DOI:10.4274/tnd.26504
Turk J Neurol 2018;24:143-152

Original Article / Özgün Araştırma

Validity, Reliability and Turkish Norm Values of the Clock Drawing Test for Two Different Scoring Systems

Saat Çizme Testinin İki Farklı Puanlama Sisteminin Türkiye Normları ve Geçerlik-Güvenirlik Çalışması

Derya Durusu Emek-Savaş[1], Deniz Yerlikaya[2], Görsev G. Yener[3]

[1]Dokuz Eylul University Faculty of Letters, Division of Psychology, Izmir, Turkey

[2]Dokuz Eylul University Institute of Health Sciences, Department of Neurosciences, Izmir, Turkey

[3]Dokuz Eylul University Faculty of Medicine, Department of Neurology, Izmir, Turkey

Abstract

Objective: The clock drawing test (CDT) version with the pre-drawn circle has been widely used in research and clinical practice without standardized Turkish norms. The present study aimed to standardize CDT scores according to the most frequently used scoring methods in the literature (Manos-Wu and Shulman) and to estimate the validity and reliability of both methods.

Materials and Methods: The norm determination phase of the study was performed with 244 healthy individuals in the age range of 50-92 years. The effects of age, education, and sex on the CDT scores were examined and normative data stratified by age and education were derived. Test-retest reliability, inter-rater reliability, and concurrent validity and criterion validity of the two scoring methods were tested. Criterion validity was assessed using the ROC analysis to examine the extent to which the CDT scores could distinguish among healthy individuals, patients with mild cognitive impairment (MCI), and patients with Alzheimer's disease (AD).

Results: In the Manos-Wu method, age and education had significant effects on CDT scores of healthy individuals; however, in the Shulman method, only an education effect was found. Both methods had high test-retest reliability (Manos-Wu, 0.81; Shulman, 0.72) and inter-rater reliability (Manos-Wu, 0.98; Shulman, 0.96) and showed high correlations with each other and with other cognitive screening tests. The area under the ROC curve had high values only in discriminating between healthy individuals and AD, and healthy individuals and MCI.

Conclusion: In the present study, detailed normative data for individuals aged 50 and older were established. Age and education levels should be taken into account when interpreting CDT scores. Both scoring methods provided similar results in terms of high test-retest and inter-rater reliability, as well as high concurrent and criterion validity.

Keywords: Clock drawing test, normative data, validity-reliability, Alzheimer's disease, mild cognitive impairment

Öz

Amaç: Saat çizme testi (SÇT) ülkemizde çok sayıda merkezde uygulanmakta olup saatin dairesinin hazır olarak sunulduğu versiyonun Türkiye popülasyonuna ait norm değerleri bulunmamaktadır. Bu çalışmada, SÇT'nin uluslararası literatürde en yaygın kullanıma sahip skorlama yöntemlerinden Manos ve Wu ile Shulman yöntemleri ile standardizasyonunun yapılması ve geçerlik ve güvenirliğinin test edilmesi amaçlanmıştır.

Gereç ve Yöntem: Çalışmanın norm belirleme aşamasına 50-92 yaş aralığındaki 244 sağlıklı birey dahil edilmiştir. Yaş, eğitim ve cinsiyet değişkenlerinin SÇT puanları üzerindeki etkisi incelenmiş, yaş ve eğitime göre norm değerleri oluşturulmuştur. Ayrıca, iki puanlama yönteminin test-tekrar test güvenirliği, değerlendiriciler arası güvenirliği, eş-zaman geçerliği ve ölçüt geçerliği test edilmiştir. Ölçüt geçerliği için ROC eğrisi yöntemi kullanılarak, SÇT puanlarının sağlıklı bireyler, hafif kognitif bozukluk (HKB) olguları ve Alzheimer hastalarını ne düzeyde ayırt edebildiği incelenmiştir.

Bulgular: Manos ve Wu yönteminde sağlıklı bireylerin SÇT puanları üzerinde yaş ve eğitim etkisi, Shulman yönteminde ise sadece eğitim etkisi olduğu saptanmıştır. Her iki yöntemin de test-tekrar test (Manos ve Wu için, 0,81; Shulman için, 0,72) ve değerlendiriciler arası güvenirliği (Manos ve Wu için, 0,98; Shulman için, 0,96) yüksek olup, birbirleriyle ve diğer kognitif tarama testleriyle yüksek düzeyde korelasyon gösterdiği bulunmuştur. ROC eğrisi altında kalan alan, sadece sağlıklı bireyler ile Alzheimer hastalarını ve sağlıklı bireyler ile HKB olgularını birbirinden ayırt etmede yüksek değerler göstermiştir.

Sonuç: Bu çalışmada, 50 yaş ve üzeri bireyler için ayrıntılı norm değerleri oluşturulmuştur. SÇT skorları yorumlanırken yaş ve eğitim düzeyi dikkate alınmalıdır. Her iki puanlama yönteminden elde edilen bulgular oldukça paralel olup, yüksek test-tekrar test ve değerlendiriciler arası güvenirlik ile yüksek eş-zaman ve ölçüt geçerliklerine işaret etmektedir.

Anahtar Kelimeler: Saat çizme testi, normatif veri, geçerlik-güvenirlik, Alzheimer hastalığı, hafif kognitif bozukluk

Address for Correspondence/Yazışma Adresi: Derya Durusu Emek-Savaş MD, Dokuz Eylul University Faculty of Letters, Division of Psychology, Izmir, Turkey
Phone: +90 536 950 20 50 E-mail: derya.emek@deu.edu.tr ORCID ID: orcid.org/0000-0001-7042-697X
Received/Geliş Tarihi: 21.09.2017 **Accepted/Kabul Tarihi:** 07.12.2017

Introduction

Neuropsychological evaluation is accepted as the gold standard method in the evaluation of cognitive functions. Tests that form the basis of neuropsychological assessment find functional application areas in diagnosis, follow-up, treatment planning, and treatment efficacy evaluation processes of diseases affecting cognitive functions. Many cognitive functions such as memory, attention, response inhibition, planning, language, abstract thinking, construction and praxis can be evaluated with neuropsychological tests (1). However, demographic variables such as age, sex, education, and cultural characteristics are known to influence neuropsychological test scores (2). Therefore, Turkish adaptation and standardization studies are of great importance for the correct interpretation of these tests.

