ARTIGO ORIGINAL/ORIGINAL ARTICLE

Validation Study of the Toulouse-Piéron Cancellation Test for Portuguese Patients with Mild Cognitive Impairment and Alzheimer's Disease

Estudo de Validação do Teste da Barragem de Toulouse-Piéron numa Amostra de Doentes com Declínio Cognitivo Ligeiro e Doença de Alzheimer

Marisa Lima[1,4]*, Diana Duro[1,2,3], Sandra Freitas[1,4], Mário R. Simões[1,4,5], Isabel Santana[2,3]
1-Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental (CINEICC), Coimbra, Portugal.
2-Serviço de Neurologia, Centro Hospitalar e Universitário de Coimbra (CHUC), Coimbra, Portugal.
3-Faculdade de Medicina da Universidade de Coimbra (FMUC), Coimbra, Portugal.
4-PsyAssessmentLab, FPCEUC, Coimbra, Portugal.
5-Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra (FPCEUC), Coimbra, Portugal.

Informações/Informations:
Artigo Original, publicado em Sinapse, Volume 19, Número 1-2, janeiro-março · abril-junho 2019. Versão eletrónica em www.sinapse.pt
Original Article, published in Sinapse, Volume 19, Number 1-2, january-march · april-june 2019. Electronic version in www.sinapse.pt

Palavras-chave:
Doença de Alzheimer; Disfunção Cognitiva; Testes Psicológicos.

Keywords:
Alzheimer Disease; Cognitive Dysfunction; Psychological Tests.

***Autor Correspondente/ Corresponding Author:**
Marisa Lima
Centro de Investigação em Neuropsicologia e Intervenção Cognitivo Comportamental
Faculdade de Psicologia e de Ciências de Educação
Universidade de Coimbra
Rua do Colégio Novo
3000-115 Coimbra, Portugal
marisalima5@hotmail.com

Recebido / Received: *2019-01-03*
Aceite / Accepted: *2019-05-02*

Resumo

Introdução: O teste da Barragem de Toulouse-Piéron (TP) é um instrumento psicométrico clássico que permite avaliar a atenção seletiva e sustentada, assim como a velocidade de processamento, fornecendo dois índices principais: Rendimento de Trabalho (RT) e Índice de Dispersão (ID).

O objetivo do estudo foi a validação do TP para doentes com declínio cognitivo ligeiro e doença de Alzheimer, através da análise das suas propriedades psicométricas e acuidade diagnóstica da prova, permitindo o estabelecimento de pontos de corte.

Métodos: A amostra é constituída por 250 participantes, divididos em: grupo de controlo (GC) (n=100), grupo com declínio cognitivo ligeiro (DCL) (n=100) e grupo com doença de Alzheimer (DA) (n=50). Os grupos clínicos cumprem os respectivos critérios de diagnóstico internacionais estandardizados e o grupo de controlo é constituído por indivíduos cognitivamente saudáveis e inseridos na comunidade. A acuidade diagnóstica do teste foi avaliada através da análise das curvas ROC (*receiver operating characteristics*). Para cada um dos índices foram selecionados os pontos de corte que forneciam o maior valor de Youden, representando uma maximização da sensibilidade e especificidade.

Resultados: O TP revelou boas propriedades psicométricas e a pontuação total nos índices diferiu entre os três grupos ($p < 0{,}001$: GC<DCL<DA). Foram calculados os valores de sensibilidade para o grupo DCL: 53% (TP-RT) e 48% (TP-ID). Em termos de especificidade, os resultados obtidos foram: 85% (TP-RT), e 83% (TP-ID), resultando numa acuidade diagnóstica de, respectivamente, 69% (TP-RT) e 65% (TP-DI). Os mesmos índices foram calculados para o grupo DA: sensibilidade 93% (TP-RT) e 93% (TP-ID), e especificidade: 95% (TP-RT) e 82% (TP-ID). O teste revelou excelente acuidade diagnóstica para doentes com DA: 94% (TP-RT) e 88% (TP-ID). Foram então estabelecidos os pontos de corte para DA: TP-RT: < 49 pontos (AUC= 0,981) e TP-ID: > 26 pontos (AUC= 0,921).

Conclusão: Os resultados confirmam a capacidade do TP para identificar a presença de défices cognitivos em doentes com DA, apresentando uma fraca sensibilidade para o grupo DCL, numa amostra Portuguesa. Os pontos de corte são de grande utilidade quer na prática clínica quer investigacional.

Abstract

Introduction: The Toulouse-Piéron Cancellation Test (TP) is a classic psychometric instrument for the assessment of selective, sustained attention and processing speed,

providing two main outcomes: Work Efficiency (WE) and Dispersion Index (DI).

Our objective was to validate the TP for patients with mild cognitive impairment (MCI) and Alzheimer's Disease (AD) through an analysis of its psychometric properties, of its classification diagnostic accuracy and the proposal of optimal cut-off points.

Methodology: Study sample included 250 participants, divided into a control group (CG) (n=100), a mild cognitive impairment (MCI) group (n=100), and an Alzheimer's disease (AD) group (n=50). The clinical groups fulfilled standard international diagnostic criteria. Controls were community-dwelling subjects without neurological or psychiatric pathologies. The diagnostic accuracy of the TP was evaluated by the receiver operating characteristics (ROC) curve analysis. The optimal cut-off points for each index of TP that generated the highest Youden value were selected, with higher values meaning the maximization of both sensitivity and specificity.

