Turkish Journal of Geriatrics
2013; 16 (1) 69-76

Banu CANGÖZ[1]
Serpil DEMİRCİ[2]
Sait ULUÇ[1]

RESEARCH

TRAIL MAKING TEST: PREDICTIVE VALIDITY STUDY ON TURKISH PATIENTS WITH ALZHEIMER DEMENTIA

ABSTRACT

Introduction: The Trail Making Test is a widely used test that assesses executive functions such as visual tracking, psychomotor speed, complex attention and mental flexibility. The main aim of this study is to evaluate the validity of the TMT in AD patients.

Materials and Method: The study group covers 50 subjects with AD and 50 healthy subjects. The mean age was 73.98 years in AD group and 74.74 years in control group. The mean scores that were obtained from 8 subscores of the TMT were compared between both groups.

Results: The TMT Time A, Time B, Time A+B, Time B-A, Correction B, Error A and Error B scores were significantly different between two groups. The estimated cut-off values for Time A, Time B, Time A+B, and Time B-A are 108 sec, 240 sec, 356 sec, and 112 sec, respectively.

Conclusion: It is known that both parts of TMT are sensitive to progressive cognitive decline. It is suggested that Part A and B may differentiate demented from normal persons and Part B may detect early stages of the AD. Our results were in accordance with related literature, and revealed that the TMT is a valid a psychometric instrument in assessing Turkish AD patients.

Key Words: Alzheimer Disease; Dementia; Neuropsychological Tests.

ARAŞTIRMA

İZ SÜRME TESTİ: TÜRK ALZHEIMER TİPİ DEMANS HASTALARI İÇİN YORDAYICI GEÇERLİK ÇALIŞMASI

ÖZ

Giriş: İz Sürme Testi (İST) görsel iz sürme, psikomotor hız, karmaşık dikkat ve zihinsel esneklik gibi yönetici işlevleri değerlendirmede yaygın olarak kullanılan bir testtir. Bu çalışmanın temel amacı, İST'nin Alzheimer hastaları (AH) için geçerliğini değerlendirmektir.

Gereç ve Yöntem: Çalışmaya 50 AH ile 50 sağlıklı yaşlı birey denek olarak katılmıştır. AH grubunun yaş ortalaması 73.98, sağlıklı yaşlı grubunun yaş ortalaması 74.74'dür. İki grup İST'den alınan 8 alt puan açısından karşılaştırılmıştır.

Bulgular: İki grup, İST A Süre, B Süre, A+B Süre, B-A Süre, B Düzeltme, A Hata ve B Hata puanları açısından anlamlı düzeyde farklıdır. A Süre, B Süre, A+B Süre ve B-A Süre alt testleri için kesme değerleri sırasıyla 108 sn, 240 sn, 356 sn ve 112 sn olarak belirlenmiştir.

Sonuç: İST'nin her iki bölümünün de ilerleyici bilişsel gerilemeye duyarlı olduğu bilinmektedir. Bölüm A ve Bölüm B sağlıklı yaşlıları demanslı bireylerden ayırdedebilirken; Bölüm B, AH'yi erken evrede belirleyebilmektedir. Elde edilen sonuçlar ilgili litertürü desteklemiş ve İST'nin Türk Alzheimer hastalarının değerlendirilmesinde kullanılabilecek, psikometrik açıdan geçerli bir araç olduğunu göstermiştir.

Anahtar Sözcükler: Alzheimer Hastalığı; Demans; Nörospikolojik Testler.

İletişim *(Correspondance)*

Banu CANGÖZ
Hacettepe Üniversitesi Edebiyat Fakültesi Psikoloji
ANKARA

Tlf: 0312 297 64 45
e-posta: banucan@hacettepe.edu.tr

Geliş Tarihi: 18/07/2011
(Received)

Kabul Tarihi: 03/11/2011
(Accepted)

[1] Hacettepe Üniversitesi Edebiyat Fakültesi Psikoloji ANKARA
[2] Süleyman Demirel Üniversitesi Tıp Fakültesi Nöroloji Anabilim Dalı ISPARTA

The Trail Making Test (TMT) (1), have been widely used as an easily and quickly administered neuropsychological test, either alone as a screening instrument or as part of a larger battery of tests. Originally, it used as part of the Army Individual Test Battery. Part A of the test is generally believed to be a test of visual search providing information about attention, visual scanning and speed of eye-hand coordination, while Part B is considered to assess the cognitive flexibility with more precision (2). All these processes are of great significance for an efficient accomplishment of executive functions (1,2).

