, Coimbra 3000-075, Portugal.

suggested the involvement of other cognitive functions, namely symbolic and graphomotor representation, auditory linguistic abilities, executive functions (organization, planning, and parallel processing), hemiattention, semantic memory, and conceptual abilities (Cosentino, Jefferson, Chute, Kaplan, & Libon, 2004; Freedman et al., 1994; Libon, Malamut, Swenson, Sands, & Cloud, 1996; Mendez, Ala, & Underwood, 1992; Rouleau, Salmon, Butters, Kennedy, & McGuire, 1992; Shulman, 2000; Strauss, Sherman, & Spreen, 2006).

One of the most common uses of the CDT in the last 20 years has been the differentiation of the cognitively normal from groups with cognitive impairment. This remains a topic of international interest, reflected in the numerous studies continuously published about the CDT capacity to detect Mild Cognitive Impairment (MCI) and multiple forms of cognitive impairment and dementia (e.g. Aprahamian, Radanovic, Nunes, Ladeira, & Forlenza, 2014; Cahn-Weiner et al., 2003; Duro, Tábuas-Pereira, Freitas, Santiago, Botelho, & Santana, 2018; Lowery et al., 2003; Mazancova, Nikolai, Stepankova, Kopecek, & Bezdicek, 2017; Parsey & Schmitter-Edgecombe, 2011; Ricci et al., 2016; Rubinová et al., 2014; Tan, Herrmann, Mainland, & Shulman, 2015; Terwindt, Hubers, Giltay, van der Mast, & van Duijn, 2016; Viscogliosi, Chiriac, Andreozzi, & Ettorre, 2016; Yoo & Lee, 2016; Vyhnálek et al., 2017).

MCI is a transitional entity between normal aging and Alzheimer's disease (AD), which makes the discrimination between normal aging and pathology a frequently difficult challenge (Petersen et al., 1999; Petersen et al., 2009; Albert et al., 2011). It is considered both an incipient stage of dementia and a situation of risk for the development of the disease, though this progression does not always occur. About 10–15% per year and 80% over 6 years of these patients develop some type of dementia (Petersen et al., 2009) and some predictors of conversion from MCI to AD have been identified, including neuropsychological data, neuroimaging and biomarkers, alone or in combination. Objectively, MCI has been clinically defined as a self or informant-reported cognitive complaint and an objective cognitive impairment that surpasses what is expectable in subjects with a certain age and education, while functional activities of daily living remain relatively intact (Petersen et al., 1999; Albert et al., 2011).

The aim of this study was to conduct a validation study of the Babins, Slater, Whitehead, and Chertkow (2008), Cahn et al. (1996) and Rouleau et al. (1992) scoring systems for MCI and AD screening. We conducted an analysis of the psychometric properties and diagnostic accuracy of the selected scoring systems, and established optimal cut-off scores for the detection of MCI and AD, as compared to healthy control subjects. Additionally, we analyzed the CDT construct validity through exploratory and confirmatory factor analysis.

Methods

Participants and procedures

We selected a convenience sample of MCI and mild AD patients previously referred for comprehensive neuropsychological assessment at the Neuropsychology Laboratory of a Portuguese tertiary center. MCI patients included in this study were of the amnestic type and the diagnosis was made in accordance with the criteria defined by

Petersen et al. (2001) and more recently the framework for MCI due to AD, proposed by NIA-AA criteria (Albert et al., 2011). Diagnostic investigation included a standard clinical evaluation, an extensive cognitive and staging assessment, laboratory tests, imaging studies (CT or MRI and SPECT), CSF analysis, APOE genotyping and eventually PiB–PET. At baseline, a neurologist completed a medical history with the patient and the caregiver and conducted a general physical, neurological and psychiatric examination, as well as a comprehensive diagnostic battery, which included: (1) cognitive instruments as the Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), Portuguese version (Guerreiro, Silva, Botelho, Leitão, & Garcia, 1994); the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005), Portuguese version (Simões et al., 2008); the Alzheimer Disease Assessment Scale-Cognitive (ADAS-Cog; Mohs, Rosen, & Davis, 1983; Rosen, Mohs, & Davis, 1984), Portuguese version (Guerreiro, Fonseca, & Barreto, 2003); and a comprehensive neuropsychological battery with normative data for the Portuguese population (the Lisbon Battery for Assessment of Dementia, Portuguese acronym BLAD; Guerreiro, 1998) (data not shown); (2) standard staging scales which provided objective information about individual performance in various domains, including the Clinical Dementia Rating scale (CDR; Berg, 1988), Portuguese version (Garret et al., 2008) for global staging; the Disability Assessment for Dementia (DAD; Gelinas, Gauthier, McIntyre, & Gauthier, 1999), Portuguese version (Leitão, 2008) for evaluation of functional status; the Neuropsychiatric Inventory (NPI; Cummings, 1997), Portuguese version (Leitão & Nina, 2003) to characterize the psychopathological profile; and the Geriatric Depression Scale (GDS-30; Yesavage et al., 1983), Portuguese version (Barreto, Leuschner, Santos, & Sobral, 2008) to exclude major depression.

All available information (baseline cognitive tests, staging scales, clinical laboratory, and imaging studies) was used to reach a consensus diagnosis. The inclusion criteria for amnestic MCI were those proposed by Petersen et al. (2001) and were operationalized as follows: (1) subjective complaints of memory decline (reported by the subject or an informant); (2) objective memory impairment (considered when scores on standard Wechsler memory tests were >1.5 *SD*s below age/education adjusted norms) with or without deficits in other cognitive domains; (3) normal global cognition suggested by normal scores in the MMSE and MoCA using the Portuguese cut off scores (Guerreiro et al., 1994; Freitas, Simões, Alves, & Santana, 2011); (4) largely normal daily life activities, evaluated with a functional scale-DAD; (5) absence of dementia, indicated by a CDR rating of 0.5. As exclusion criteria for enrolment, we considered a MMSE score <19; a significant underlying medical or neurological illness revealed by laboratory tests or imaging; a relevant psychiatric disease, including major depression, suggested in the medical interview and confirmed by the GDS; CT or MRI demonstration of significant vascular burden (Roman et al., 1993).