The clock drawing test (CDT) is an easy-to-use, fast and reliable paper-and-pencil test that has been used for over 30 years for the evaluation of planning, ranking, and abstract thinking skills and visual-motor functions (3,4,5,6,7).

Many different application and scoring methods have been developed since the first use of CDT (8). In a classic CDT application, an individual is instructed to draw a clock, put in all the numbers, and set the hands for a specified time (e.g. 11:10, 15:00, 18:45). Each step in the application is associated with a distinct cognitive function. For this reason, the differences in the application of CDT may result in focusing on the cognitive process being assessed or ignoring one of these functions. Many studies in the literature have compared CDT application and scoring methods. In different CDT applications, the patient may be provided a pre-drawn circle (3,7,8,9), or the patient may be asked to draw the circle (e.g., 10). In CDT application, an individual may be asked to show one (3,4,5,6) or several different times (8), or the clock representation may not be evaluated at all (6,9). In addition to the application differences of the test, there are also many different scoring systems in the literature (11). Shulman, Manos-Wu, and Sunderland are the most frequently used scoring methods in the studies. In the Shulman scoring method, the individual is provided with a pre-drawn circle and asked to set the hands for 11:10. The method has a score range of 0-5 and 5 is the highest score possible. The Manos-Wu method is based on a total of 10 points and is similar to the Shulman method in terms of application. In the Sunderland method, there is an evaluation over 10 points, in which the clock circle is drawn by the person and this level is included in the scoring as well.

Cangoz et al. (12) conducted a norm determination and validity-reliability study of CDT in Turkish population aged 50 years and over. In this study, a four-point scoring method for CDT, in which the individual is asked to draw the circle, became the standardized sample for Turkey. Cangoz et al. (12) reported that they preferred this method because of its functionality in distinguishing patients with dementia and its easy scoring system. However, the CDT with the pre-drawn circle is used in many clinics in Turkey and there is no standardized scoring system in Turkish regarding this application format. The fact that an effect of age (13,14,15) and education (14,16,17) on CDT has been demonstrated in many studies has increased the need for normative data, especially in the clinical field.

The CDT has been reported to have high sensitivity and specificity ratios, especially in distinguishing dementia syndromes (13,18). The sensitivity and specificity of the Shulman method were reported as 86% and 72%, respectively, and 76% and 78%, respectively, for the Manos-Wu method (18). In a study conducted in Turkey, three different CDT scoring methods were used to differentiate between elderly patients with and without dementia and high sensitivity and specificity values were reported for the Shulman method (19). In the same study, it was stated that the use of CDT in conjunction with the mini mental state examination (MMSE) increased the sensitivity to 100%. However, the discriminatory validity of the CDT was found to be lower for Alzheimer's disease (AD)-related mild cognitive impairment (MCI) (13).

In this study, we aimed to standardize the test and evaluate the validity and reliability of the CDT format currently applied in many centers in Turkey, which has no norm values for the Turkish population, based on the two most commonly used scoring systems in the international literature, the Shulman and Manos-Wu scoring methods. In addition, the sensitivity and specificity values of these scoring systems were determined for MCI and AD groups, and their discrimination levels were compared.

Materials and Methods

Sample

Two hundred forty-four healthy individuals aged 50 years and over, who were determined to have no cognitive impairment according to the results of clinical examination, brain imaging, and neuropsychological evaluations, and who were followed up longitudinally between 2011-2017 at Dokuz Eylul University Institute of Health Sciences Neuroscience Department, were included in the study.

In neuropsychological tests performed by specialist neuropsychologists routinely with one year interval; (1) general cognitive status is evaluated with MMSE (20,21), (2) verbal and visual memory with the Öktem Verbal Memory Processes Test (22) and Wechsler Memory Scale Revised (WMS-R) visual production subtest (23), (3) attention with WMS-R forward and reverse number range test (23), executive functions with Wechsler Adult Intelligence Scale (WAIS-IV) similarities subtest (24), verbal fluency tests (semantic and phonemic) (25), CDT and Stroop test (26), (5) visual-spatial functions with simple shape copying tests, and (6) language skills with the Boston Nomenclature test (27).

The level of education for the study population was at least literate. Healthy individuals with an MMSE score below 27 were not included in the study. Also, uncontrolled disease history, head trauma, and uncorrected visual impairment that might affect cognitive skills were identified as exclusion criteria. All subjects included in the study had been assessed in previous studies both by a physician and with the Yesavage Geriatric Depression Scale (GDS) (28) for depressive symptoms and depressive symptoms. Individuals with a Yesavage GDS score of 13 and above were not included in the study. Demographic characteristics of healthy subjects from whom normative data were obtained are presented in Table 1.

In order to perform validity and reliability analyses, 128

patients with amnestic MCI and 109 patients with Alzheimer's disease who had been routinely followed with neurological examinations, neuropsychological tests, and structural magnetic resonance imaging in every 12 months on average were included in this study. The diagnostic criteria of Albert et al. (29) for MCI and McKhann et al. (30) for AD were used. At different stages of the study, different numbers of patients with MCI and AD were included. The CDT performances of healthy individuals, and patients with MCI and Alzheimer's from past years were scored and included in the analyses for this study. Written informed consent was obtained from all participants from the previous research studies and the study protocol was approved by the ethics committee of Dokuz Eylul University on 13.07.2017 (reference number: 2017/18-01).

Materials

In this study, the previous CDT performances of participants were evaluated using the Manos and Wu (5) and Shulman (4) scoring methods. In the CDT version used, an empty circle with a diameter of 10 cm is given on paper, and the individual is asked to place the clock numbers and set the clock at 11:10 hrs. It is thought that faults related to clock representation might be functional in detecting early cognitive impairment (31). For this reason, particular attention is paid not to mention terms such as short hand, long hand or clock hands regarding clock representation, when the instructions are given. When needed, all or parts of the instructions are repeated in the same way for reminder purposes. The selected scoring methods are preferred because the circle is pre-drawn, the clock is set at 11:10 hrs and there is no need for copying. The Manos-Wu and Shulman scoring methods are introduced in detail in the subheadings.