Administration and scoring procedures of TMT have changed over the years. In the original test, each part was terminated after three uncorrected errors and each part received a score on a ten-point scale depending on the time taken to complete it. On later administrations of TMT, patients were allowed to complete the test regardless of the number of errors and were given a score of zero to performances in which errors were left uncorrected. In 1955, Reitan proposed a different application by requiring the examiner to score on time alone. As noted by Lezak, in Neuropsychological Assessment-third edition, this method is also criticized that the price for a simplified scoring system might have been paid in diminished reliability, for the measured amount of time includes the examiner's reaction time (in noticing errors) and speed in pointing them out and the speed with which the patient comprehends and makes the correction. This method penalizes for errors directly but does not control for differences in response times and correction styles that can conceivably result in significant biases in the time scores obtained with different examiners.

Despite the widespread use of TMT, it follows that there exist few comprehensive and/or large-scale normative studies for ages 50 and over. Most published normative data for TMT contained subjects with a limited age range, especially age 50 and over (55-74, 75-98 years etc.) (3). Similarly, normative and/or validation data for elderly people were collected from time to completion, difference or a ratio scores (Time A, Time B, B-A, B/A) (3,4). Frequency of errors, while often recorded and reported clinically, has not been empirically evaluated in prior TMT normative studies. Some researchers have pointed out that examiner's correction of errors adds additional time to the total score, thus accounting for difficulties reflected in the number of errors (5).

The challenging nature of the test Part B is expressed with slow response time . Age and education are two important factors that have impact on the time to complete the test (1,4).

In order to remove the effect of velocity component, the difference between the times required to complete the Part B and (Time B-A) was used. This score was found to be highly correlated with mental ability tests and various cognitive disorders (5).

The absolute cut-off values of a test must be estimated to identify organic brain damage. Davis (6) compared young, middle-aged and brain-damaged groups of subjects, and suggested that '*different cutting points were required for a young and for a middle-aged group. Without establishing these different cutting points, normals are more likely to be misclassified as brain-damaged as age increases' (p.98)*.

However, the TMT error rate is difficult to interpret in isolation, particularly because errors are common among cognitively normal adults. Ruffolo, Guilmette and Willis (7) observed that 34.7% of control participants committed at least one error on TMT Part B. In brain-lesion persons (*i.e.*, frontal vs. non-frontal) for a cut-off >1 error, a higher positive predictive factor was shown for a frontal lesion (*i.e.*, high error rate suggests presence of frontal lesion vs. other lesion) but poor negative predictive power (*i.e.*, less than 2 errors does not necessarily implicate a nonfrontal lesion) (5). Several studies noted that head-injured individuals are more likely to commit errors and less likely to correct them without prompting, whereas some failed to spot head-injured participants from controls based on errors (5,7).

Dementia, particularly AD was first emphasized as a major public health problem 30 years ago. Neuropsychological studies of individuals defined as neither normal nor demented demonstrate progressive declines in cognition over time. These are particularly striking in the area of episodic memory, but other domains appear to be affected as well (8,9). The finding is consistent with the fact that the clinical criteria (*i.e.*, NINCDS-ADRDA) for dementia require impairment in two or more cognitive domains (10). In this context, the differentiation of early stage AD cognitive changes from those observed in normal aging is important. Therefore, neuropsychological tests are important tools though not enough alone.

In AD, which is characterized by a general decline in cognitive processes, impairment in attentional tasks may be observed in early stages of the disease. However, some investigators propose that the problems on attentional tasks might be related with the memory impairments. Attention is the first non-memory aspect of cognition that declines in AD

prior to any deficits in language or visuo-spatial abilities. Attentional deficits might be the underlying problem in some of the difficulties in activities of daily living (11). For that reason, in the present study, besides validity testing the relationship between TMT sub scores and Functional Activities Questionnaire (FAQ) that measure dependency in daily life activities was evaluated.