As for the mild AD clinical group, diagnosis was established according to the Diagnostic and Statistical Manual of Mental Disorders-fourth edition criteria for dementia (DSM-IV-TR; APA, 1994) and specific criteria for AD (McKhann et al., 1984, 2011). These patients should have had: (1) objective evidence of dementia by cognitive testing (using the MMSE, MoCA, and the ADAS-COG scores and qualitative evaluation, i.e. impairment of memory plus another domain); (2) a MMSE score ≥ 15; (3) a global CDR

rating from 0.5 to 1, confirming the cognitive profile of dementia and loss of autonomy.

The control group was composed by healthy volunteer subjects living in the community. In brief, they were recruited according to the following criteria: (a) informed consent; (b) Portuguese as mother language and formal education received in Portuguese schools; (c) normal score (according to age and education) on two cognitive screening instruments validated for the Portuguese population, the MMSE and the MoCA; (d) preserved independence and functionality; (e) no severe depressive symptomatology (GDS-30 $\leq$ 20); (f) no history of psychiatric, neurologic or other diseases with a negative impact in cognition; (g) no medication with a negative impact in cognition; (h) no significant motor, visual or auditory deficits with a possible negative impact in cognition; (i) no present or past history of alcoholism or drug abuse. For full description of recruitment methods, please consult Santana, Duro, Freitas, Alves, and Simões (2013).

MCI and AD patients included in this study were recruited after the diagnosis was established; the evaluation session was conducted as part of their ambulatory follow up and consisted of the CDT, MMSE, and MoCA (by order of application). The clock drawings from all participants were later scored by the same neuropsychologist according to the Babins et al. (2008), Rouleau et al. (1992), and Cahn et al. (1996) scoring systems (by this order).

An informed consent was obtained from all participants after the research aims and procedures were fully explained by a member of the research team. The present research complied with the ethical guidelines on human experimentation stated on the Declaration of Helsinki and was approved by the Portuguese Foundation for Science and Technology and by the Faculty of Medicine of the University of Coimbra Scientific Committee.

Clock Drawing Test scoring systems

The 18-points Babins scoring system assesses in detail several clock components: clock face (2 points); center (2 points); type and organization of numbers (6 points); clock hands (6 points); and global gestalt (2 points) (Babins et al., 2008). The Rouleau et al. (1992) scoring system is a 10 point quantitative system that encompasses the three major clock components: clock face (2 points), numbers (4 points), and hands (4 points). The Cahn scoring system combines the Rouleau quantitative score with a qualitative analysis of the eight types of error most commonly found in clock drawing as described by Freedman: (1) stimulus-bound response (placement of minute hand towards the number "10" instead of "2"; (2) conceptual deficit (multiple errors which reflect a difficulty in assessing specific characteristics of the clock related with the numbers or the hands); (3) perseveration (numbers beyond 12 or repetitions of the same number); (4) left hemispace neglect (all numbers placed in one side of the clock face, usually the right hemispace); (5) planning deficit (errors in the placement of 12, 3, 6, and 9); (6) nonspecific spatial error (errors in the placement of numbers with no specific pattern referred in the previous errors); (7) numbers outside the clock; (8) numbers counter clockwise (Cahn et al., 1996; Freedman et al., 1994).

All study participants were assessed by two certified and experienced neuropsychologists (DD and SF). The clock drawing instructions given to the participants were the following: "I want you to draw a round clock, place all the numbers, and set the time for ten past eleven." The words "hands" and "minutes" were preferably avoided as they constituted hints for the execution process. All clock drawings from both the clinical and control groups were scored by an experienced neuropsychologist (DD); for the purpose of analyzing interrater reliability, a subgroup of randomly selected clock drawings from MCI patients was scored by a neurologist with no previous experience in neuropsychology (except for training related with the CDT and the selected scoring methods).

Statistical analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS, version 19.0) and a p value of .05 was considered statistically significant. Descriptive statistics were used for sample's characterization and two sample t-test as well as independence chi-square (χ^2) test allowed the group comparisons. The convergent validity was determined using Pearson's correlation coefficients between the CDT scoring systems, MoCA, and MMSE scores. Interrater reliability was determined by Pearson correlations between CDT scores of two independent raters; we included Cohen's kappa as an additional measure of interrater reliability. The group differences were examined with independent samples t-test, analysis of variance (ANOVA) (with Bonferroni post hoc test) and analysis of covariance (ANCOVA).

The diagnostic accuracy of the three CDT scoring systems for the prediction of a clinical diagnosis of MCI and AD was assessed through receiver operating characteristics (ROC) curve analysis. In this analysis, the area under the curve (AUC) can vary between 0.5 and 1 and a larger AUC indicates better diagnostic accuracy. The optimal cutoff points for each scoring system that yielded the highest Youden's index were selected, with the highest value indicating maximization of sensitivity and specificity. For the analysis of the predictive value of this test we calculated, for each cutoff point, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and classification accuracy.

Exploratory factor analysis (EFA) (principal axis factor analysis with varimax rotation) was used in order to determine the factorial structure of the CDT scoring systems. The proposed models, when applicable, were analyzed through confirmatory factor analysis (CFA), conducted with Analysis of Moment Structures (AMOS) version 19.0 (Arbuckle, 2010). The following guidelines were used concerning model fit criteria and acceptable fit interpretation (Bentler & Bonett, 1980; Brown, 2006; Cochran, 1952; Cramer, 1998; Hu & Bentler, 1999; Macmann & Barnett, 1994; Maruyama, 1998; Schumacker & Lomax, 2004): Goodness-of-Fit Index (GFI) and Adjusted Goodness-of-Fit Index (AGFI) values greater than 0.90 and close to 0.95 reflected a good fit; the Comparative Fit Index (CFI) with values closer to 1.0 implied good model fit and had the advantage of reflecting the degree of fit relatively well at all sample sizes; the Incremental Fit Index (IFI) attempted to correct the possibility of the chosen model being improved (closer to 1.0) by merely adding parameters; Root Mean Square Error

Table 1. Characterization of the study sample.