The Manos-Wu Scoring Method

The scoring method developed by Manos and Wu (5) is listed in the literature as the Ten-Point Clock Test. In this method, an evaluation is made over a total of 10 points, and a transparent circle template is used to help evaluate eight equal parts (Annex 1). One point is given for each of the following numbers that falls in its proper zone of the circle: 1, 2, 4, 5, 7, 8, 10 and 11. One point is given each to a short hand pointing at the number eleven, and a long hand pointing at the number two. If the lengths of the arms are incorrect, no points are awarded. The scoring formula can be expressed as [8 (numbers) + 1 (short hand) + 1 (long hand)=10].

In addition to the total CDT scores of the individuals, the number placement scores over 8 points and the clock representation scores over 2 points were also calculated in order to examine the CDT performances in detail.

The Shulman Scoring Method

In this method, CDT performance is scored in the range of 0-5, and the highest score that can be achieved from the test is 5. Five points are awarded for a perfect clock representation and number placement; 4 points for a clock containing minor visuospatial errors; 3 points for acceptable visuospatial organization but inaccurate representation of a clock; 2 points for visuospatial disorganization of numbers that will not allow the right clock representation; 1 point for a severe level of visuospatial disorganization; and 0 point for no clock representation.

Only the total scores of the individuals were calculated and evaluated because there are no sub-score types in the Shulman method.

Statistical Analysis

Statistical analyses were performed using the SPSS 20.0 program. ROC curve analyses were performed using the MedCalc 15.8 program.

Norm determination: First, the effect of age, education, and sex variables on CDT scores was assessed through stepwise multiple regression analysis in the norm determination step. Norm values were based on age and education variables that had an effect on CDT scores. Statistical differences between age and education groups were examined using two-way ANOVA. According to a 3x3 ANOVA design, age [(3 levels): 50-59, 60-69, and over 70 years] and education [(3 levels): 0-5 years, 6-11 years, and 12 years and over] were included in the analysis. Two-way ANOVA was repeated for both total scores of two scoring methods and sub-score types of Manos-Wu method. In post hoc analyses, the main effects of age and education were examined using the independent sample t-test, and the interaction effect of age x education using the paired- samples t-test.

The norm values for the total scores of two scoring methods and sub-scores of Manos-Wu method were based on three age groups (50-59, 60-69, and over 70 years) and three education groups (0-5 years, 6-11 years, and 12 years and over), and the mean and standard deviation values for the groups for each score type are reported.

Test-retest reliability: In order to evaluate the test-retest reliability, the first and second CDT performances, with an average interval of 12 months, of 50 randomly selected healthy participants among longitudinally monitored individuals were evaluated

Table 1. Demographic characteristics of normative sample

	Total (n=244)		Age groups					
			50-59 years (n=58)		60-69 years (n=97)		+70 years (n=89)	
	Range	Mean±SD	Range	Mean±SD	Range	Mean±SD	Range	Mean±SD
Age (years)	50-92	66.06±8.67	50-59	55.07±2.82	60-69	64.31±3.25	70-92	75.12±4.96
Education (years)	0-23	12.11±4.32	7-21	13.62±2.62	0-23	11.87±4.57	0-21	11.40±4.73
Sex (M/F)	89/155		15/43		42/55		32/57	
Handedness (Right/Left)	239/5		58/0		94/3		87/2	
MMSE	27-30	29.04±1.09	27-30	29.22±0.90	27-30	29.05±1.07	27-30	28.92±1.19

Data are presented as mean ± standard deviation. M: Male, F: Female, MMSE: Mini mental state examination, SD: Standard deviation

within the scope of the current study. The agreement between the first and second assessments was evaluated using interclass correlation (ICC) analysis. In the ICC analyses, an absolute fit was tested with a two-way mixed model.

Inter-rater reliability: In order to test the inter-rater reliability, the CDT performance of a total of 100 participants, 35 healthy individuals, 33 patients with MCI, and 32 patients with Alzheimer's, was independently rated by a second rater. The absolute agreement between the two evaluators was analyzed using ICC analysis. In the ICC analyses, absolute agreement was tested with a two-way mixed model. The inter-rater reliability results were reported following calculation of ICC coefficients for all groups together and for each group separately.

Concurrent validity: Correlation between CDT scores (for both scoring methods) and concurrent MMSE scores of all participants were examined. Pearson's correlation coefficient was calculated for healthy participants, MCI and Alzheimer's disease groups together, and for each group seperately. In addition, the correlation of scoring methods with each other was examined.

Criterion validity: The Manos-Wu and Shulman methods were evaluated using ROC analysis in terms of distinguishing healthy-MCI, healthy-AD, and MCI-AD groups from each other. The healthy-MCI, healthy-AD, and MCI-AD groups were formed and compared including education and age-matched individuals because CDT scores are known to be affected by education and age variables. For the groups, cut-off points, sensitivity and specificity values were determined and criterion validity was tested.

Results

Norm Determination

The effect of age, sex, and educational levels on CDT scores was examined using stepwise regression analysis. First, total score, clock representation, and figure placement scores according to the Manos-Wu scoring system were evaluated. The predictive contribution of age and education levels was found as 6.9% in the total CDT scores calculated according to the Manos-Wu method [$F_{(2,244)}$=9.111, p<0.001]. Sex, as a variable, was found to have no significant contribution to the model. The total CDT scores calculated according to the Manos-Wu method were found to decrease by 0.011 points with each year of age and increased by 0.030 points with each yearly increase in education level. Of the total variance of the clock representation scores, 4.6% was explained by education level [$F_{(1,245)}$=11.848, p=0.001] and there was no significant contribution of age or sex to the model on this score type; 3.1% of the total variance of the number placement scores was explained by age [$F_{(1,245)}$=7.814, p=0.006] and this time there was no significant contribution of education or sex to the model for this score type.

Regression analyses were repeated for the total CDT scores calculated using the Shulman method and only a significant contribution of the education level was determined; education level explained 3.5% of total Shulman scores [$F_{(1,245)}$=9.006, p=0.003].