Results: The TP revealed good psychometric indicators, and the total scores significantly differ between the three groups ($p < 0.001$: Control<MCI<AD). For the mild cognitive impairment group, the values of sensitivity were 53% (TP-WE) and 48% (TP-DI). Regarding specificity, the values were 85% (TP-WE) and 83% (TP-DI) and the diagnostic accuracy was, respectively, 69% (TP-WE) and 65% (TP-DI). Concerning Alzheimer's disease patients, the values of sensitivity were 93% (TP-WE; TP-DI) and of specificity were 95% (TP-WE) and 82% (TP-DI). The TP revealed high diagnostic accuracy for the Alzheimer's disease group, respectively, 94% (TP-WE) and 88% (TP-DI). The optimal cut-off points for Alzheimer's disease group were then established: TP-WE: < 49 points (AUC= 0.981) and TP-DI: > 26 points (AUC= 0.921).

Conclusion: Our findings confirmed the capacity of the TP to identify cognitive impairment in Alzheimer's disease patients, with poor sensitivity for MCI patients, in a Portuguese population. The cut-off points will be useful in clinical and research contexts.

Introduction

The impact of dementia on worldwide public health and its underlying economic costs is one of the major challenges for the following years.[1] The increase of life expectancy and reduced birth-rate has brought about relevant demographic changes in the last decades, leading to a worrying trend to a worldwide aged population. At the same time, age represents a main risk factor for cognitive impairment and dementia, particularly Alzheimer's disease (AD).[2,3] This form of dementia is responsible for at least 60% of all cases,[4] affecting 5% - 7% of people over the age of sixty.[5]

Attentional deficits are responsible for difficulties in information processing at various levels and may explain possible functional losses in both normal aging and dementia.[6,7] The cognitive profile of early-stage AD is mainly characterized by a deterioration of episodic memory eventually with a decline of other functions such as attention. Similar to AD, the evaluation of attention is fundamental in prodromal or pre-dementia states of dementias globally designed for mild cognitive impairment (MCI) as it may be responsible for variations in performance and constraints to an effective assessment of other cognitive functions.[8]

The Toulouse-Piéron Cancellation Test (TP)[9] was first developed in 1904 by Édouard Toulouse and is the most known and used psychometric test for the assessment of perceptive and attentional abilities. In Portugal,[10] the TP is also the third most used neuropsychological test after the Rey Complex Figure Test[11,12] and the D2 Test of Attention (D2).[13] This popularity is explained because it allows the assessment of two main attentional domains that are frequently impaired in ageing and dementia: sustained and selective attention. Additionally, a recent study comparing AD patients to cognitively healthy subjects, revealed that impairment in attentional measures comprises lower performances in divided, sustained and selective attention tasks and also in processing speed,

mechanisms which are not explained by age or educational level.[14] These findings support the utility of a complex attention tool such as the TP for the evaluation of MCI and AD patients, since it is a psychometric instrument that encompasses not only perception and attentional control but also, visual search, working memory skills, psychomotor abilities, switching, and cognitive flexibility.

We present a validation study of the TP for MCI and mild AD patients. More specifically, we conducted an exploratory analysis of its psychometric properties, considering the cognitive performance of the study groups and set up its diagnostic accuracy based on optimal cut-off points.

Methodology

Participants and procedures

We used a convenience sample composed by 250 participants (aged ≥ 45 years) distributed between three groups: (I) MCI group with 100 patients; (II) AD group with 50 patients; and (III) Control Group (CG) with 100 cognitively healthy adults. MCI and AD patients were recruited at the Memory Clinic of of the Neurology Department of a central hospital. All patients underwent a cognitive screening with the Mini-Mental State Examination (MMSE)[15,16] and the Montreal Cognitive Assessment (MoCA)[17,18] as well as a comprehensive neuropsychological assessment with the *Bateria de Lisboa para Avaliação da Demência* (BLAD).[16] A medical exam was conducted by a neurologist and routine complementary diagnostic exams were performed: laboratory analysis, Apolipoprotein E (ApoE) genotyping and structural and functional imaging exams, namely computed tomography - CT-scan or magnetic resonance imaging (MRI) and SPECT (*single photon emission computed tomography*). Most of these patients were further investigated with PiB-PET (Pittsburgh compound B-positron emission tomography), and/or cerebrospinal fluid analysis, allowing an AD diagnosis supported by biomarkers. A final diagnosis was established by a multidisciplinary team following international criteria for MCI[19,20] and probable AD (NINCDS-ADRDA criteria).[21] Considering AD, we selected patients, in mild dementia stages (MMSE ≥ 18). For MCI we only included amnesic-MCI type (a-MCI), classified as single-domain or multi-domain-MCI.

The CG was composed by cognitively healthy community-dwelling subjects. For the initial selection of participants the defined inclusion and exclusion criteria included: a) Portuguese as native language and having had at least one year of formal education; b) normal scores according to the normative values defined for the Portuguese population on the MMSE[29,30] and MoCA[31,32]; c) preserved independence and functionality on activities of daily living; d) no medication intake that could interfere with normal cognitive functioning; e) absence of neurological or psychiatric disorders; f) no significant motor, visual or auditory deficits with a possible negative influence in cognitive performance; g) no present or past history of alcoholism or drug abuse. The exclusion criteria included: illiteracy; functional deficits with a recognized influence in daily living autonomy; and clinically significant depressive symptomatology (determined by a GDS-30 score ≥ 11 points).

All subjects participated voluntarily and gave their informed written consent for the study, following the tenets of the Declaration of Helsinki, after clarification of the nature and possible implications of the study. For AD patients a written consent by the caregiver was also obtained.