The available literature indicates that various subcomponents of attention may be differentially affected in AD. For example, there is some evidence that focused attention appear to be relatively preserved in early stage AD (12). In this sense, TMT is regarded as an important tool in evaluation of AD suspected patients since it covers various subcomponents of attention.

It is known that both parts of TMT are sensitive to progressive cognitive decline. It is suggested that Part A may differentiate demented from normal persons and Part B may detect early stages of the AD (13). It is shown that TMT Part B and Stroop Test (interference card) were impaired in more than 40% in mild AD patients (14). In this study, complex attention skills were found to be affected more frequently than other executive functions; however, there was a considerable heterogeneity among AD patients in the pattern of executive dysfunction.

Patients with various dementias at mild, moderate and severe stages, need more time to complete the given tests as the disease severity increases; and studies showed a statistically significant difference among groups. The poor TMT error performance of patients with AD, supports the hypothesis of an inhibitory dysfunction. Similarly, it has been shown that TMT Part B time and error scores are differentiating measures in clinical assessment of AD patients (15,16).

However, it can be seen in the literature that validity studies comparing healthy persons and AD patients have usually analyzed the time to complete the test and error scores, and neglected the correction scores (16). This may arise from the practical difficulties in estimation of correction scores. In the present study, we compared groups for correction scores as well time and error scores.

Cangöz, Karakoç and Selekler (17) have evaluated the standardization and reliability of TMT in normal Turkish persons aged 50 and over in a previous study. The present study has mainly three aims: (a) to determine the predictive validity of TMT in Turkish AD patients. We also evaluated different kinds of scores (Time A, Time B, Time A+B, Time B-A, Error A, Error B, Correction A, Correction B) obtained from TMT which puts our study apart from the previous ones, (b) to determine the diagnostic accuracy of TMT error and correction scores in a sample consisting of controls and patients with AD, (c) to show the relation between TMT scores and daily living activities performance measured by FAQ.

METHOD

Participants

One hundred subjects participated in the study, including 50 persons with mild-moderate AD and 50 healthy controls that had at least 5 years education. Participants in the AD group were selected among consecutive patients admitted to Neurology Department of the Medical School and given the diagnosis of AD according to NINCDS-ADRA criteria (10). The mean age of this group was 73.98±7.35 (range 62-89 years) and there were 31 women and 19 men. Healthy controls were selected randomly among those who are working in state offices or retired ones. These participants were selected to individually match each AD participant on gender, education and hand preference. The mean age of this group was 74.74±7.24 (range 60-86 years) and there were 31 women and 19 men. Groups were similar in context of mean age (p = 0.604).

The inclusion and exclusion criteria for AD are consistent with the National Institute of Neurological and Communicative Diseases and Stroke-Alzheimers Disease and Related Disorders Association (NINCDS-ADRDA) criteria (10). The severity of dementia was staged according to the Clinical Dementia Rating (CDR) Scale (18). According to this scale, scores of 0, 0.5, 1, 2, and 3 represent to no dementia, very mild, mild, moderate, and severe dementia respectively. The other tools used to evaluate cognition, behavior and functional abilities were Mini Mental State Examination (MMSE)(19), Geriatric Depression Scale (GDS) (20), Three Words Three Shapes Test (3W-3S) (21), Functional Activities Questionnaire (17), and Clock Drawing Test (22). All tests and/or scales used in this study have already been subjected to adaptation, norm determination and standardization studies for Turkish cultural background. Those with a GDS score>14 and FAQ score>9 were excluded as well as those with a diagnosis of cerebrovascular disease, Parkinson's disease, multiple sclerosis, psychiatric disorder, hypothyroidisms, chronic pulmonary and renal disease or those taking any psychotropic drugs. Demographic features and screening test scores of participants were given in Table 1.