	Total sample	MCI	AD	CNT	Group differences	Effect size (Cohen's *d*)
N	1100	450	250	400	–	–
Gender (% females)	698 (63.5%)	291 (64.7%)	147 (58.8%)	260 (65.0%)	χ^2= 3.033, p=.220	–
Age	69.47 (8.89) [50–91]	70.35 (8.40) [50–91]	74.14 (7.69) [51–91]	65.56 (8.46) [50–91]	$F_{(2,1097)}$=87.340, $p < .001$ CNT < MCI < AD	AD > MCI: 0.47 AD > CNT: 1.06 MCI > CNT: 0.57
Education (years of formal education)	6.69 (4.14) [1–15]	7.13 (4.42) [1–15]	6.08 (4.05) [1–15]	6.61 (3.93) [1–15]	$F_{(2,1097)}$=6.054, $p < .01$	MCI > AD: 0.25
Age of onset	–	66.90 (8.36) [45–88]	70.17 (7.61) [49–90]	N/A	$t_{(469)}$=4.251, p<.001	AD > MCI: 0.41
MMSE	26.40 (3.55) [15–30]	27.15 (2.34) [19–30]	21.63 (3.38) [15–29]	28.54 (1.41) [23–39]	$F_{(2,1097)}$=697.401, p<.001	CNT > MCI: 0.72 CNT > AD: 2.67 MCI > AD: 1.90
MoCA	19.32 (5.92) [3–30]	18.87 (4.28) [8–29]	11.03 (4.17) [3–21]	23.30 (3.59) [15–30]	$F_{(2,1097)}$=575.048, p<.001	CNT > MCI: 1.12 CNT > AD: 3.15 MCI > AD: 1.86
CDT Rouleau	6.83 (2.55) [1–10]	7.14 (2.19) [2–10]	4.38 (2.00) [1–10]	8.01 (2.16) [2–10]	$F_{(2,1097)}$=229.491, p<.001	CNT > MCI: 0.40 CNT > AD: 1.74 MCI > AD: 1.32
CDT Cahn	5.30 (3.44) [−2 − 10]	5.55 (3.04) [−2 − 10]	1.96 (2.44) [−2 − 10]	7.12 (2.84) [0–10]	$F_{(2,1097)}$=256.706, p<.001	CNT > MCI: 0.53 CNT > AD: 1.95 MCI > AD: 1.30
CDT Babins	11.58 (4.87) [0–18]	12.15 (4.33) [2–18]	6.87 (3.77) [0–16]	13.87 (3.99) [2–18]	$F_{(2,1097)}$=233.393, p<.001	CNT > MCI: 0.41 CNT > AD: 1.80 MCI > AD: 1.30

MCI, Mild Cognitive Impairment; *AD*, Alzheimer's disease; *CNT*, control group; *MMSE*, Mini-Mental State Examination; *MoCA*, Montreal Cognitive Assessment; *CDT*, Clock Drawing Test. Results are presented as mean (standard deviation) [range] except for gender (characterized by female's *n* and respective percentage).

Table 2. Analysis of covariance of Clock Drawing Test scores as a function of diagnosis.

	F	*df*	*Sig.*	Partial eta^2	Mean difference	95% C.I.
Babins	177.930	2	<.001	.246	CNT vs MCI: 1.450 ± .264	.816–2.084
					CNT vs AD: 5.934 ± .323	5.160–6.707
					MCI vs AD: 4.484 ± .299	3.768–5.200
Rouleau	170.632	2	<.001	.238	CNT vs MCI: .718 ± .140	.381–1.054
					CNT vs AD: 3.077 ± .171	2.666–3.488
					MCI vs AD: 2.359 ± .159	1.979–2.740
Cahn	192.874	2	<.001	.261	CNT vs MCI: 1.352 ± .186	.916–1.799
					CNT vs AD: 4.425 ± .228	3.879–4.971
					MCI vs AD: 3.072 ± .211	2.568–3.577

Note: $N = 1100$. F testes the effect of Diagnosis. Results for mean difference are presented as corrected mean difference ± standard error. Covariates: age and education.

of Approximation (RMSEA) values closer to 0.06 or very close to 0 suggested good model fit; finally, a χ^2 test lower value represented a better adjustment.

Results

The final sample included 1100 subjects, divided by the three groups as follows: 450 MCI, 250 AD, and 400 healthy controls. The mean age of the participants was 69.47 (±8.89) years, and 6.69 (±4.14) years of education. 63.5% of the subjects were female. There were no differences between the groups regarding gender distribution (χ^2=3.033, p=.220). As for age, the control group was slightly younger than the MCI group and they were both younger than the AD group [$F_{(2,1097)}$=87.340, p<.001]. There were also differences regarding education: the MCI group had more years of formal education than the AD group [$F_{(2,1182)}$=6.054, p<.01]. The complete characterization of the study sample is described in Table 1.

Group differences

There were significant differences between the groups in all cognitive measures, including the three CDT scoring systems (see Table 1). In order to assess the existence of group differences once the effects of age and education were controlled, we performed an ANCOVA with these variables as covariates. The results showed that the variable diagnosis maintained a significant effect (Table 2).

Psychometric properties

The CDT scoring systems had values of internal consistency ranging from .439 to .901; Cronbach's alpha values were significantly higher for the Babins system. Cronbach's alpha values according to each group and each scoring system are presented in Table 3.

The Babins scoring system presented high correlations with the Rouleau (r=.964, p< .001) and Cahn (r=.948, p< .001) scoring systems. There were also moderate significant correlations with the MMSE and the MoCA, which can be considered indicative of convergent validity. The lowest correlations were found between the CDT scoring systems and the MMSE (Table 3). Additionally, we correlated the CDT scores with the

Table 3. Psychometric properties of the Clock Drawing Test.