Study materials

The evaluation procedure was implemented individually within a one-hour session through an established fixed order. For this particular study we developed a short protocol composed by the MMSE and MoCA (two measures of cognitive screening), followed by the TP, the Trail Making Test (TMT A/B)[22] (a double-task test that provides information about attention, visual search, eye-hand coordination, processing speed, sequencing capability and cognitive flexibility – Part A, and also evaluates executive functions – Part B), the Geriatric Depression Scale (GDS-30)[23-25] (a straightforward screening measure for depression symptomatology in elderly populations), and the Subjective Memory Complaints scale (SMC)[26-28] (a scale that aims to characterize memory complains with multiple levels of response, according with the complains severity). Additionally, in the control group we also performed the D2 Attention Test (D2), in order to further explore the convergent validity with the TP. The D2 is a timed cancellation test that allows the assessment of selective attention, processing speed, rule compliance and quality of performance, through two of its indexes: E% (percentage of errors), a variable that measures the qualitative aspect of the sub-

ject's performance; TN-E (total performance) which is a measure of the quantity of work completed after a single correction for errors and omissions.

The TP[9] is a timed paper-and-pencil cancellation test that assesses sustained and selective attention, processing speed, visuo-perceptive and inhibition abilities and also demands high concentration levels and fatigue resistance. Its administration can be done individually or in group and takes exactly 10 minutes. The test consists of a blank sheet of paper with twenty-five lines and forty small squares per line. The squares are distinguished from each other by the orientation of the rows on the outer surface: in each square the stroke is oriented in eight possible directions, similarly to the wind rose. The subject is required to mark the three models proposed in the header. For each line the evaluator must register the total number of hits (H), i.e., the number of items correctly selected by the subject, errors (E), i.e., the number of items wrongly selected (false positives), and omissions (O), i.e., the number of correctly items that the subject had not selected (false negatives). The TP yet presents two main outcomes: the first one is Work Efficiency (WE),[33] which sets up a measure of both the attentional and perceptual abilities of the subject with a maximum score of 375 points calculated by the following formula [WE=H-(E+O)]; the second outcome is the Dispersion Index (DI),[33] corresponding to the percentage of failures committed by the subject during the test, calculated by [DI=(E+O)/H*100]. This index allows a higher characterization of the result obtained on WE, namely if it is mainly influenced by a pattern of global response slowing or selection mistakes - impulsivity (errors) or distractibility (omissions).

Statistical analysis

Statistical analyses were developed using the Statistical Package for the Social Sciences (SPSS), version 22.0 for Windows.[34] Descriptive statistics were used for sample's characterization followed by independent-samples t-test for group comparisons. The influence of sociodemographic characteristics, as age and education level, in TP scores was addressed with multiple linear regression (MLR) analysis (Enter method). To assess test-retest reliability we calculated the correlations between the scores at baseline and at follow-up six months later (only for the CG). The convergent validity was determined using Pearson correlation coefficients between the TP (WE and DI) and the MMSE, MoCA, TMT-A/B, and D2. The group differences were examined using two-sample t-test and analysis of variance (ANOVA). Estimates of effect size were also calculated through analysis of eta squared (η^2).[35] The diagnostic accuracy of the TP (WE and DI) for the prediction of the clinical diagnosis of MCI and AD was evaluated by the receiver operating characteristics (ROC) curve analysis. The optimal cut-off points for each index of TP that generated the highest Youden value were selected, with higher values meaning the maximization of both sensitivity and specificity. For the analysis of the predictive values of these indexes we calculated, for each of the cut-off points, the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and classification accuracy. The comparative analysis between the AUC values was performed through the statistical software MedCalc for Windows, version 18.[36]

Results

Sample characterization

The characteristics of the total study sample as well as for each group are presented in Table 1. We present data on the sample size, age, education level, gender, MMSE score, MoCA score, TMT A/B scores, GDS scores and SMC scores. There were no statistically significant differences between the three groups on mean age (65.23 ± 11.19, ranging from 45 to 83 years) and mean educational level (7.85 ± 4.48, ranging from 3 to 18 years). According to *Post-hoc* test analysis, we obtained the same pattern for each comparison (mean age: control = MCI = AD; mean education level: control = MCI = AD). As expected, we found statistically significant differences between the groups on cognitive performance on the MMSE, MoCA, TMT-A, TMT-B, GDS-30 and SMC. *Post-hoc* test analysis confirmed that the control group had a better cognitive performance than both clinical groups in all measures except for TMT-A, between CG and MCI. Besides, the MCI group revealed higher performance levels than the AD group in all described instruments.

In order to characterize the influence of age, education and gender in the results of TP, we correlated these variables with the work efficiency and dispersion index total scores. There were no statiscally significant correlations between TP indexes and gender. We found that age was significantly negatively correlated with the total

Table 1. Sociodemographic and cognitive characterization of the study sample.

	CG	MCI	AD	
n	**100**	**100**	**50**	**Differences between groups**
Education	6.70±4.33	6.97±4.35	6.30±4.33	$F_{(1,249)}$ = .058; p=.944
Age	71.28±9.53	71.40±9.29	71.83±9.97	$F_{(1, 249)}$ = .271; p=.763
Gender	54 (54.0)	58 (58.0)	29 (58.0)	$F_{(1, 249)}$ = .201; p=.796
MMSE score	28.89±1.41	28.08±1.54	22.37±3.17	$F_{(2, 249)}$ = 92.936, p<.01
MoCA score	25.40±3.27	20.00±3.51	11.54±4.23	$F_{(2, 249)}$ = 92.086, p<.01
GDS score	2.67±2.88	11.13±6.61	6.69±4.77	$F_{(2,235)}$ = 49.005, p<.01
SMC score	5.00±5.93	8.17±4.27	6.46±4.17	$F_{(2,235)}$ = 21.880, p<.01
TMT A score	80.21±34.80	89.22±37.99	165.44±90.18	$F_{(2,228)}$ = 20.06, p<.01
TMT B score	210.50±118.91	233.28±98.92	364.82±204.02	$F_{(2, 201)}$ = 6.14, p<.01
D2 (TN-E) score	103.70±67.91	-	-	-
D2 (E%) score	30.87±41.54	-	-	-