Table 1— The Mean and Standard Deviations of Scores Obtained From Screening Tests and Scales

				Control (n=50)		AD (n=50)	
	T	df	p	M	SD	M	SD
Age	-0.52	98	.60	73.98	7.35	74.74	7.24
MMSE	14.84*	98	.000	27.90	1.37	20.92	3.03
GDS	-4.10*	98	.000	5.22	3.09	8.36	4.44
FAQ	-9.15*	98	.000	2.54	2.63	7.90	3.20
CDT	6.91*	98	.000	3.62	0.53	2.32	1.22
3W-3S Scores Copying	6.32*	98	.000	29.80	0.45	28.26	1.66
Incidental	12.57*	98	.000	26.08	1.97	13.18	6.99
05 min delay	18.32*	98	.000	27.52	0.99	8.10	7.43
15 min delay	20.04*	98	.000	27.22	1.13	6.94	7.07
30 min delay	21.87*	98	.000	26.10	1.57	6.28	6.21
Multiple choice	35.71*	98	.000	27.48	0.76	14.82	2.39

*p< .001.

MATERIALS

Trail Making Test (TMT)

Participants were administered the two parts of the TMT. Part A consisted of encircled numbers from 1 to 25, which were randomly distributed on a A4 page (21 x 29 cm), while part B consisted of encircled numbers from 1 to 13 and encircled letters from A to İ, which were randomly distributed on a similar A4 page. Apart from the use of the first 12 letters of the Turkish alphabet, the Turkish version of the TMT followed the original format of the English version. Recently, Bowie and Harvey (2) have reviewed the administration and interpretation of the TMT. In a previous study, authors studied the standardization of TMT in a Turkish sample which comprised subjects older than 50 years age. According to the data obtained from that study, validity of the TMT components was as follows: Part A time score (Time A) 0.78, Part B time score (Time B) 0.73. The inter-rater validity was 0.99 for Time A and 0.93 for Time B (23).

Various scoring methods were proposed while assessing the TMT. The features we used in this study were as follows: 1) Turkish characters that do not exist in the original form such as 'Ç','Ğ' and 'İ' were added on the work sheet; 2) time to complete the test was not limited; 3) in addition to time scores number of errors and corrections were also scored. By the way, we measured 8 sub scores driven by the TMT performance.

Mini Mental State Examination (MMSE)

This test was developed by Folstein et al. in 1975. MMSE is the most widely used test for scanning general cognitive functions. It contains 30 items that cover orientation, memory, attention and language. Each correct answer is graded, so the total score ranges from 0 to 30. Güngen et al. (19) studied it's reliability and validity for Turkish society, and found high sensitivity and specificity (respectively, 0.91 and 0.95) with high positive and negative predictive values (0.90 and 0.95, respectively). The inter-rater validity showed a high correlation (r = 0.99) with a *Kappa* value of 0.92.

Geriatric Depression Scale (GDS)

This depression scale was developed by Yesavage et al. in 1983, contains 30 questions. Total score ranges between 0 and 30. The score is obtained based on a caregiver interview. Ertan and Eker (20) have evaluated the validity and reliability of this test for Turkish society. Internal consistency was 0.91 and test-retest validity was 0.74.

Functional Activities Questionnaire (FAQ)

This questionnaire was developed by Pfeffer et al. in 1982. It contains 10 items addressing complex functional daily living activities. FAQ is administered to one of the family members or caregivers. Each item is scored between 0 and 3, so the total score varies between 0 and 30. A score of nine or more in at least 2 items points to a dependency on functional activities. Its adaptation and standardization for Turkish sample was studied by Selekler, Cangöz and Karakoç (17).

Three Words-Three Shapes (3W-3S) Test

3W-3S was designed as an easy test for elderly patients that assess verbal and nonverbal memory within the same modali-

ty. The test comprised copying, incidental recall, learning trails to reach criterion, acquisition, delayed recall (5 min, 15, min, 30 min.), multiple choice recognition for words and shapes. Standardization of this test for Turkish sample was studied by Kudiaki and Aslan (21).