		Total sample ($N = 1100$)	MCI ($N = 450$)	AD ($N = 250$)	CNT ($N = 400$)
Reliability (Cronbach's alpha)	Babins	.901	.878	.833	.879
	Rouleau	.560	.439	.442	.476
	Cahn	.559	.441	.442	.469
Convergent validity (Correlation coefficients)	Rouleau/Cahn	.978	.979	.953	.972
	Rouleau/Babins	.964	.949	.941	.953
	Cahn/Babins	.948	.931	.906	.932
	Rouleau/MMSE	.606	.366	.380	.420
	Rouleau/MoCA	.711	.564	.584	.577
	Cahn/MMSE	.606	.366	.328	.422
	Cahn/MoCA	.723	.536	.549	.574
	Babins/MMSE	.608	.355	.368	.457
	Babins/MoCA	.725	.571	.583	.608
	MMSE/MoCA	.817	.647	.677	.614

MCI, Mild Cognitive Impairment; AD, Alzheimer's disease; CNT, control group; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment.
All correlation values are significant at $p<.001$ level.

MoCA subscores according to cognitive domains. We found significant correlations between all CDT scores and the six cognitive domains. Specifically, the Babins total score correlated with memory ($r=.379$, $p< .001$), visuospatial ability ($r=.813$, $p< .001$), executive functions ($r=.575$, $p< .001$), attention/working memory ($r=.507$, $p< .001$), language ($r=.461$, $p< .001$), and orientation ($r=.451$, $p< .001$). Similarly, the Rouleau score correlated with memory ($r=.365$, $p< .001$), visuospatial ability ($r=.814$, $p< .001$), executive functions ($r=.551$, $p< .001$), attention/working memory ($r=.500$, $p< .001$), language ($r=.446$, $p< .001$), and orientation ($r=.452$, $p< .001$). Finally, the Cahn score correlated with memory ($r=.402$, $p< .001$), visuospatial ability ($r=.814$, $p< .001$), executive functions ($r=.554$, $p< .001$), attention/working memory ($r=.506$, $p< .001$), language ($r=.440$, $p< .001$), and orientation ($r=.455$, $p< .001$).

Inter-rater reliability of the CDT was assessed in a subgroup of 70 randomly selected MCI subjects that were scored by two independent raters. We found high significant correlations between total scores of both raters in the Babins ($r=.897$, $p<.001$), Rouleau ($r=.895$, $p<.001$), and Cahn scoring systems (total score: $r=.871$, $p<.001$; qualitative errors: $r=.691$, $p<.001$). Cohen's κ was run as an alternative measure to determine if there was agreement between the two independent raters. The results showed that there was a fair agreement between the two raters using the Babins system ($\kappa=.289$, $p<.001$), the Rouleau system ($\kappa=.346$, $p<.001$), and the Cahn system ($\kappa=.237$, $p<.001$).

Cut-off points and diagnostic accuracy

Receiver operating characteristics (ROC) curve analysis was performed to determine the CDT diagnostic accuracy to discriminate MCI and AD patients from healthy control subjects. The Babins system had an AUC of .638 for MCI ($p<.001$; S.E.=.019; 95% C.I.=.601, .675) and .886 for AD detection ($p<.001$; S.E.=.013; 95% C.I.=.861, .911). The optimal cutoff scores for detection of MCI and AD were calculated based on Youden's index. A cutoff of $\leq$15 points showed 61% diagnostic accuracy for MCI (sensitivity

Table 4. Exploratory factor analysis: factor loadings for the rotated factors.

	Factor loading			
Item	1	2	3	Communality
Two recognizable hands	.940			.92
Hands are joined	.938			.91
Hour hand towards correct number	.895			.87
Center	.845			.74
Minute hand towards correct number	.803			.74
Size difference of hands is respected	.762			.65
Gestalt	.578	.425	.422	.69
Arrows are drawn	.488			.30
Numbers all the same		.783		.62
Numbers inside circle		.766		.59
Numbers clockwise and correct sequence		.692		.61
No missing or added numbers		.577		.55
Spacing equal (3, 6, 9, and 12)			.688	.51
Clock face			.649	.43
Spacing equal (1, 2, 4, 5, 7, 8, 10, and 11)			.629	.51
Eigenvalues	5.431	2.359	1.851	
% of variance	36.21	15.73	12.34	

$N = 1100$. Loadings $<.40$ are omitted.

=60%; specificity =62%; PPV =61; NPV =61); for AD detection, a cutoff of $\leq$12 points had 81% accuracy (sensitivity =84%; specificity =78%; PPV =79; NPV =83).

The Rouleau system had an AUC of .635 for MCI ($p<.001$; S.E.=.019; 95% C.I.=.597, .673) and .874 for AD detection ($p<.001$; S.E.=.013; 95% C.I.=.848, .901). A cutoff of $\leq$9 points showed 61% diagnostic accuracy for MCI (sensitivity =64%; specificity =58%; PPV =60; NPV =62); for AD detection, a cutoff of $\leq$7 points had 81% accuracy (sensitivity =84%; specificity =78%; PPV =79; NPV =83).

The Cahn system had an AUC of .657 for MCI ($p<.001$; S.E.=.019; 95% C.I.=.620, .694) and .897 for AD detection ($p<.001$; S.E.=.012; 95% C.I.=.873, .921). A cutoff of $\leq$8 points showed 63% diagnostic accuracy for MCI (sensitivity =65%; specificity =61%; PPV =63; NPV =64); for AD detection, a cutoff of $\leq$6 points had 83% accuracy (sensitivity =90%; specificity =76%; PPV =79; NPV =88).

For comparison purposes a similar analysis was performed for the MMSE and MoCA. The MMSE presented an AUC of .679 ($p<.001$) with 48% sensitivity and 78% specificity for MCI detection (63% diagnostic accuracy), and an AUC of .969 ($p<.001$) with 85% sensitivity and 97% specificity for AD (91% diagnostic accuracy). The MoCA revealed an AUC of .779 ($p<.001$) with 71% sensitivity and 70% specificity for MCI detection (71% diagnostic accuracy), and an AUC of .983 ($p<.001$) with 95% sensitivity and 91% specificity for AD (93% diagnostic accuracy).