Notes: Gender is presented by female's n and its respective percentage (%). The others variables are presented as means ± standard deviation. CG = control group; MCI = mild cognitive impairment; AD = Alzheimer's disease; MMSE = Mini Mental State Examination; MoCA = Montreal Cognitive Assessment; GDS = Geriatric Depression Scale; SMC = subjective memory complains; TMT A = Trail Making Test A; TMT B = Trail Making Test B; D2 = D2 Test of Attention.

scores on TP-WE (r=-0.508, $p < 0.01$) and significantly positively correlated with TP-DI (r=0.232, $p < 0.01$). Regarding educational level we observed that there was a significant positive correlation with TP-WE (r=0.554, $p < 0.01$) and a significant negative correlation was found with TP-DI (r=-0.306, $p < 0.01$). Furthermore, we performed a MLR analysis (enter method) to analyse the influence of these two variables in TP results (Table 2). The results presented on Table 2 showed that both age and education are significant contributors to the prediction of the TP. To this model, the adjusted R^2 value was 0.436, which means that 43.6% of the variance on the TP-WE was explained by these two sociodemographic variables. The same was observed for TP-DI, with an adjusted R^2 value of 0.114 (11.4%).

Table 2. MLR analysis for age and educational level.

Variable	B	SEB	β
TP-WE $_1$			
Age	-3.445	.350	-.508
Education	26.688	4.223	.554
TP-DI $_2$			
Age	1.142	.287	.232
Education	-18.704	3.506	-.306
$_1R^2$=.436, $F_{(2,249)}$=103.670, p<.01; $_2R^2$=.114, $F_{(2,249)}$=17.641, p<.01;			

WE = work efficiency; DI = dispersion index; SEB = standard error of B.

Psychometric properties

The convergent validity was determined through Pearson correlations between the TP and the MMSE, MoCA, TMT, and D2 (this one specifically in CG group). We observed significant negative correlations between the total scores of the TP-DI and the MMSE (r=-0.617; $p < 0.01$) and the MoCA (r=-0.620, $p < 0.01$) and conversely positive correlations between TP-WE and the MMSE (r= 0.868; $p < 0.01$) and the MoCA (r=0.891; $p < 0.01$), once that the more errors and omissions subjects commit, the worst will be the scores obtained on the cognitive screening measures. The results also showed significant negative correlations between the total scores of the TP-WE and the TMT-A (r=-0.565, $p < 0.01$) and B (r=-0.524, $p < 0.01$). Positive significant correlations were found between the total scores of TP-DI and the TMT-A (r=0.486, $p < 0.01$) and B (r=0.422, $p < 0.01$). These results show that the higher the work efficiency on TP, the less time subjects take to complete both TMT A and B. Regarding D2 test, we observed significant positive correlations between D2 (TN-E) and the TP-WE (r= 0.959, $p < 0.01$) and significant negative correlations between the D2 (TN-E) and TP-DI (r=-0.920, $p < 0.01$). These results were expected since D2 (TN-E) and TP-WE are both indicative of the quantity of subjects' work and TP-DI is a measure of inattention what explains its negative correlations with D2 (TN-E).

Test-retest reliability was performed for a convenience sub-sample of the control group (n=30) and both the assessments were performed by the same evaluator. The follow-up period was fixed in six months according with the methodologies used in other studies for MCI

and AD patients.[37] The reliability was measured through Pearson's correlation coefficient between the baseline and the follow-up data and the obtained values were r=0.877 (TP-DI) and r=0.903 (TP-WE), all $p < 0.01$. These results indicate a good test-retest reliability.[38]

Group differences

We conducted an ANOVA in order to analyse differences between groups on TP indexes as well as the respective effect sizes (η^2). There were statistically significant differences in TP main indexes concerning the three groups: $F_{(2, 247)}$= 54.696, $p < 0.01$, η^2= 0.282 (TP-WE); $F_{(2, 247)}$= 55.883, $p < 0.01$, η^2= 0.287 (TP-DI). In Table 3 we present in detail TP performances for all groups and for all the five indexes. According to *post-hoc* analysis, the control group obtained the higher total scores and the AD group obtained the lowest total scores.

Cut-off points

The accuracy of the test depends on how well the test distinguishes the population being tested into those with and without the disease. It is measured by the area under the receiver operating curve (ROC) curve. An area of 1 represents a perfect test and an area of 0.5 represents a worthless test. ROC curve analysis was calculated to evaluate the diagnostic accuracy of the TP to distinguish MCI and AD patients from healthy older adults. Graphic representations of the ROC curves are delivered in Figs. 1 and 2. The AUCs for MCI in the two main indexes of TP were for TP-WE 0.739 [95% confidence interval (CI) = 0.631-0.847] and for TP-DI 0.629 [95% CI = 0.506-0.753]. For AD, the AUCs were 0.981 [95% CI = 0.960-1.000] for TP-WE and 0.921 [95% CI = 0.861-0.980] for TP-DI. So, the obtained values showed that there was a fair diagnostic accuracy for the MCI group and a good-excellent diagnostic accuracy for AD patients.