Clock Drawing Test (CDT)

CDT is a widely used and easily administered test which is used for screening for cognitive impairment and dementia as a measure of spatial dysfunction and neglect. Doing the test requires verbal understanding, memory and spatially coded knowledge in addition to constructive skills. This test developed by Freedman et al. in 1994. CDT can be performed in different ways and the scoring also varies. The one (4 point version) used in this study is described elsewhere, and test-retest correlation coefficient was found to be 0.88 and the inter-rater validity coefficient was 0.74 (24). Its adaptation and standardization for Turkish sample was studied by Cangöz, Karakoç and Selekler (22).

PROCEDURE

The examination took place at the Neurology Clinic. We first informed all participants about the aims of the study and the procedure, and asked them to sign a voluntary consent which was approved by the local hospital ethics committee. Before the examination, we obtained individual demographic data and a brief medical history for each participant. Each participant was assessed according to the TMT administration procedure and was also given the MMSE, CDR, GDS, FAQ, CDT and 3W-3S. Performance on the TMT was scored in terms of time required to successfully complete each part (A and B), (A+B), (B-A), the number of errors and corrections. Time A and Time B scores were determined by the tester with the help of a *CAGITA C6-503* digital hand chronometer (in seconds). Time B-A and Time A+B scores were calculated using Time A and Time B (in seconds). Error A, Error B, Correction A and Correction B scores were calculated by the tester upon counting the errors and corrections during the test procedure (numerical values). Error was described as false reactions given and corrected by the tester "*with interference by the tester*" within the time to complete Part A and B. Correction was described as false reactions given and corrected spontaneously by the participant "*without any interference by the tester*" within the time to complete Part A and B.

Statistical Analysis

For analyses, we used descriptive statistics, *t-test* and *Mann Whitney-U test*. As well, we accomplished *receiver operating characteristic* (*ROC*) analyses for each of the TMT trails as well as the other indexes produced from the TMT scores. We used the *ROC* analyses to determine the sensitivity and specificity of the TMT scores in distinguishing between the AD and healthy control groups. We analyzed the relationship of FAQ scores with TMT time scores with *Pearson* correlation test; and the relationship of FAQ scores with the TMT error and correction scores with *Spearman* correlation test. For *ROC* analyses we used *Med Calc* Statistical Analysis Program and for the others *SPSS 15.0* .

RESULTS

The tests that were used to discriminate between the demented and healthy subjects showed a statistically significant difference between groups. The healthy control group showed a higher performance in comparison to the AD group on MMSE (t = 14.84, p< 0.001), CDT (t = 6.91, p < 0.001) and on all 3W-3S subcomponents that were copying (t = 6.32, p < .000), incidental memory (t = 12.57, p < .000), 5 minute delay recall (5 min delay) (t = 18.32, p < .000), 15 minute delay recall (15 min delay) (t = 20.04, p < .000), 30 minute delay recall (30 min. delay) (t = 21.87, p < .000) and multiple choice (t = 35.71, p < .000). Whereas, AD patients were recorded to have greater GDS (t = -4.10, p < .000) and FAQ scores (t = -9.15, p < .000).

Comparison of scores obtained on the TMT showed a significant difference between groups (Table 2 and Table 3). AD patients required more time than controls to complete both parts of the TMT (Time A (t = -4.80, p < .000), Time B (t = -7.76, p < .000), Time A+B (t = -8.099, p < .000) and Time B-A (t = -6.412, p < .000)). Regarding the number of errors AD group stood apart from the controls [Error A (U =519.0, p <.000) and Error B (U =146.5, p < .000)]. Healthy controls made more spontaneous corrections on Part B (U = 984.0, p < 0.030), whereas healthy group and AD patients was similar in terms of Correction A (U = 1201.5, p = 0.417) (Table 4).

According to *ROC* analysis results, threshold value for Time A is estimated as 108 seconds. At this level, the test has a sensitivity of 0.66 and a specificity of 0.80. The positive and negative predictive values are 0.77 and 0.70, respectively.

Threshold value for Time B was 240 seconds. At this level, the test had a sensitivity of 0.84 and a specificity of 0.90. The positive and negative predictive values were 0.89 and 0.85, respectively.