Exploratory factor analysis (EFA)

Principal axis factor analysis with varimax rotation was used in order to determine the factorial structure of the three CDT scoring systems. For the Rouleau and Cahn scoring systems, only one factor could be extracted; the corresponding factors explained 57 and 60% of the variability of results for Rouleau and Cahn respectively. As for the Babins scoring system, after rotation we found three factors that explained 64% of variability: the first factor (time-setting) accounted for 36.2% of the variance; the

Table 5. Confirmatory factor analysis: adjustment indexes.

Model	χ^2	d.f.	*p*	GFI	AGFI	CFI	IFI	RMSEA
Model 1	1347.5	87	<.001	.839	.778	.896	.896	.115
Model 2	1173.1	87	<.001	.859	.805	.910	.911	.107
Model 2a	**289.1**	**71**	**<.001**	**.966**	**.942**	**.982**	**.982**	**.053**

$N = 1100$. Bold values enlighten the CFA model with best adjustment indexes.
χ^2, Chi-square; d.f., degrees of freedom; GFI, Goodness-of-Fit Index; AGFI, Adjusted Goodness-of-Fit Index; CFI, Comparative Fit Index; IFI, Incremental Fit Index; RMSEA, Root Mean Squared Error of Approximation.

Table 6. Group differences according to Babins factors based on confirmatory factor analysis.

Factor	CNT	MCI	AD	Mean difference	Sig.	95% CI	Effect size (*d*)
Time-setting	5.54 (2.862)	5.01 (3.008)	1.78 (2.553)	CNT > MCI: 0.527	.022	0.06–1.00	0.180
				CNT > AD: 3.764	.000	3.21–4.32	1.387
				MCI > AD: 3.237	.000	2.70–3.78	1.158
Display characteristics	5.43 (1.064)	4.88 (1.353)	3.34 (1.684)	CNT > MCI: 0.545	.000	0.32–0.77	0.452
				CNT > AD: 2.081	.000	1.82–2.34	1.484
				MCI > AD: 1.536	.000	1.28–1.79	1.008
Planning	2.91 (0.866)	2.26 (0.837)	1.72 (0.641)	CNT > MCI: 0.648	.000	0.51–0.78	0.763
				CNT > AD: 1.188	.000	1.03–1.34	1.562
				MCI > AD: 0.540	.000	0.39–0.69	0.724

$N = 1100$. Results are presented as mean (standard deviation).
CFA, confirmatory factor analysis; CNT, control group; MCI, Mild Cognitive Impairment; AD, Alzheimer's disease; CI, confidence interval.

second factor (display characteristics) explained 15.7% of the variance of results; finally, a third factor (planning) explained 12.3% of the variance of results. Table 4 displays the items and factor loadings for the rotated factors, with loadings less than .40 omitted to improve clarity.

Confirmatory factor analysis (CFA)

As a final measure of construct validity of the CDT, we applied CFA to the Babins scoring system. Our goal was to determine the possibility of creating alternative scoring methods based on the 18 points and explore their capacity to detect different types of cognitive impairment (instead of a single total score). As starting point for CFA, we used the factorial structure previously obtained through EFA. We analyzed different models according to several adjustment indexes. The first model (Model 1) included all items grouped according to the results of EFA in three factors. After following several modification indexes proposed by the system, we defined the final three factor model (Model 2): the items composing each factor kept the original organization except "gestalt" that was transposed to the second factor (display characteristics). The adjustment indexes are presented in Table 5; Model 2a represents the optimal model and resulted from the establishment of intercorrelations between the items (as suggested by the software).

As a final analysis, we were interested in determining if the participants' scores differed according to each factor. We performed an ANOVA with Bonferroni correction for multiple comparisons and found statistically significant differences between the three groups in all factors, following the same pattern: CNT > MCI > AD. However, the analysis of effect size showed that the differences between CNT and MCI patients were medium for display characteristics and small for time-setting. Results are presented in detail in Table 6.

Discussion

The main purpose of this study was to validate the CDT as a cognitive screening measure for MCI and AD. Specifically, we conducted a thorough validation study of three scoring systems (Babins, Rouleau, and Cahn) and determined optimal cutoff scores for the identification of MCI and AD. The results confirmed that the CDT is a valid and reliable measure for the screening of MCI and AD patients, and can be used as a cognitive screening tool when more complex and time-consuming instruments are not available. This assumption is also supported by the moderate to high correlations with the MMSE and the MoCA, including correlations with the MoCA cognitive domains, suggesting convergent validity. The CDT has a clear advantage relative to the most common cognitive screening tools: it is easy to administer and score, well accepted by patients from all age ranges, and less time-consuming. We selected quantitative and qualitative scoring systems with the same administration method and they all have been described in the literature as having good psychometric properties and good sensitivity and specificity values for MCI and AD. Also, the CDT, unlike the MMSE or MoCA, does not require any overt spoken response nor does the CDT rely much on episodic memory and temporal orientation. Moreover, while the MoCA and MMSE contain visuospatial construction items, particularly the MoCA, which includes a clock drawing, a detailed empirically-based method for analyzing errors on the visuospatial construction items on these measures does not exist.

The analysis of interrater reliability, particularly the high correlations obtained between the ratings of an experienced neuropsychologist versus a neurologist with no specific experience in neuropsychology supported the broad use of the CDT in clinical practice (specifically with these scoring systems). Our results are in line with other international studies that have suggested that the CDT has high interrater reliability regardless of the scoring system used (e.g. Shulman, 2000; Aprahamian, Martinelli, Neri, & Yassuda, 2009; Pinto & Peters, 2009; Duro et al., 2015; Jorgensen, Kristensen, Waldemar, & Vogel, 2014; Mazancova et al., 2017; Vyhnálek et al., 2017). This is a very important aspect as it is crucial to guarantee that the patient will receive the same score regardless of the health professional that administers the test.