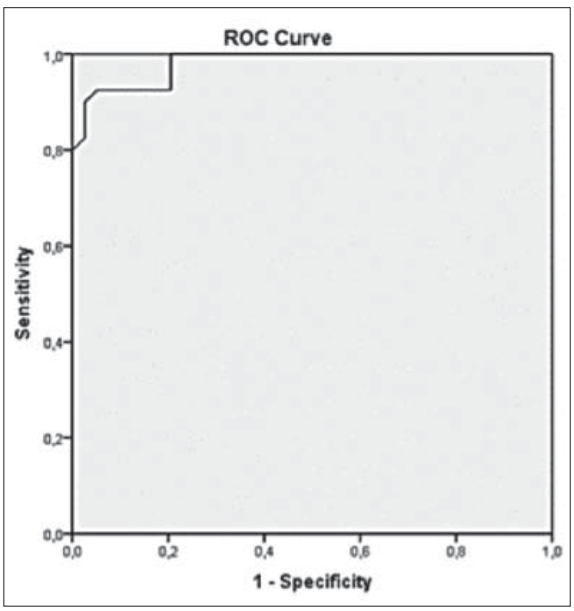

Figure 1. ROC curve analysis of the TP-WE to detect AD.

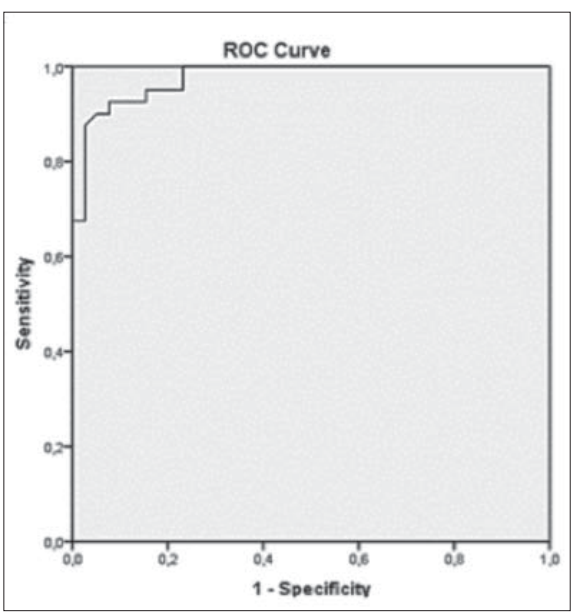

Figure 2. ROC curve analysis of the TP-DI to detect AD.

Table 3. Cognitive characterization of the groups on TP.

	CG	MCI	AD	
n	**100**	**100**	**50**	**Diferences between groups**
WE	124.48±55.50	79.50±43.14	-3.55±64.06	$F_{(2, 247)}$ = 54.696, p<.01, η^2= .282
DI	19.20±16.16	27.32±21.37	113.91±106.04	$F_{(2, 247)}$) = 55.883, p<.01, η^2= .287
H	163.64±74.95	102.13±39.94	53.05±23.60	$F_{(2, 247)}$ = 51.957, p<.01, η^2= .398
E	0	1.70±3.92	9.35±14.91	$F_{(2, 247)}$ = 14.971, p<.01, η^2= .160
O	19.94±16.58	21.88±14.75	48.08±66.00	$F_{(2, 247)}$ = 5.017, p=.008, η^2= .06

Notes: The scores are presented as means ± standard deviation. WE = work efficiency; DI = dispersion index; TR= total result; H = hits; E = errors; O = omissions.

Table 4. Diagnostic classification accuracy of the TP.

	TP	Cut-off	AUC	Sensitivity	Specificity	PPV	NPV	Classifcation Accuracy
MCI	WE	<73	0.739	53	85	78	64	69
	DI	>27	0.629	48	83	73	61	65
AD	WE	<49	0.981	93	95	95	93	94
	DI	>26	0.921	93	82	84	92	88

AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value; TP = Toulouse-Piéron; WE = work efficiency; DI = dispersion index; TR = total result; H = hits; E = errors; O = omission. Note: Values of sensitivity, specificity, PPV, NPV, and classification accuracy were expressed in percentage. Cut-off points indicate the minimum score required for presence of signal.

The optimal cut-off points for maximum accuracy and the respective values of sensitivity, specificity, PPV, NPV, and classification accuracy for both MCI and AD were further calculated (Table 3), with the ideal value regarding a perfect test being one (100%), and the worst possible value being zero.

Discussion

Several studies developed in our country with different neuropsychological instruments revealed the critical influence of sociodemographic variables on performance, influencing normative and validation data.[10,13,29,31] Like those studies our results confirm that the significant sociodemographic variables are age and education. Considering the TP indexes, MLR analysis showed that educational level was the most significant predictor of the TP results which is congruent with other studies for the Portuguese population.[29,31] Similarly, the model encompassing both age and education explained 43.6% of the variance of results in TP-WE and 11.4% in TP-DI.

In order to validate the TP as a measure of sustained and selective attention we explored its psychometric properties. This is essential to guarantee that the results are adequate to the focus of the evaluation and to ensure that the test provides adequate values of diagnostic accuracy allowing its optimal use in comprehensive neuropsychological assessment protocols. As so, we explored both convergent validity and test-retest reliability. Regarding convergent validity, we obtained statistically significant correlations between TP indexes and both MMSE and MoCA total scores, as well as with TMT-A/B and D2 scores, which are more specific instruments for the assessment of attentional abilities. These results show a convergent validity for both modalities of evaluation in dementia.[35] We also explored TP's test-retest reliability. We revaluated 30 subjects of the control group six months after the baseline assessment and concluded that their performances remained stable across the time. This was shown by the high Pearson correlation values ($r > 0.80$) obtained when the two evaluations were compared.[38] This property is of extreme importance once it shows that TP performances remain stable over assessments in controls, being therefore appropriate for use in longitudinal research to identify pathological decline in dementia.

We further explored differences between groups, as well as the optimal cut-off points for maximum accuracy.