Table 2— Mean and Standard Deviations of TMT Time Scores

				Control (n=50)		AD (n=50)	
	T	df	p	M	SD	M	SD
Time A	-4.80*	98	.000	88.04	32.54	125.20	44.02
Time B	-7.76*	98	.000	178.18	51.52	345.76	143.76
Time A+B	-8.09*	98	.000	266.22	75.42	470.96	162.06
Time B-A	-6.41*	98	.000	90.14	41.68	220.56	137.65

*p< .001.

We conducted pairwise comparisons among curves to determine if the difference between AUC were significant (Hanley & Mc Neil, 1982). A significant difference was found between the AUC for the Time A and Time B, indicating that AUC for the Time B was significantly larger than for the Time B; and so we can deduce that Time B determines AD patients with a higher specificity and sensitivity than Time A.

The estimated threshold value for (Time A +B) was 356 seconds. At this level, the test had a sensitivity of 0.82 and a specificity of 0.90. The positive and negative predictive values were 0.89 and 0.83, respectively.

The estimated threshold value for Time B-A was 112 seconds. At this level, the test had a sensitivity of 0.92 and a specificity of 0.78. The positive and negative predictive values were 0.81 and 0.91, respectively.

Table 4 outlines the sensitivity, specificity, positive and negative predictive values of time scores and Table 5 shows the number of subjects on True Positive (*TP*), False Positive (*FP*), True Negative (*TN*), False Negative (*FN*) areas or various threshold values for TMT subscores.

Regarding the relation of FAQ scores with the time to complete the TMT Part A and B, Time A+B and Time B-A, we observed no significant correlation in normal healthy group ($N = 50$, $p > 0.05$). However, in AD patients, we found a statistically significant correlation between FAQ scores and Time A score ($N = 50$, $r = 0.30$, $p < 0.05$); Time B score ($N = 50$, $r = 0.29$, $p < 0.05$) and Time A+B scores ($n = 50$, $r = 0.21$, $p < 0.05$). There were no significant correlation between FAQ scores and differential time scores ($N = 50$, $p > .05$). We observed that Time A, B and A+B scores increased in parallel to FAQ scores. There were no significant relationship between FAQ scores and error and correction frequencies of TMT Part A and B in both groups ($N = 50$, $p > .05$).

Discussion

Although memory seems the primary affected cognitive domain in AD, various components of attention and executive functions are also proposed to be involved. From this point of view, either Part A that is sensitive to simple atten-

Table 3— Results of TMT Correction and Error Scores (Mann Whitney-U test) in AD and Control Groups

	N	M	Total	U	p
Correction A					
Normal-Control	50	49,53	2476,50	1201,50	.41
ATD	50	51,47	2573,50		
Correction B					
Normal-Control	50	45,18	2259,00	984,00*	.03
ATD	50	55,82	2791,00		
Error A					
Normal-Control	50	35,88	1794,00	519,00**	.000
ATD	50	65,12	3256,00		
Error B					
Normal-Control	50	28,43	1421,50	146,50**	.000
ATD	50	72,57	3628,50		

**p <.001; *p <.05.

Table 4— Sensitivity, Specificity, Positive and Negative Predictive Values of Time Scores

Cut-off Value	Sensitivity	Specificity	+LR	-LR	+PV	-PV
Time A						
>97	.66	.72	2.36	.47	.70	.68
>108*	.66	.80	3.30	.42	.77	.70
>110	.62	.82	3.44	.46	.78	.68
Time B						
>227	.84	.82	4.67	.20	.82	.84
>240*	.84	.90	8.40	.18	.89	.85
>241	.82	.90	8.20	.20	.89	.83
Time A+B						
>351	.82	.88	6.83	.20	.87	.83
>356*	.82	.90	8.20	.20	.89	.83
>363	.80	.90	8.00	.22	.89	.82
Time B-A						
>111	.92	.76	3.83	.11	.79	.91
>112*	.92	.78	4.18	.10	.81	.91
>113	.90	.78	4.09	.13	.80	.89

Note. +LR, Positive Likelihood Ratio; -LR, Negative Likelihood Ratio; +PV= Positive predictive value; -PV, Negative predictive value; * cut-off point.

tion or Part B that is sensitive to executive functions such as sustained attention, set shifting and interference, that are short, easily administered neuropsychological tests are usable in the differential diagnosis of AD. With this feature, TMT is a tool that might be included in the evaluation of attention and executive functions in AD.