Statistical differences between the levels of age (50-59, 60-69, and over 70 years) and education (0-5 years, 6-11 years, and over 12 years) variables were assessed using two-way ANOVA. Sex variation was not included in the ANOVA design as there was no sex effect on the total scores of both methods and sub-scores of Manos-Wu method, and the analyses were performed only on the 3x3 design with age and education levels included.

In the ANOVA analysis, the main effect on the Manos-Wu total scores was observed with age [$F_{(2,236)}$=4.969, p=0.008]. In post hoc analyses, the CDT scores of patients aged 70 years and over were found to be significantly lower than the 50-59 years group (p=0.003) and the 60-69 years group (p=0.006). There was no significant difference between the 50-59 and 60-69 years groups (p=0.633).

In addition, the main effect on the total scores of Manos-Wu was observed with education level [$F_{(2,236)}$=4.253, p=0.015]. In the post hoc analyses, it was determined that the total scores of individuals with 12 years and over education were found to be higher than those of individuals with 0-5 years' (p=0.016) and 6-11 years' (p=0.026) education. There was no significant difference between the scores of individuals with 0-5 years and 6-11 years' education (p=0.379).

Although the main effect on Manos-Wu number placement scores was found with age [$F_{(2,236)}$=3.748, p=0.025], there was no major effect of education [$F_{(2,236)}$=2.415, p=0.092]. In the post hoc analyses, it was found that the number placement scores of the individuals aged 70 years and over were significantly lower than those of the 50-59 years group (p=0.022) and the 60-69 years group (p=0.010). There was no significant difference between the 50-59 and 60-69 years groups (p=0.763). In addition, there was an age x education interaction effect on number placement scores [$F_{(3,236)}$=4.020, p=0.008]. The number placement scores of individuals aged 70 years or older with 6-11 years of education level were found to be lower than those of the 50-59 years (p=0.024) and 60-69 years (p=0.030) groups. There was no significant difference between the other education levels and age groups (p>0.05).

Education was found to have a main effect on Manos-Wu clock representation scores [$F_{(2,236)}$=3.809, p=0.024], but age had no major effect [$F_{(2,236)}$=2.409, p=0.092]. The clock representation scores of individuals with 0-5 years of education level were significantly lower than those of with 12 years and over education (p=0.030). There was no significant difference between the other education levels (p>0.05).

Age [$F_{(2,236)}$=3.248, p=0.041] and education level [$F_{(2,236)}$=3.510, p=0.031] were found to have the main effects on the total CDT scores calculated using the Shulman method. The total scores of individuals aged 70 years and over were lower than those of the 50-59 years group (p=0.030) and the 60-69 years group (p=0.013). In addition, the scores of individuals with 0-5 years' education were significantly lower than those of with 12 years and over education (p=0.032). The ANOVA findings of the total scores of both methods are summarized in Table 2.

In this study, it was difficult to reach a population within the age range of 50-59 years and with 0-5 years' education, so the data of 3 healthy individuals in this group were excluded from the analyses. This situation was thought to be related to the fact that compulsory primary school education was introduced in Turkey. The norm values for the mentioned age group have not been established and the norms of the individuals between the ages of 50-59 years have

been reported for only education levels of 6-11 years and 12 years' and over. The norm values obtained are presented in Tables 3, 4, 5 and 6 for the Manos-Wu and Shulman methods.

Test-Retest Reliability

The test-retest reliability coefficient of total CDT scores calculated using the Manos-Wu method was found as 0.806, and the test-retest reliability coefficient of CDT scores calculated with the Shulman method was found as 0.720. Reliability coefficients and 95% confidence intervals for detailed score types are presented in Table 7.

Inter-rater Reliability

Inter-rater reliability was calculated for two independent raters scoring the CDT scores of a total of 100 participants, consisting of healthy subjects, and patients with MCI and AD. When all the participants were evaluated together, the inter-rater agreement coefficient of total scores calculated using the Manos-Wu method was 0.892, and the inter-rater agreement coefficient of total scores calculated with the Shulman method was found as 0.957. Reliability coefficients and 95% confidence intervals for detailed score types are presented in Table 8.

Concurrent Validity

Correlation analyses were performed by examining healthy participants, and patients with MCI and AD together and groups separately. A high correlation was found between MMSE scores and Manos-Wu total scores ($r=0.797$, $p<0.001$) and Shulman total scores ($r=0.760$, $p<0.001$) in the analyses involving all participants. The correlations of all scores and groups are presented in Table 9. In addition, there was a strong correlation between the total CDT scores calculated using the Manos-Wu method and the Shulman method ($r=0.929$, $p<0.001$).

Criterion Validity

The total scores calculated according to both scoring methods were examined using ROC analysis in terms of distinguishing healthy-MCI, healthy-AD and MCI-AD groups from each other. Demographic information of age and education-matched groups is presented in Table 10.

The ROCs with which two methods were compared were prepared for all groups (Figure 1). The area under the curve and 95% confidence intervals for both methods are shown in Table 11. Manos-Wu differentiated the healthy-MCI ($p<0.0001$) and healthy-AD ($p<0.0001$) groups significantly better than the Shulman method. There was no significant difference between the two methods in terms of differentiating MCI-AD groups ($p=0.721$).

In the ROC analysis, no group comparisons contained 0.5 values regarding 95% confidence intervals. This means that the discriminating powers of the scoring methods do not depend on chance.

In the analyses, cut-off points and corresponding sensitivity and specificity values were calculated for the Manos-Wu and Shulman scoring methods. When the cut-off point was determined as 9/10 in the Manos-Wu scoring method, the sensitivity of the test discriminating patients with MCI from healthy individuals was 80.8% and its specificity was 80.0%, and its sensitivity and specificity for discriminating patients with AD from healthy individuals were 89.8% and 76.8%, respectively. The sensitivity and specificity values were found as 62.9% and 61.9%, respectively, when the cut-off point was determined as 7/10 for distinguishing patients with MCI and AD.