The analysis of group differences suggested that TP was able to distinguish between clinical and control groups, as well as between the clinical conditions targeted in this study. Furthermore, we observed statistically significant differences between the performances of the three groups in each one of the five indexes of TP, which reinforces its discriminative power. The results showed the projected tendency to obtain worse performances in AD group in comparison with both MCI and controls. These findings were congruent with previous studies concerning the assessment of attention in early AD and showed that after an early amnesic stage, attention is one of the first non-memory domains to become impaired, even before the emergence of deficits in language and visuospatial skills.[2,14] More specifically, we found that all three groups committed omission mistakes, possibly due to a pattern of distractibility but that the presence of errors was only verified in the AD group and in a-MCI-multidomain patients who have the lowest scores in cognitive screening measures. This pattern of impulsivity is congruent with what was previously described in these patients which typically reveal attentional and oculomotor abnormalities that can have a negative impact on visual processing and associated cognitive functions.[39] Through an overall performance analysis concerning the three groups, we observed that AD patients seem to have severe difficulties to under-

stand task instructions and are not able to memorize the models proposed in the header. These features lead to a high slow response pattern and also a very low speed of attentional switching, culminating in very low work efficiency scores. This was also demonstrated in previous studies,[40,41] showing that early AD patients have significant deficits in sustained attention and visuo-perceptive skills which may affect performances on global cognitive testing and therefore should be taken into account during clinical assessments. This impulsivity pattern was verified not only when patients committed errors but also while they were completing the test itself. They seem to be very confused regarding what they have to do and frequently start to mark all the squares repeatedly. This can also explain the weak performances in WE index as well as higher levels in the dispersion index. The overall performance of AD patients in TP is also congruent with an impairment of inhibitory control and processing speed abilities, that was also observed in TMT A/B scores and illustrated by the high Pearson correlation values obtained when correlating both the instruments, providing a mark of working memory and attentional dysfunction in AD.[42]

As far as we know, there is no international study encompassing both AD patients and the TP, making cultural or geographical comparisons unfeasible. Regarding other attentional tests, it was recently found that the direct assessment of attention through different tasks can help to identify AD patients at a prodromal stage, showing that, there is a gradual impairment in tests of working memory and attentional control in both multidomain MCI and AD, as above-mentioned.[8] It was also found that MCI patients may present deficits related with response inhibition, switching and cognitive flexibility skills, confirming the presence, in some cases, of deficits in attention and executive functions besides memory,[43] which is congruent with the results we found in some of the a-MCI-multidomain patients.

One of the main focus of our study was to explore the diagnostic accuracy of the TP in the spectrum of AD. As expected, the analysis of the diagnostic validity of the TP suggests a higher discriminative potential of the TP indexes for AD than for MCI patients. Effectively, for the optimal cut-off points established, the respective AUC and diagnostic parameters were higher for AD patients. Regarding the capacity of the test to distinguish between controls and MCI's, the two main indexes showed a fair sensitivity and, consequently, a not-so-good classification accuracy (percentage of correct predictions). On the other hand, for AD patients we observed high sensitivity and specificity values, leading to an excellent diagnostic accuracy. So, for the MCI group, the observed sensitivity and classification accuracy should be viewed as an indicator of high likelihood to have false-negative cases and the obtained cut-off points should be used carefully. At the same time, better effect sizes were observed for AD (TP-WE: $\eta^2 = 0.54$; TP-DI: $\eta^2 = 0.29$), when compared to MCI patients (TP-WE: $\eta^2 = 0.17$; TP-DI: $\eta^2 = 0.05$), corroborating the more suitable use of this neuropsychological instrument for evaluating attentional processes during disease progression. As so, results offer strong evidence that TP is an excellent and reliable psychometric tool to distinguish between cognitively healthy adults and AD patients.

The main limitation of our study was the exclusion of illiterate subjects, which represent 5% of all Portuguese population and correspond to 25.7% of the population over 65 years old.[44] However, it has been extensively described that the educational level has a strong influence on cognitive performance, especially in complex tests like the TP, leading to floor effects.[45] Furthermore, we only included the amnestic subtype of MCI (single and multi-domain) patients so the generalization of the results for other types of MCI should be done carefully.

Besides these boundaries, this study has an important set of strengths: (1) we believe that its added value is the rigorous and meticulous methodology used; (2) it included homogeneous clinical groups (including an MCI group that allows to afford the knowledge about the discriminant ability of the TP within the spectrum of AD); (3) equivalent sample-sizes reducing the probability of occurrence of biases in statistical analysis; and (4) the overall matching between the three groups regarding age and education (the most significant sociodemographic variables influencing test results).

Looking forward, we intend to continue focused on TP's psychometric properties, increasing the number of subjects in test-retest reliability and further exploring interrater reliability. Additional directions for future research include studies regarding sub-classification according to the MCI-cognitive profile, namely distinguishing between amnesic single-domain and multidomain MCI, since the last one seems to presents a higher risk of future conversion to AD.[46,47] Similarly, it is important

to investigate other forms of dementia namely Dementia with Lewy bodies, Parkinson's disease dementia, or vascular dementia where attentional deficits are a key feature explicitly considered in clinical criteria.

Conclusion

In conclusion, the novelty of our results for the Portuguese population highlights the excellent diagnostic accuracy of the TP indexes for dementia, specifically its excellent discriminative power for AD. However, it is relevant to mention the need for a careful use of the TP with MCI patients, once it revealed a fair sensitivity and classification accuracy corroborating the fact that the TP should not be used as a single neuropsychological assessment instrument for the screening of cognitive decline. ■

Responsabilidades Éticas

Conflitos de Interesse: Os autores declaram a inexistência de conflitos de interesse na realização do presente trabalho.