ROC curve analysis of the selected scoring systems showed that all have a significantly higher diagnostic accuracy for AD detection (81–83%) than for MCI (58–61%). Several studies have addressed the CDT's poor diagnostic accuracy for the identification of MCI subjects (e.g. Pinto & Peters, 2009; Ehreke, Luppa, Konig, & Riedel-Heller, 2010), while others revealed more optimistic results. Parsey and Schmitter-Edgecombe (2011) used a modified version of the Rouleau system and showed that it had higher sensitivity for AD and MCI when they included the analysis of qualitative errors in the final score: they reported 39% sensitivity/88% specificity to distinguish MCI from controls and 58% sensitivity/100% specificity to distinguish AD from MCI. In the original study by Babins et al. (2008), the proposed scoring system revealed 76% specificity and 90% sensitivity in differentiating AD patients from cognitively normal subjects and 78% sensitivity in the identification of the MCI subgroup of subjects who later developed dementia. Placement of the hands proved to be the most discriminative task among the four analyzed groups, even between MCI subgroups (non progressors versus progressors), and could therefore be used as an indicator of future cognitive

decline (Babins et al., 2008). Our results with this scoring system are closer to the ones described in a German population-based longitudinal study: the authors found 60% sensitivity and 70% specificity for the Babins system, for the screening of MCI (Ehreke et al., 2010). Other recent studies continued to present similar values concerning MCI and AD detection (e.g. Mazancova et al., 2017; Reiner, Eichler, Hertel, Hoffman, & Thyrian, 2017; Vyhnálek et al., 2017) regardless of the scoring systems used. The lower sensitivity for MCI can be explained by several factors. MCI is a heterogeneous group: there can be single domain or multidomain cognitive impairment and this fact can justify the difficulty of tests such as the CDT to have better diagnostic accuracy in large samples. Also, a significant percentage of MCI patients remain stable over time, not developing AD or other forms of dementia. As we were able to confirm, even other cognitive screening tests such as the MMSE or the MoCA, despite having specific items for different cognitive domains, present sensitivity and specificity values for MCI closer to the CDT (although slightly higher for the MoCA). In future research it will be interesting to compare MCI patients with single and multidomain deficits longitudinally and analyze if there are differences between these groups regarding the development of cognitive deficits (including the performance on the CDT).

Despite having comparable diagnostic accuracy, the Babins system has revealed better psychometric properties than the Rouleau and the Cahn systems in the Portuguese population (e.g. Duro et al., 2015, Duro et al., 2018). In line with this fact, we decided to further explore the validity of these scoring systems regarding their construct validity. While the Rouleau and Cahn scoring systems resulted in a single factor solution, the Babins systems revealed a three factor solution, a fact that prompted us to conduct CFA. The practical implications of this CFA study allowed us to propose a three factor factorial structure for the Babins scoring system: Time-setting, Display Characteristics and Planning. Such levels of confirmatory model fit offer plausibility and trustworthiness of the hypothesized three factor structure (Model 2a, 3 factors), especially given the fact that the competing factor models, also tested, specifying alternative dimensionalities, presented evidently lesser levels of fit. The results of the CFA suggest possible future studies where we can consider correlating these findings with other neuropsychological tests or domains, or event neuroimaging data. Also, it showed us that different clock components may be related with similar cognitive processes. As future research, it will be interesting to confirm if we can establish a correlation with specific regions of interest or cognitive functions, eventually developing alternative scoring methods within the existent ones in order to increase the diagnostic accuracy for milder forms of cognitive impairment.

This study makes a valuable contribution by allowing a more precise and accurate use of the CDT in clinical practice, for several reasons: (1) we used thoroughly studied MCI and AD patients, having excluded patients with uncertain diagnoses; (2) we included the participants' first cognitive assessment data for analysis (after the establishment of the diagnosis by our clinical team)—our purpose was to determine the CDT's usefulness as a cognitive screening measure at an early stage; (3) we used a rigorous methodology—all participants were assessed by experienced neuropsychologists and all clock drawings were scored by a single neuropsychologist (DD), thus minimizing interpersonal bias; (4) we presented cutoff scores for MCI and AD (and not a

single cutoff score for global cognitive impairment); (5) we presented original construct validity data. This is the first known study of CFA with the CDT, particularly the Babins scoring system. It is important to acknowledge the lack of other CFA studies which would allow a more specific comparative analysis of our results.

However, this study has some limitations. The clinical groups and the control group do not match according to age and education, a consequence of the use of convenience samples from our memory clinic. There is a higher concentration of younger control subjects (although with age ranges equivalent to the clinical groups); MCI patients were also younger than AD patients, which can be explained by the natural course of the disease (MCI state usually precedes a diagnosis of AD or other forms of dementia). Also, we excluded all participants with significant depressive symptomatology, which limits the applicability of our data with this particular segment of the population.

As closing remarks, this study is part of a more global project with the main goal of validating the CDT as a cognitive screening instrument for use not only in specialized centers but in primary care where neuropsychological assessment is not available. We believe we produced important data regarding the utility of CDT as a cognitive screening instrument for amnestic MCI and AD. As shown by our results, the CDT (regardless of the scoring system used) presented a fair diagnostic accuracy for MCI; the same scoring systems revealed good to excellent diagnostic accuracy for mild AD. In face of such results, we can recommend the CDT as a complementary screening instrument for AD, while the use of this instrument with MCI patients must be made as part of a more complex assessment protocol. Once again, our data allow us to conclude that the CDT should never replace a comprehensive neuropsychological assessment but it can produce valuable data for referral purposes.

Disclosure statement

The authors declare no conflicts of interest.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (Fundação para a Ciência e Tecnologia, FCT) through a PhD fellowship (SFRH/BD/52289/2013).

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., … Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 270–279.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed. TR). Washington DC: American Psychiatric Association.