In this study, we analyzed not only the time required to complete the test but also error and correction scores (frequencies) which is not assessed in previous studies. It is observed that all components of the TMT on time; namely Time A, Time B, Time A+B, and Time B-A; differentiates healthy adults from the AD patients. This finding is consistent with the previous studies (10,13). *ROC* analyses for Time A and B revealed that Time B score differentiates AD patients from healthy controls with a higher specificity and sensitivity. We observed the same finding for also Time A+B score.

Our findings showed that, TMT error and correction scores, Error A, Error B and Correction B are useful in identifying AD patients. Our results confirm the previously reported proposals that not only the time scores but also error and correction scores obtained from the TMT are efficacious in detecting functional brain disorders. However, the previous validity studies did not evaluate correction scores (12,14,16). The originality of our study is that AD patients and healthy persons were compared for correction scores as well as time and error scores. Our findings are in line with previous validity studies that have demonstrated that besides time scores (Time B superior to Time A), Part A and B error scores (Error B superior to Error A) are important measures for clinical diagnosis of AD (14,15).

In conclusion, in this study we looked at the time to complete the test while scoring the TMT and also the number of the errors and corrections. Our findings support that error (Error A and B) and correction (Correction B) scores are as significant as the TMT time scores (Time A, B, A+B and B-A) in clinical diagnosis of AD. However, we did not analyze the error and correction subtypes (perseveration, sequential etc.) that might be handled in future studies.

Table 5— Cut-off Values for TMT Time Scores

Cut-off value	TP	FP	TN	FN
Time A				
>108	33	10	40	10
Time B				
>240	42	5	45	8
Time A+B				
>356	41	9	45	5
Time B-A				
>112	46	4	39	11

Note. TP, True Positive; FP, False Positive; TN, True Negative; FN, False Negative.

We also estimated specificity, sensitivity and cut-off values for the time to complete the test. We found all time scores to have a high specificity and sensitivity in the diagnosis of AD in the elderly. Our results confirm that Turkish version of the TMT can be used as a useful tool in assessing attention and executive functions in Turkish AD patients.

In summary, the present study described TMT scores in a sample of well-characterized Turkish healthy elderly and AD patients. , The relationship between FAQ scores and eight TMT subscores were evaluated as well. Our findings show that in AD patients TMT (Part A and B) scores were related with FAQ scores, and an increase in FAQ scores may predict the increase of time required to complete the TMT (Part A and B). This finding indicates that this relationship can be used as a screening tool not only for AD patients but also for healthy elderly population.