When the cut-off point was determined as 4/5 in the Shulman scoring method, the sensitivity of the test to discriminate patients with MCI from healthy individuals was 81.5% and its specificity

Table 2. The effect of demographic variables on the total scores of clock drawing test

	Variables	Mean ± SD	F	Post hoc
Manos-Wu total score	**Age**			
	A: 50-59	9.90±0.35		A=B, p=0.633
	B: 60-69	9.87±0.42	$F_{(2.236)}$=4.969, p=0.008	A>C, p=0.003*
	C: >70	9.57±0.90		B>C, p=0.006*
	Education			
	A: 0-5	9.51±0.85		A=B, p=0.379
	B: 6-11	9.66±0.80	$F_{(2.236)}$=4.253, p=0.015	B<C, p=0.026*
	C: >12	9.89±0.44		A<C, p=0.016*
Shulman total score	**Age**			
	A: 50-59	4.90±0.35		A=B, p=0.858
	B: 60-69	4.91±0.35	$F_{(2.236)}$=4.488, p=0.012	A>C, p=0.030*
	C: >70	4.72±0.62		B>C, p=0.013*
	Education			
	A: 0-5	4.63±0.73		A=B, p=0.222
	B: 6-11	4.80±0.50	$F_{(2.236)}$=3.188, p=0.043	B=C, p=0.084
	C: >12	4.91±0.36		A<C, p=0.032*

Data are presented as mean ± standard deviation. SD: Standard deviation, *p<0.05

was 79.6%, and its sensitivity and specificity in discriminating patients with AD from healthy individuals were 61.6% and 83.2%, respectively. The cut-off point of 3/5 to differentiate between MCI and AD groups showed low sensitivity and specificity (61.9% and 59.8%, respectively). The cut-off points for healthy-MCI, healthy-AD, and MCI-AD groups are summarized in Table 12.

Table 3. Time representation scores calculated using Manos-Wu method

Education level (years)	Age groups (year) 50-59	60-69	70+	All age groups
0-5 years	-	1.80±0.56	1.55±0.83	1.66±0.73
(n)		(15)	(20)	(35)
6-11 years	1.94±0.24	1.84±0.45	1.77±0.59	1.84±0.47
(n)	(17)	(31)	(26)	(74)
+12 years	1.95±0.31	1.98±0.14	1.88±0.45	1.94±0.32
(n)	(41)	(51)	(43)	(135)
All education groups	1.95±0.29	1.91±0.36	1.78±0.60	1.87±0.45
(n)	(58)	(97)	(89)	(244)

Data are presented as mean ± standard deviation

Table 4. Number placement scores calculated using Manos-Wu method

Education level (years)	Age groups (year) 50-59	60-69	70+	All age groups
0-5 years	-	7.80±0.41	7.90±0.31	7.86±0.36
(n)		(15)	(20)	(35)
6-11 years	7.94±0.24	7.97±0.18	7.58±0.81	7.82±0.53
(n)	(17)	(31)	(26)	(74)
+12 years	7.95±0.22	8.00±0.00	7.88±0.39	7.95±0.25
(n)	(41)	(51)	(43)	(135)
All education groups	7.95±0.22	7.96±0.20	7.80±0.55	7.90±0.376
(n)	(58)	(97)	(89)	(244)

Data are presented as mean ± standard deviation

Table 5. Clock drawing test total scores calculated by Manos-Wu method

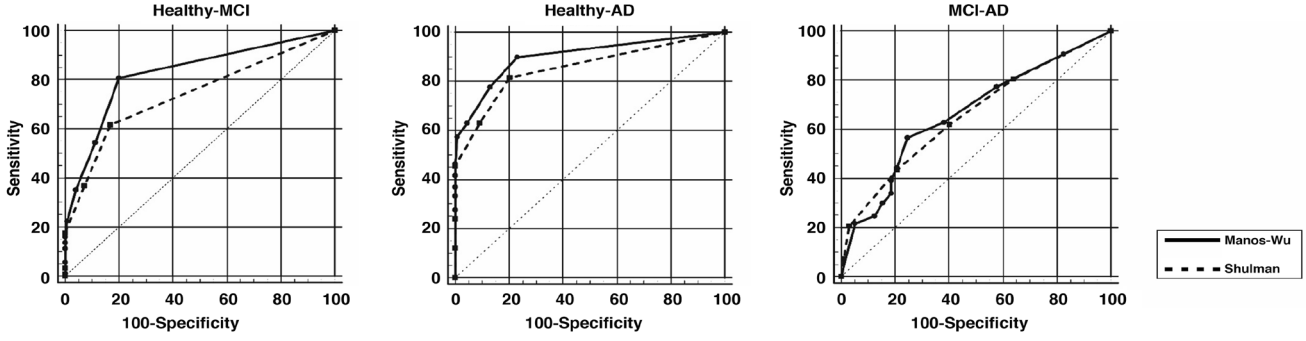
Education level (year)	Age groups (year) 50-59	60-69	70+	All age groups
0-5 years	-	9.60±0.74	9.45±0.95	9.51±0.85
(n)		(15)	(20)	(35)
6-11 years	9.88±0.33	9.81±0.48	9.35±1.16	9.66±0.80
(n)	(17)	(31)	(26)	(74)
+12 years	9.90±0.37	9.98±0.14	9.77±0.65	9.89±0.44
(n)	(41)	(51)	(43)	(135)
All education groups	9.90±0.36	9.87±0.42	9.57±0.90	9.77±0.65
(n)	(58)	(97)	(89)	(244)

Data are presented as mean ± standard deviation

Table 6. Clock drawing test total scores calculated by Shulman method

Education level (years)	Age groups (year) 50-59	60-69	70+	All age groups
0-5 years	-	4.73±0.59	4.55±0.83	4.63±0.73
(n)		(15)	(20)	(35)
6-11 years	4.94±0.24	4.84±0.45	4.65±0.63	4.80±0.50
(n)	(17)	(31)	(26)	(74)
+12 years	4.88±0.40	5.00±0.00	4.84±0.49	4.91±0.36
(n)	(41)	(51)	(43)	(135)
All education groups	4.90±0.36	4.91±0.36	4.72±0.62	4.84±0.48
(n)	(58)	(97)	(89)	(244)

Data are presented as mean ± standard deviation



Şekil 1. ROC curves for two different clock drawing test scoring methods
MCI: Mild cognitive impairment, AD: Alzheimer's disease

Discussion

The CDT, as in the rest of the world, is a widely used test by researchers and physicians in Turkey; however, the lack of a normative study data of the most commonly used version of the test in our country is remarkable. The aim of this study was to determine CDT norm values of the Turkish population aged 50 years and over according to the Manos and Wu (5) and Shulman (4) scoring methods, and to make it available for the relevant physicians and researchers. In addition, the validity and reliability of the scoring methods used were tested.