Aprahamian, I., Martinelli, J. E., Neri, A. L., & Yassuda, M. S. (2009). The clock drawing test: A review of its accuracy in screening for dementia. *Dementia & Neuropsychologia*, *3*(2), 74–81.

Aprahamian, I., Radanovic, M., Nunes, P. V., Ladeira, R. B., & Forlenza, O. O. (2014). The use of the clock drawing test in bipolar disorder with or without dementia of Alzheimer's type. *Arquivos de Neuro-Psiquiatria*, *72*(12), 913–918.

Arbuckle, J. L. (2010). *IBM SPSS AmosTM 19.0 User's guide*. Chicago: Amos Development Corporation.

Babins, L., Slater, M.-E., Whitehead, V., & Chertkow, H. (2008). Can an 18-point clock-drawing scoring system predict dementia in elderly individuals with mild cognitive impairment? *Journal of Clinical and Experimental Neuropsychology*, *30*(2), 173–186.

Barreto, J., Leuschner, A., Santos, F., & Sobral, M. (2008). Escala de Depressão Geriátrica. In Grupo de Estudos de Envelhecimento Cerebral e Demência (Ed.), *Escalas e Testes na Demência* (pp. 69–72). Lisbon: Novartis.

Battersby, W. S., Bender, M. B., Pollack, M., & Kahn, R. L. (1956). Unilateral spatial agnosia (inattention) in patients with cerebral lesions. *Brain: A Journal of Neurology*, *79*(1), 68–93.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606.

Berg, L. (1988). Clinical dementia rating (CDR). *Psychopharmacology Bulletin*, *24*(4), 637–639.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford: New York.

Cahn, D. A., Salmon, D. P., Monsch, A. U., Butters, N., Wiederholt, W. C., Corey-Bloom, J., & Barrett-Connor, E. (1996). Screening for dementia of Alzheimer type in the community: The utility of the clock drawing test. *Archives of Clinical Neuropsychology*, *11*(6), 2, 529–539.

Cahn-Weiner, D. A., Williams, K., Grace, J., Tremont, G., Westervelt, H., & Stern, R. A. (2003). Discrimination of dementia with Lewy bodies from Alzheimer disease and Parkinson disease using the clock drawing test. *Cognitive and Behavioral Neurology*, *16*(2), 85–92.

Cochran, W. G. (1952). The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, *23*(3), 315–345.

Cosentino, S., Jefferson, A., Chute, D., Kaplan, E., & Libon, D. J. (2004). Clock drawing errors in dementia: Neuropsychological and neuroanatomical considerations. *Cognitive and Behavioral Neurology*, *17*(2), 74–84.

Cramer, K. M. (1998). Analysis and comparison of WISC-III weighted sum score factor models using two independent samples. *Educational and Psychological Measurement*, *58*(5), 754–767.

Cummings, J. L. (1997). The Neuropsychiatric inventory: assessing psychopathology in dementia patients. *Neurology*, *48*, 10–16.

Duro, D., Carneiro, D., Freitas, S., Pereira, M., Santiago, B., & Santana, I. (2015). Validation studies of the clock drawing test in mild cognitive impairment. *European Journal of Neurology*, *22*(suppl 1), 496.

Duro, D., Tábuas-Pereira, M., Freitas, S., Santiago, B., Botelho, M. A., & Santana, I. (2018). Validity and clinical utility of different clock drawing test scoring systems in multiple forms of dementia. *Journal of Geriatric Psychiatry and Neurology*. doi:10.1177/0891988718774432.

Ehreke, L., Luppa, M., Konig, K. K., & Riedel-Heller, S. G. (2010). Is the clock drawing test a screening tool for the diagnosis of mild cognitive impairment? A systematic review. *International Psychogeriatrics*, *22*(01), 56–63.

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

Freedman, M., Leach, L., Kaplan, E., Winocur, G., Shulman, K. L., & Delis, D. C. (1994). *Clock drawing: A neuropsychological analysis* (vol.6). New York: Oxford University Press.

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.

Garret, C., Santos, F., Tracana, I., Barreto, J., Sobral, M., & Fonseca, R. (2008). Avaliação Clínica da Demência. In Grupo de Estudos de Envelhecimento Cerebral e Demência (Ed.), *Escalas e Testes na Demência* (pp. 17–32). Lisbon: Novartis.

Gelinas, I., Gauthier, L., McIntyre, M., & Gauthier, S. (1999). Development of a functional measure for persons with Alzheimer's disease: The disability assessment for dementia scale. *American Journal of Occupational Therapy*, *53*(5), 471–481.

Guerreiro, M. (1998). *Contributo da neuropsicologia para o estudo das demências* [Contribute of neuropsychology for the study of dementia]. Unpublished Doctoral Thesis, University of Lisbon, Lisbon, Portugal.

Guerreiro, M., Fonseca, S., & Barreto, J. (2003). Escala de Avaliação da Doença de Alzheimer. In Grupo de Estudos de Envelhecimento Cerebral e Demência (Ed.), *Escalas e Testes na Demência* (pp. 33–49). Lisbon: Novartis.

Guerreiro, M., Silva, A. P., Botelho, M. A., Leitão, O., & Garcia, C. (1994). Adaptação à população portuguesa da tradução do "Mini Mental State Examination" (MMSE). *Revista Portuguesa de Neurologia*, *1*, 9–10.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Models*, *6*(1), 1–55.

Jorgensen, K., Kristensen, M. K., Waldemar, G., & Vogel, A. (2014). The six-item Clock Drawing Test – reliability and validity in mild Alzheimer's disease. *Aging, Neuropsychology and Cognition*, *22*(3), 301–31. doi:10.1080/13825585.2014.932325.

Leitão, O. (2008). Avaliação da Incapacidade Funcional na Demência. In Grupo de Estudos de Envelhecimento Cerebral e Demência (Ed.), *Escalas e Testes na Demência* (pp. 107–110). Lisbon: Novartis.

Leitão, O., & Nina, A. (2003). Inventário Neuropsiquiátrico. In Grupo de Estudos de Envelhecimento Cerebral e Demência (Ed.), *Escalas e Testes na Demência* (pp. 67–86). Lisbon: Novartis.