Fontes de Financiamento: Não existiram fontes externas de financiamento para a realização deste artigo.

Confidencialidade dos Dados: Os autores declaram ter seguido os protocolos da sua instituição acerca da publicação dos dados de doentes.

Proveniência e Revisão por Pares: Não comissionado; revisão externa por pares.

Ethical Disclosures

Conflicts of Interest: The authors have no conflicts of interest to declare.

Financing Support: This work has not received any contribution, grant or scholarship

Confidentiality of Data: The authors declare that they have followed the protocols of their work center on the publication of data from patients.

Protection of Human and Animal Subjects: The authors declare that the procedures followed were in accordance with the regulations of the relevant clinical research ethics committee and with those of the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Provenance and Peer Review: Not commissioned; externally peer reviewed.

References

1. Santana I, Farinha F, Freitas S, Rodrigues V, Carvalho A. Epidemiologia da demência e da Doença de Alzheimer em Portugal: Estimativas da prevalência e dos encargos financeiros com a medicação. Act Med Port. 2015, 28: 182-8. doi: 10.20344/amp.6025
2. Prince M, Ali GC, Guerchet M, Prina AM, Albanese E, Wu YT. Recent global trends in the prevalence and incidence of dementia, and survival with dementia. Alzheimers Res Ther. 2016, 8.1: 1-13. doi: 10.1186/s13195-016-0188-8
3. Wortmann M. Dementia: a global health priority-highlights from an ADI and World Health Organization report. Alzheimers Res Ther. 2012, 4.5: 40.
4. Lobo A, Launer LJ, Fratiglioni L, Andersen K, Di Carlo A, Breteler MB, et al. Prevalence of dementia and major subtypes in Europe: a collaborative study of population-based cohorts. Neurology. 2000, 54.5: S4.
5. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: A systematic review and metanalysis. Alzheimers Dement. 2013; 9:63-75.e2. doi: 10.1016/j.jalz.2012.11.007.
6. Kirova AM, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. Biomed Res Int. 2015;2015:748212. doi: 10.1155/2015/748212
7. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: a clinical review. JAMA Neurol. 2014; 312: 2551-2561. doi: 10.1001/jama.2014.13806
8. Sharma S, Kaur S, Tripathi M, Talwar A, Sharma R. differential deficits in attention, working and semantic memory discriminates between mild cognitive impairment and Alzheimer's disease. Indian J Physiol Pharmacol. 2017; 61: 348-356.
9. Amaral JR. O teste da barragem de Toulouse e Piéron: Elementos de aferição para a população portuguesa. Lisboa: Fundação Calouste Gulbenkian; 1967.
10. Almeida LS, Simões MR, Almiro P, Santos MJ. Práticas de avaliação psicológica em Portugal: Resultados gerais de inquérito realizado sobre a utilização de testes e outros instrumentos. Paper presented at: 10º congresso AIDAP/AIDEP: Diagnóstico e avaliação psicológica; September, 2018; Coimbra, Portugal.
11. Rey A. Manuel: Test de Copie d`une Figure Complexe. Paris: Centre de Psychologie Appliqueé; 1959.
12. Espírito-Santo H, Lemos L, Ventura L, Moitinho S, Pinto AL, Rodrigues F et al. Teste da Figura Complexa de Rey-Osterrieth. In: Mário R. Simões, Isabel Santana, & Grupo de Estudos de Envelhecimento Cerebral e Demência (GEECD) (Coords.), Escalas e Testes na Demência. 3ª ed. Lisboa: Novartis; 2015.
13. Brickenkamp R. Teste de Atenção D2: Manual técnico (adaptação portuguesa: Carla Ferreira e António M. Rocha). Lisboa: Cegoc; 2007.
14. Kaiser A, Kuhlmann BG, Bosnjak M. A meta-analysis of inhibitory-control deficits in patients diagnosed with Alzheimer's dementia. Neuropsychology. 2018, 18: 615-33. doi: doi.org/10.1037/neu0000460
15. Folstein MF, Folstein SE, McHugh PR. Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res. 1975, 12: 189-98.
16. Guerreiro M. Contributo da neuropsicologia para o estudo das demências.Lisboa: Faculdade de Medicina de Lisboa.
17. Nasreddine Z, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I,et al. The Montreal Cognitive Assessment, MoCA: A Brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005; 53: 695–9. doi: 10.1111/j.1532-5415.2005.53221.x.
18. Simões MR, Freitas S, Santana I, Firmino H, Martins C, Nasreddine Z,et al. Montreal Cognitive Assessment (MoCA). Coimbra: Serviço de Avaliação Psicológica, Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra; 2008.
19. Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: Clinical characterization and outcome. Arch Neurol. 1999; 56: 303–8. doi: 10.1001/archneur.56.3.303.
20. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC,et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging and Alzheimer's Association workgroup. Alzheimers Dement. 2011; 7: 270-9. doi: 10.1016/j.jalz.2011.03.008.
21. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack Jr CR, Kawas C, et al. 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. Alzheimers Dement. 2011; 7: 263-9. doi: 10.1016/j.jalz.2011.03.005.
22. Cavaco S, Gonçalves A, Pinto C, Almeida E, Gomes F, Moreira I,et al. Trail Making Test: Regression-based Norms for the Portuguese Population. Arch Clin Neuropsychol. 2013; 28: 189-98. doi:10.1093/arclin/acs115.

23. Yesavage J A, Brink TL, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. J Psychiatr Res. 1983; 17: 37-49. doi: 10.1016/0022-3956(82)90033-4.