Although CDT is fairly simple, understandable, and feasible, concurrent involvement of different cognitive functions is required so that the task can be successfully performed (4,32,33). Thus, different CDT error profiles are observed in injuries of various brain regions and/or diseases associated with these regions. Studies have reported that visual-spatial errors are related to right hemisphere injury, and clock representation errors are due to left hemisphere injury (33,34). On the basis of this information, the effects of demographic variables on the total score as well as on the number placement and clock representation sub-score types of the Manos-Wu scoring method were examined. In this study, there were age and education level effects on Manos-Wu total scores, whereas education level showed an effect only on number placement scores, and age had an effect only on clock representation scores. It was found that total score and clock representation scores decreased

Table 7. Test-retest agreement coefficients of the two methods

	Agreement	95% Confidence interval
Manos and Wu (5)		
Number placement	0.950***	0.914-0.971
Time presentation	0.330*	0.063-0.558
Total score	0.806***	0.681-0.885
Shulman (4)		
Total score	0.720**	0.554-0.831

Interclass correlation coefficient: *Weak, **Good, ***Excellent

Table 8. Inter-rater agreement coefficients of two methods

Scoring methods	Groups	Reliability coefficient	
Manos and Wu (5)		**Agreement**	**95% Confidence interval**
Number placement			
	A	0.979	0.968-0.986
	B	0.892	0.797-0.944
	C	0.958	0.913-0.979
	D	0.983	0.966-0.992
Time presentation			
	A	0.931	0.899-0.953
	B	0.460	0.165-0.683
	C	0.965	0.930-0.982
	D	0.910	0.825-0.955
Total score			
	A	0.892	0.974-0.988
	B	0.825	0.683-0.907
	C	0.970	0.932-0.986
	D	0.984	0.968-0.992
Shulman (4)		**Agreement**	**95% Confidence interval**
Total score			
	A	0.957	0.922-0.975
	B	0.790	0.625-0.888
	C	0.950	0.897-0.976
	D	0.948	0.814-0.980

A: All participants (n=100), B: Healthy subjects (n=35), C: Mild cognitive impairment patients (n=33), D: Alzheimer patients (n=32)

Table 9. Correlation between clock drawing test and mini mental state examination scores

	All participants (n=484)	MCI patients (n=128)	Alzheimer patients (n=109)
	MMSE		
MW time presentation	0.619	0.283	0.535
MW number placement	0.774	0.644	0.687
MW total score	0.787	0.614	0.706
SH total score	0.760	0.519	0.705

MCI: Mild cognitive impairment, MW: Manos-Wu scoring method, SH: Shulman scoring method, MMSE: Mini mental state examination

Table 10. Demographic characteristics of paired groups for ROC analysis

	Healthy participants (n=125)	MCI patients (n=125)	p
Age	72.31±6.20	73.72±6.26	0.075
Education	10.66±4.74	10.09±4.93	0.347
	Healthy participants (n=108)	**Alzheimer patients (n=108)**	**p**
Age	70.39±8.61	70.57±8.76	0.876
Education	9.48±4.31	8.55±4.39	0.115
	MCI patients (n=97)	**Alzheimer patients (n=97)**	**p**
Age	73.36±6.46	72.51±7.17	0.384
Education	9.11±4.43	8.70±4.39	0.516

Data are presented as mean ± standard deviation. MCI: Mild cognitive impairment

with increasing age, whereas total scores and number placement scores increased with increasing education levels.

Age and education level effects were also determined on CDT performances scored using the Shulman method. No sex effect was observed on any score type. In other words, these findings indicate that increasing age in healthy individuals negatively affects abstract thinking and conceptualization skills more, and that lower education level negatively affects ranking and planning skills more.

The norm values for both scoring methods were based on age and education levels. In both scoring methods, the CDT scores of individuals aged 70 years and over were significantly lower than those of other age groups. Similar findings have been reported in previous studies, and the age effect has been reported to be particularly evident in the population aged over 70 years (13,14,15).

It has been reported in several studies that CDT scores were adversely affected by lower levels of education (35,36). In the present study, it was found that the CDT scores of individuals with 0-5 years' education were lower than those of high school and above educated individuals. In addition, the CDT total scores obtained by using the Manos-Wu method also differed between the junior high/high school and high school and above education groups. In the direction of these findings, it can be concluded that the Manos-Wu method is more sensitive than the Shulman method in terms of educational effect. This difference between the two methods is thought to be related to the scoring of the CDT over a wider range of parameters in the Manos-Wu method.

It is thought that the type of error in CDT (i.e., the placement of numbers, the representation of clock, the presence of perseverative errors) may be distinctive for dementia syndromes. Perseverative error types are frequently observed in Parkinson's disease dementia, whereas conceptual errors (errors related to clock representation) are reported in Alzheimer-type dementia (33,37). In one study, uneducated healthy individuals were also found to make conceptual errors similar to patients with AD (36). These findings are consistent with the fact that only an education effect was detected in the clock representation scores in our study.

In the present study, CDT performances of healthy subjects were examined to assess the test-retest reliability for both methods. The agreement between the first and second applications was found to be at a perfect level for the Manos-Wu method, and at a good level for the Shulman method. In our study, the test-retest reliability of the Shulman method was lower than that of the Manos-Wu method, which may be related to the method being semi-quantitative and involving subjective evaluation. Fuzikawa et al. (38) noted that scoring of good test performance with the Shulman method (decision to give a performance of 4 or 5 points) is more challenging than scoring poor performances and that it may decrease test-retest reliability. In the study of Manos and Wu (5), the test-retest reliability of CDT administered at two-day intervals was reported as 0.94. In our study, CDT was administered with an average of 12-month intervals and the test-retest reliability was found as 0.81. This difference is thought to be related to the longer time interval between the two applications in our study. All neuropsychological test findings (e.g., memory, attention, executive functions) of healthy individuals that were repeated after 12 months were consistent with age and education norms and did not differ significantly from initial evaluations. Nevertheless, in the interpretation of test-retest reliability it should be considered that long intervals might reduce reliability. However, the values reported in both studies for Manos and Wu (5) indicated very high reliability coefficients.