Libon, D. J., Malamut, B. L., Swenson, R., Sands, L. P., & Cloud, B. S. (1996). Further analyses of clock drawings among demented and nondemented older subjects. *Archives of Clinical Neuropsychology*, *11*(3), 193–205.

Lowery, N., Giovanni, L., Mozley, L. H., Arnold, S. E., Bilker, W. B., Gur, R. E., & Moberg, P. J. (2003). Relationship between clock-drawing and neuropsychological and functional status in elderly institutionalized patients with schizophrenia. *American Journal of Geriatric Psychiatry*, *11*(6), 621–628.

Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the verbal-performance dichotomy of the Wechsler scales. *School Psychology Quarterly*, *9*(3), 161–197.

Maruyama, G. M. (1998). *Basics of structural equation modelling*. Thousands Oaks, CA; Sage.

Mazancova, A. F., Nikolai, T., Stepankova, H., Kopecek, M., & Bezdicek, O. (2017). The reliability of clock drawing test scoring systems modelled on the normative data in healthy aging and nonamnestic mild cognitive impairment. *Assessment*, *24*(7), 945–957.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's disease. *Neurology*, *34*(7), 939–944.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., … Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer & Dementia*, *7*(3), 263–269.

Mendez, M., Ala, T., & Underwood, K. (1992). Development of scoring criteria for the clock drawing task in Alzheimer's disease. *Journal of the American Geriatrics Society*, *40*(11), 1095–1099.

Mohs, R. C., Rosen, W. G., & Davis, K. L. (1983). The Alzheimer's disease assessment scale: An instrument for assessing treatment efficacy. *Psychopharmacological Bulletin*, *19*, 448–450.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., … Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699.

Parsey, C. M., & Schmitter-Edgecombe, M. (2011). Quantitative and qualitative analyses of the clock drawing test in mild cognitive impairment and Alzheimer disease: Evaluation of a modified scoring system. *Journal of Geriatric Psychiatry and Neurology*, *24*(2), 108–118. doi:10.1177/0891988711402349.

Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., … Jack, C. R. (2009). Mild cognitive impairment: Ten years later. *Archives of Neurology*, *66*(12), 1447–1455.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*(3), 303–308.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review): Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*, 1132–1142.

Pinto, E., & Peters, R. (2009). Literature review of the clock drawing test as a tool for cognitive screening. *Dementia and Geriatric Cognitive Disorders*, *27*(3), 201–213. doi:10.1159/000203344.

Reiner, K., Eichler, T., Hertel, J., Hoffman, W., & Thyrian, J. R. (2017). The Clock Drawing test: A reasonable instrument to assess probable dementia in primary care? *Current Alzheimer Research*. doi:10.2174/1567205014666170908101822.

Ricci, M., Pigliautile, M., D'Ambrosio, V., Ercolani, S., Bianchini, C., Ruggiero, C., … Mecocci, P. (2016). The clock drawing test as a screening tool in mild cognitive impairment and very mild dementia: A new brief method of scoring and normative data in the elderly. *Neurological Sciences*, *37*(6), 867–873.

Roman, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., … Scheinberg, P. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, *43*(2), 250–260.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American Journal of Psychiatry*, *141*(11), 1356–1364.

Rouleau, I., Salmon, D. P., Butters, N., Kennedy, C., & McGuire, K. (1992). Quantitative and qualitative analyses of clock drawings in Alzheimer's and Huntington's disease. *Brain and Cognition*, *18*(1), 70–87.

Rubinová, E., Nikolai, T., Marková, H., Siffelová, K., Laczó, J., Hort, J., & Vyhnálek, M. (2014). Clock drawing test and the diagnosis of amnestic mild cognitive impairment: Can more detailed scoring systems do the work? *Journal of Clinical and Experimental Neuropsychology*, *36*(10), 1076–1083.

Santana, I., Duro, D., Freitas, S., Alves, L., & Simões, M. R. (2013). The clock drawing test: Portuguese norms, by age and education, for three different scoring systems. *Archives of Clinical Neuropsychology*, *28*(4), 375–387.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.

Simões, M. R., Freitas, S., Santana, I., Firmino, H., Martins, C., Nasreddine, Z., & Vilar, M. (2008). *Montreal Cognitive Assessment (MoCA): Versão final portuguesa*. Serviço de Avaliação Psicológica, Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra. Coimbra, Portugal.

Shulman, K. (2000). Clock-drawing: Is it the ideal cognitive screening test? *International Journal of Geriatric Psychiatry*, *15*(6), 548–561.

Strauss, E., Sherman., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (vol 3). New York: Oxford University Press.

Tan, L. P., Herrmann, N., Mainland, B. J., & Shulman, K. (2015). Can clock drawing differentiate Alzheimer's disease from other dementias? *International Psychogeriatrics*, *27*(10), 1649–1660.

Terwindt, P. W., Hubers, A. A., Giltay, E. J., van der Mast, R. C., & van Duijn, E. (2016). Screening for cognitive dysfunction in Huntington's disease with the clock drawing test. *International Journal of Geriatric Psychiatry*, *31*(9), 1013–1020.

Viscogliosi, G., Chiriac, I. M., Andreozzi, P., & Ettorre, E. (2016). The metabolic syndrome predicts longitudinal changes in clock drawing test performance in older nondemented hypertensive individuals. *American Journal of Geriatric Psychiatry*, *24*(5), 359–363.

Vyhnálek, M., Rubínova, E., Marková, H., Nikolai, T., Laczó, J., Andel, R., & Hort, J. (2017). Clock drawing test in screening for Alzheimer's dementia and mild cognitive impairment in clinical practice. *International Journal of Geriatric Psychiatry*, *32*(9), 933–939.

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49.

Yoo, D. H., & Lee, J. S. (2016). Clinical usefulness of the clock drawing test applying Rasch analysis in predicting of cognitive impairment. *Journal of Physical Therapy Science*, *28*(7), 2140–2143.