24. Simões, MR, Prieto G, Pinho, MS, Sobral M, Firmino H, Grupo de Estudos de Envelhecimento Cerebral e Demência. Geriatric Depression Scale (GDS-30). Coimbra: Laboratório de Avaliação Psicológica e Psicometria, Faculdade de Psicologia e de Ciências da Educação da Universidade de Coimbra; 2013.

25. Simões MR, Sousa LB, Pinho MS, Vilar M, Prieto G, Firmino H. Escala de Depressão Geriátrica (GDS). In: Gonçalves MM, Simões MR, Almeida LS, coordenadores. Psicologia Clínica e da Saúde: Instrumentos de Avaliação. Lisboa: Pactor: 2017; p. 219-33.

26. Schmand B, Jonker C, Hooijer C, Lindeboom J. Subjective memory complaints may announce dementia. Neurology. 1996; 46: 121-5.

27. Ginó S, Guerreiro M, Garcia C. Escalas de Queixas de Memória (SMC). In: Mendonça A, Guerreiro M, Grupo de Estudos de Envelhecimento Cerebral e Demência, editores. Escalas e Testes na Demência. 2ª ed. Lisboa: Novartis; 2008. p. 117-20.

28. Ginó S, Mendes T, Mendonça A, Guerreiro M. Escala de Queixas Subjectivas de Memória (QSM). In: Simões MR, Santana I, Grupo de Estudo de Envelhecimento Cerebral e Demência, editores. Escalas e Testes na Demência. 3ª ed. Porto Salvo: Novartis; 2015. p. 44-9.

29. Freitas S, Simões MR, Alves L, Santana I. Mini Mental State Examination (MMSE): Normative study for the Portuguese population in a community stratified sample. Appl Neuropsychol: Adults. 2014; 22: 311-9. doi: 10.1080/23279095.2014.926455.

30. Freitas S, Simões MR, Alves L, Santana I. Mini Mental State Examination (MMSE). In: Simões MR, Santana I, Grupo de Estudo de Envelhecimento Cerebral e Demência, editores. Escalas e Testes na Demência. 3ª ed. Porto Salvo: Novartis; 2015. p. 18-23.

31. Freitas S, Simões MR, Alves L, Santana I. Montreal Cognitive Assessment (MoCA): Normative study for the Portuguese population. J Cli Exp Neuropsychol. 2011; 33: 989-96. doi: 10.1080/13803395.2011.589374.

32. Freitas S, Simões MR, Alves L, Santana I. Montreal Cognitive Assessment (MoCA). In: Simões MR, Santana I, Grupo de Estudos de Envelhecimento Cerebral, editores. Escalas e Testes na Demência. 3ª ed. Porto Salvo: Novartis; 2015. p. 24-31.

33. Baeta É. Bateria para avaliação neuropsicológica de adultos com epilepsia. Psicologia. 2002; 16: 79-96.

34. IBM Corp. IBM SPSS Statistics for Windows (Version 22.0). Armonk: IBM; 2013.

35. Cohen RJ. Statistical power analysis for the behavioral sciences. 2ª ed. Hillsdale: Erlbaum; 1988.

36. MedCalc Software. MedCalc—user-friendly statistical software. Belgium: Ostend; 2018.

37. Freitas S, Simões MR, Alves L, Santana I. Montreal cognitive assessment: validation study for mild cognitive impairment and Alzheimer disease. Alzheimer Dis Assoc Disord. 2013; 27: 37-43. doi: 10.1097/WAD.0b013e3182420bfe.

38. Vaz S, Falkmer T, Passmore, AE, Parsons R, Andreou P. The case for using the repeatability coefficient when calculating test-retest reliability. PLoS One. 2013; 8: e73990. doi: 10.1371/journal.pone.0073990.

39. Fernández G, Orozco D, Agamennoni O, Schumacher M, Sañudo S, Biondi J, et al. Visual Processing during Short--Term Memory Binding in Mild Alzheimer's Disease. J Alzheimers Dis. 2018;63:185-94. doi: 10.3233/JAD-170728.

40. Perry RJ, Watson P, Hodges JR. The nature and staging of attention dysfunction in early (minimal and mild) Alzheimer's disease: relationship to episodic and semantic memory impairment. Neuropsychologia. 2000; 38: 252-71. doi: 10.1016/S0028-3932(99)00079-2.

41. Huntley JD, Hampshire A, Bor D, Owen AM, Howard RJ. The importance of sustained attention in early Alzheimer's disease. Int J Geriatr Psychiatry. 2017; 32: 860-7. doi: 10.1002/gps.4537.

42. Crawford TJ, Higham S, Mayes J, Dale M, Shaunak S, Lekwuwa G. The role of working memory and attentional disengagement on inhibitory control: effects of aging and Alzheimer's disease. Age. 2013; 35: 1637-50.

43. Petersen RC. Mild cognitive impairment. CONTINUUM: Lifelong Learning in Neurology. 2016; 22: 404-8. doi: 10.1212/CON.0000000000000313.

44. INE & PORDATA. População residente com 15 a 64 e 65 e mais anos: Por nível de escolaridade completo mais elevado (%). [Accessed September 24, 2018] Available at: https://www.pordata.pt/Portugal. [Accessed September 24, 2018]

45. Brucki SM. Illiteracy and dementia. Dement Neuropsychol. 2010; 3: 153-7.

46. Saunders NL, Summers MJ. Longitudinal deficits to attention, executive, and working memory in subtypes of mild cognitive impairment. Neuropsychology. 2011; 25.2: 237. doi: 10.1037/a0021134.

47. Roberts RO, Geda YE, Knopman DS, Cha RH, Pankratz VS, Boeve BF, et al.The incidence of MCI differs by subtype and is higher in men: The Mayo Clinic Study of Aging. Neurology. 2012; 78: 342-51.