Table 11. The area under the ROC curve, standard error and 95% confidence intervals for two different clock drawing test scoring methods in discriminating between healthy subjects, MCI and AD patients

	Manos-Wu			Shulman		
	AUC	Standard error	95% Confidence interval	AUC	Standard error	95% Confidence interval
Healthy-MCI	0.83	0.03	0.773-0.871	0.74	0.03	0.680-0.792
Healthy-AD	0.90	0.02	0.853-0.937	0.85	0.02	0.799-0.897
MCI-AD	0.66	0.04	0.591-0.729	0.66	0.04	0.585-0.723

MCI: Mild cognitive impairment, AD: Alzheimer's disease, AUC: Area under the curve

Table 12. The cut-off scores and the corresponding sensitivity and specificity values of the clock drawing test scoring methods in discriminating between healthy subjects, MCI and AD patients

	Manos-Wu			Shulman		
	Cut-off score	Sensitivity (%)	Specificity (%)	Cut-off score	Sensitivity (%)	Specificity (%)
Healthy-MCI	9	80.8%	80.0%	4	61.6%	83.2%
Healthy-AD	9	89.8%	76.8%	4	81.5%	79.6%
MCI-AD	7	62.9%	61.9%	3	61.9%	59.8%

MCI: Mild cognitive impairment, AD: Alzheimer's disease

Inter-rater reliability was found to be excellent for both scoring methods. However, in the Manos-Wu method, the inter-rater reliability in the clock representation scores of healthy individuals was lower than all other types of points, while being acceptable. In this method, as well as representing the clock correctly, it is important to have the correct length of clock hands for the scoring. The short hand is not scored, even if it points to 11, if it is not drawn shorter than the long hand. During the evaluation of the test performance, it was sometimes difficult to distinguish the lengths of the hands. This is thought to affect the scoring when the clock representation is accurate but the hand lengths are not clear.

In our study, there was a strong correlation between CDT scores obtained by two different scoring methods and MMSE scores assessing general cognitive status. This finding is consistent with Shulman et al.'s (3) study and shows that the concurrent validity of scoring methods is high. In addition, the validity of the two scoring methods has also been proven by a high degree of agreement between the methods.

The CDT was first used to evaluate cases of hemispheric neglect (18,39,40), but many studies in the last 20 years have shown that the CDT may be a useful screening test for detecting dementia syndromes. In this study, both scoring methods were found to have a high power to differentiate patients with AD and MCI from healthy individuals. However, the sensitivity and specificity of the Manos-Wu method for distinguishing patients with MCI from healthy individuals were higher than those of Shulman's method. In both methods, reliable results were not obtained for the distinction between patients with MCI and AD. These findings indicate that the CDT can be used as a screening test for MCI and AD in clinical use but may not be able to distinguish MCI from AD alone. In the literature, it has been reported that the use of CDT in conjunction with verbal fluency tests in detecting dementia in elderly individuals was the most beneficial combination compared to other test combinations (4).

As a result, it is thought that an individual with a CDT performance below norm values might require more detailed tests to investigate the source of the cognitive impairment. Literature findings suggesting that CDT performance might be affected from the early stages of the disease on is supported by this study, even though this alone is not sufficient for AD diagnosis. In the longitudinal follow-up of neuropsychiatric diseases affecting cognitive functions, use of CDT along with other neuropsychological tests is recommended.

Conclusion

The presence of different methods of application and evaluation of CDT calls for separate norm values for each method. For clinical evaluation and research, the most frequently used CDT version in our country is used without norm values. In this study, CDT norm values according to age and level of education for the Turkish population were established for the most commonly used scoring methods in the international studies, the Manos-Wu and Shulman methods. In the validity and reliability analyses, the results obtained from both scoring methods were highly parallel and showed that both methods had high test-retest and inter-rater reliability as well as high concurrent and criterion validity.

With this study, Turkish norm values of CDT, which has been widely used to evaluate visual-spatial skills and executive functions in our country and worldwide for more than 30 years and which is a rapid screening test, easy to practice and score, also suitable for bedside examination, are presented for the use of relevant physicians and researchers.

Ethics

Ethics Committee Approval: The study protocol was approved by the Ethics Committee of Dokuz Eylul University on 13.07.2017 (reference number: 2017/18-01).

Informed Consent: Written informed consent was obtained from all participants from the previous research studies.

Peer-review: Externally and internally peer-reviewed.