REFERENCES

1. Spreen O, Strauss E. A compendium of neuropsychological tests: Administration, norms and commentary. New York: Oxford University Press, 1991, pp 98–121.
2. Bowie CR, Harvey PD. Administration and interpretation of the Trail Making Test. Nature Protocols 2006;1(5):2277-81. (PMID:17406468).
3. Ivnik JR, Malec JF, Smith GE, Tangalos EG, Petersen RC. Neuropsychological tests norms about age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. The Clinical Neuropsychologist 1996;10:262-78.
4. Tombaugh TN. Trail Making Test A and B. Normative data stratified by age and education. Archieves of Clinical Neuropsychology 2004 Mar;19(2):203-14. (PMID:15010086).
5. Steinberg BA, Breliauskas LA, Smith GE, Ivink RJ. Mayo's older Americans normative studies: Age and IQ-adjusted norms for the Trail Making Test, The Stroop Test and controlled Oral Word Association Test. The Clinical Neuropsychologist 2005;19:329-77. (PMID:16120535).
6. Davis A. The influence of age on Trail Making Test performance. Journal of Clinical Psychology 1986;24:96-8. (PMID:5639473).
7. Ruffolo LF, Guilmette TJ, Willis WG. Comparison of time and error rates on the Trail Making Test among patients with head injuries, experimental malingerers, patients with suspect effort on testing, and normal controls. The Clinical Neuropsychologist 2000;14:223-30. (PMID:10916197).
8. Stuss DT, Bisschop SM, Alexander MP, Levine B, Izukawa D. The Trail Making Test: A study in focal lesion patients. Psychological Assessment 2001;13:230-9. (PMID:11433797).
9. Albert MS, Blacker D. Mild cognitive impairment and dementia. Annual Review of Psychology 2006;2:379-88. (PMID:17716075).
10. McKhann O, Drachman D, Folstein M, Katzman R, Price D, Stradian EM. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspcies of department of health and human services task force on Alzheimer's disease. Neurology 1984;34:939-44. (PMID:6610841).
11. Duchek D, Balota DA. Failure to control prepotent pathways in early dementia of Alzheimer's type: Evidence from dichotic listening. Neuropsychology 2005;19(5):687-95. (PMID:16187887).
12. Baddely AD, Baddely HA, Bucks RS, Wilcock GK. Attentional control in Alzheimer's disease. Brain 2011;124:14492-508. (PMID:11459742).
13. Meguro K, Constans JM, Shimada M, Yamaguchi S, Ishizaki J, Ishı H, Yamdori A, Sehita Y. Corpus callosum atrophy, white matter lession, and frontal executive dysfunction in normal agning and Alzheimer's disease. A community based study: The Tajiri Project. International Psychogeriatrics 2003;15(1):9-25. (PMID:12834197).
14. Stokholm J, Vogel A, Gode A, Voldemor G. Heterogenity in executive impairment in patients with very mild Alzheimer's disease. Dementia and Geriatric Cognitive Disorders 2006;22:54-9. (PMID:16682794).
15. Ashendorf L, Jefferson A, O'Connor MK, Chaisson C, Green RC, Stern RA. Trail Making Test errors in normal aging, mild cognitive impairment, and dementia. Archieves of Clinical Neuropsychology 2008;23:129-37. (PMID:18178372).
16. Amieva H, Lafont S, Auriacombe S, Rainville C, Orgogozo JM, Dortigues JF, Fabrigoule C. Analysis of error types in the Trail Making Test evidences an inhibitory deficit in dementia. Journal of Clinical and Experimental Neuropsychology 1988;20(2):280-5. (PMID:9777482).
17. Selekler K, Cangöz B, Karakoç E. İşlevsel Faaliyetler Anketi'nin 50 yaş ve üzeri grupta Türk Kültürü İçin Uyarlama ve Norm Belirleme Çalışması. Türk Nöroloji Dergisi 2004;10(2):102-7.
18. Morris JC. The Clinical Dementia Rating (CDR): Current version and scoring rules. Neurology 1993;43:2412-4. (PMID:8232972).
19. Güngen C, Ertan T, Eker E, Yaşar R, Engin F. Standardize mini mental testin Türk toplumunda hafif demans tanısında geçerlik ve güvenilirliği. Türk Psikiyatri Dergisi 2002;13(4):273–81. (PMID:12794644).
20. Ertan T, Eker E. Reliability, validity, and factor structure of the geriatric depression scale in Turkish elderly: are there different factor structures for different cultures? Int Psychogeriatr 2000 Jun;12(2):163-72. (PMID:10937537).
21. Kudiaki Ç, Aslan A. The Three Words-Three Shapes Test: Normative data for Turkish elderly. Archieves of Clinical Neuropsychology 2007;22(5):637-45. (PMID:17521867).
22. Cangöz B, Karakoç E, Selekler K. Saat Çizme Testi'nin Türk yetişkin ve yaşlı örneklemi üzerindeki norm belirleme, geçerlik ve güvenirlik çalışmaları. Turkish Journal of Geriatrics 2006;9(3):136-42.
23. Cangöz B, Karakoç E, Selekler K. Trail Making Test: Normative data for Turkish elderlys by age, sex and education. Journal of the Neurological Sciences 2009;283(1-2):73-8. (PMID:19264326).