Authorship Contributions

Concept: D.D.E.S., Design: D.D.E.S., G.G.Y., Data Collection or Processing: D.Y., Analysis or Interpretation: D.D.E.S., D.Y., Literature Search: D.D.E.S., D.Y., Writing: D.D.E.S., D.Y., G.G.Y.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Öktem Ö. Nöropsikolojik testler ve nöropsikolojik değerlendirme. Türk Psikoloji Dergisi 1994;9:33-44.
2. Strauss E, Sherman EM, Spreen O. A compendium of neuropsychological tests: Administration, norms, and commentary. Oxford University Press, USA, 2006.
3. Shulman KI, Shedletsky R, Silver IL. The challenge of time: clock-drawing and cognitive function in the elderly. Int J Geriatr Psychiatr 1986;1:135-140.
4. Shulman KI. Clock-drawing: is it the ideal cognitive screening test? Int J Geriatr Psychiatry 2000;15:548-561.
5. Manos PJ, Wu R. The ten point clock test: a quick screen and grading method for cognitive impairment in medical and surgical patients. Int J Psych Med 1994;24:229-244.
6. Watson YI, Arfken CL, Birge SJ. Clock completion: an objective screening test for dementia. J Am Geriatr Soc 1993;41:1235-1240.
7. Tuokko H, Hadjistavropoulos T, Miller JA, Beattie BL. The clock test: a sensitive measure to differentiate normal elderly from those with Alzheimer's disease. J Am Geriatr Soc 1992;40:579-584.
8. Goodglass H, Kaplan E, Barresi B. Boston diagnostic aphasia examination (BDAE). Philadelphia: Lea and Febiger, 1983.
9. Wolf-Klein GP, Silverstone FA, Levy AP, Brod MS. Screening for Alzheimer's disease by clock drawing. J Am Geriatr Soc 1989;37:730-736.
10. Sunderland T, Hill JL, Mellow AM, Lawlor BA, Gundersheimer J, Newhouse PA, Grafman JH. Clock drawing in Alzheimer's disease. A novel measure of dementia severity. J Am Geriatr Soc 1989;37:725-729.
11. Agrell B, Dehlin O. The clock-drawing test. Age and Ageing 1998;27:399-403.
12. Cangoz B, Karakoc E, Selekler K. The norm determination and validity-reliability studies of clock drawing test on Turkish adults and elderlys (ages 50 and over). Turk J Geriatr 2006;9:136-142.
13. Mazancova AF, Nikolai T, Stepankova H, Kopecek M, Bezdicek O. The Reliability of Clock Drawing Test Scoring Systems Modeled on the Normative Data in Healthy Aging and Nonamnestic Mild Cognitive Impairment. Assessment 2016;pii:1073191116632586.
14. Bozikas VP, Giazkoulidou A, Hatzigeorgiadou M, Karavatos A, Kosmidis MH. Do age and education contribute to performance on the clock drawing test? Normative data for the Greek population. J Clin Exp Neuropsychol 2008;30:199-203.
15. Sugawara N, Yasui-Furukori N, Umeda T, Sato Y, Kaneda A, Tsuchimine S, Saito M, Nakaji S, Kaneko S. Clock drawing performance in a community-

dwelling population: Normative data for Japanese subjects. Aging Ment Health 2010;14:587-592.
16. Hubbard EJ, Santini V, Blankevoort CG, Volkers KM, Barrup MS, Byerly L, Chaisson C, Jefferson AL, Kaplan E, Green RC, Stern RA. Clock drawing performance in cognitively normal elderly. Arch Clin Neuropsychol 2008;23:295-327.
17. Lourenço RA, Ribeiro-Filho ST, Moreira Ide F, Paradela EM, Miranda AS. The Clock Drawing Test: performance among elderly with low educational level. Rev Bras Psiquiatr 2008;30:309-315.
18. Mainland BJ, Shulman KI. Clock drawing test. Cognitive Screening Instruments. Springer London 2013:79-109.
19. Can SS, Özel Kızıl ET, Varlı M, Turan E, Atlı T. Demans hastalarında üç farklı saat çizme testinin Türkçe uyarlamalarının psikometrik özellikleri. Nöropsikiyatri Arşivi 2010;47:91-95.
20. 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;12:189-198.
21. Güngen C, Ertan T, Eker E, Yaşar R, Engin F. Reliability and validity of the standardized Mini Mental State Examination in the diagnosis of mild dementia in Turkish population. Turk Psikiyatri Derg 2002;13:273-281.
22. Öktem Ö. Sözel Bellek Süreçleri Testi. Bir ön çalışma. Nöropsikoloji Arşivi 1992;29:196-206.
23. Wechsler D. Wechsler memory scale-revised (WMS-R). Psychological Corporation 1987.
24. Wechsler D. Wechsler adult intelligence scale-Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson, 22, 2008;498
25. Tumaç A. Normal deneklerde frontal hasarlara duyarlı bazı testlerde performansa yaş ve eğitimin etkisi. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Psikoloji Bölümü, Yüksek Lisans Tezi, İstanbul, 1997.
26. Stroop JR. Studies of interference in serial verbal reactions. Journal of Experimental Psychology 1935;18:643.
27. Kaplan E, Goodglass H, Weintraub S. Boston naming test. Pro-ed, 2001.
28. Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1983;17:37-49.
29. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH. 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. Alzheimers Dement 2011;7:270-729.
30. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR Jr, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH. 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;7:263-269.
31. Teixeira Fabricio A, Aprahamian I, Sanches Yassuda M. Qualitative analysis of the Clock Drawing Test by educational level and cognitive profile. Arq Neuropsiquiatr 2014;72:289-295.
32. Freedman, M. Clock drawing: A neuropsychological analysis. Oxford University Press, USA. 1994.
33. Caffarra P, Gardini S, Zonato F, Concari L, Dieci F, Copelli S, Freedman M, Stracciari A, Venneri A. Italian norms for the Freedman version of the Clock Drawing Test. J Clin Exp Neuropsychol 2011;33:982-988.
34. Tranel D, Rudrauf D, Vianna EP, Damasio H. Does the Clock Drawing Test have focal neuroanatomical correlates? Neuropsychology 2008;22:553-562.
35. von Gunten A, Ostos-Wiechetek M, Brull J, Vaudaux-Pisquem I, Cattin S, Duc R. Clock-drawing test performance in the normal elderly and its dependence on age and education. Eur Neurol 2008;60:73-78.
36. Kim H, Chey J. Effects of education, literacy, and dementia on the Clock Drawing Test performance. J Int Neuropsychol Soc 2010;16:1138-1146.
37. Lee AY, Kim JS, Choi BH, Sohn EH. Characteristics of clock drawing test (CDT) errors by the dementia type: quantitative and qualitative analyses. Arch Gerontol Geriatr 2009;48:58-60.
38. Fuzikawa C, Lima-Costa MF, Uchoa E, Barreto SM, Shulman K; Bambuí Health and Ageing Study. A population based study on the intra and inter-rater reliability of the clock drawing test in Brazil: the Bambuí Health and Ageing Study. Int J Geriatr Psychiatry 2003;18:450-456.
39. Critchley M. The Parietal Lobes. Hafner: New York, NY, 1953.
40. Hazan E, Frankenburg F, Brenkel M, Shulman K. The test of time: a history of clock drawing. Int J Geriatr Psychiatry 2018;33:e22-e30.

Annex 1. Clock drawing model with a diameter of 10 cm that can be used for Manos and Wu (5) scoring method