Jurnal Psikologi
Volume 49, Number 2, 2022: 104– 122
DOI: 10.22146/jpsi.68953

ISSN 0215-8884 (print)
ISSN 2460 867X (Online)

Indonesian Trail Making Test: Analysis of Psychometric Properties, Effects of Demographic Variables, and Norms for Javanese Adults

Lucia Trisni Widhianingtanti*[1], Gilles van Luijtelaar[2], Angela Oktavia Suryani[3], Yohana Ratrin Hestyanti[3], Augustina - Sulastri[1]

[1]*Faculty of Psychology, Soegijapranata Catholic University, Indonesia* [2]*Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen Netherlands,*
[3]*Faculty of Psychology, Atma Jaya Catholic University of Indonesia*

Submission 12 September 2021 Accepted 20 July 2022 Published 26 August 2022

Abstract. Trail Making Test (TMT) has been widely used in Indonesia as an Executive Function (EF) test, however, no studies reported validity and reliability of the test. In this study, we analyzed TMT's psychometric properties, including reliability and validity, effect of demographic variables such as education, age, and sex on the TMT, and propose norm scores for the urban population of Java island. Four hundred ninety persons (aged 16-80 years) with varying education levels participated in the study. Four additional EF tests (Digit Span, Five Point, phonemic Verbal Fluency, and Stroop) were administered. Principal Component Analysis was used to test whether the structure of TMT supported theoretical foundations of EF. The test-retest reliability was estimated in different sample (N = 50). Results suggest that education increased and age decreased the performance, however there were no differences between high school graduates and undergraduates in the age range 16-39 years. The intraclass correlation as a reliability measure showed good results (TMTA, rs=0.76; TMT B, rs= 0.86; TMT B-A, rs=0.74). The PCA revealed that the TMT, Digit Span, and Five Point scores loaded highly on one construct, while Stroop, phonemic Verbal Fluency, and errors on Five Point test loaded highly on others. It can be concluded that TMT is a valid neuropsychological instrument measuring EF with high reliability score and has a high reliability and sensitivity for education, and age effects. The study also provides TMT norms and cut-off scores for Javanese population between 16 and 39 years old with senior high school and undergraduate level of education.

Keywords: executive function; javanese adults; psychometric properties; trail making test

The Trail Making Test (TMT), first developed in 1944 by United States Army psychologist, is a visual motor test part of the Army Individual Test Battery. It is a pen and paper test and consists of two parts, TMT-A and TMT-B, with various difficulties. In TMT-A, the instruction is to connect 25 randomly distributed numbers on a sheet of paper as quickly as possible by drawing lines between them in ascending order with a pencil (1, 2, 3,...). In TMT-B, the instruction is to connect 25 alternating numbers and letters presented on a sheet of paper in ascending and alphabetical order (i.e., 1–A–2–B) (Bezdicek et al., 2012). The participant is instructed to complete both parts as fast and accurately as possible without lifting the pencil from the paper (Bowie & Harvey, 2006; Varjacic et al., 2018). The

*Address for correspondence: trisni@unika.ac.id

score of these tests is computed based on the completion time (in seconds) of the TMT-A and the TMT-B (errors count only by increasing performance time) (Reitan & Wolfson, 1985, 2004).

The TMT is appealing to clinicians because it measures psychomotor speed (TMT-A), visual-motor skills, and its versatility of executive functions (TMT-B) aspects, among others planning, mental flexibility, inhibition/interference control, and attentional set-shifting (Arbuthnott & Frank, 2000; Bouattour et al., 2017; Bowie & Harvey, 2006; Fernandez & Marcopulos, 2008; Seo et al., 2006; Siciliano et al., 2018; Tombaugh, 2004). The difference in completion time between the TMT-B and TMT-A (TMT (B-A) is thought to partially account for the influence of baseline motor speed and is supposed to measure more basic cognitive abilities, particularly in the executive function domain. compared to the TMT-A. Therefore, the difference score based on completion time of both tests reflects higher-order executive demands placed on participants. The TMT is also popular due to its simplicity and versatility although it is not completely culture-free (Fernandez & Marcopulos, 2008; Ojeda et al., 2014). The TMT is also clinically used and sensitive to the presence of cognitive impairments in traumatic brain injury and stroke patients, also for patients suffering from dementia (Demakis, 2004; Periáñez et al., 2007) although poor performance on the test is relatively nonspecific (Lezak, 1995).

The TMT can be administered independently as a single instrument assessing performance in healthy subjects as part of an assessment for admission to higher education, of for job selection, but also clinically to quantify the impact of neurological diseases and neuropsychological functional impairments (Ashendorf et al., 2008; Bezdicek et al., 2012) or as a part of a larger test battery like the Halstead-Reitan Neuropsychological Test Battery (HNTB) (Jarros et al., 2017; O'Bryant et al., 2017; Rivera & Arango-Lasprilla, 2017). Interpretation of performances on the TMT relies on the presence of normative scores. Normative studies on TMT have been carried out across time and countries (Cavaco et al., 2013; Fernandez & Marcopulos, 2008; Seo et al., 2006; Siciliano et al., 2018; Tombaugh, 2004). These authors concluded that even the most elementary inspection of normative TMT data as collected in eight countries with a "Western" style of education affirm that norms are not interchangeable between countries and cultures. Tombaugh (2004) updated the normative scores of his 1998's version by providing a more comprehensive set of norms for the Canadian population. Changes in normative data across time were found in Italy, showing that updating normative scores is imperative because of changes in living conditions, longer life expectancies, and higher mean education levels (Siciliano et al., 2018). Therefore, normative scores must be based on recently collected data and are dependent on a country's composition of the population regarding age, level of education, and health conditions. Despite its widespread use throughout the world, no studies on the TMT normative scores were published for the Indonesian population, to the best of our knowledge. We initiated the collection of normative data from the TMT in Java, the most densely populated island in Indonesia, and we also evaluated the reliability of the TMT using the test-retest method in an independent study.

Next to the mean and *SD* of the TMT-A, TMT-B, and TMT (B-A), percentile scores (5 and 95%) will be presented, since they would often be more appropriate in the daily work of practitioners than presenting normative information using means score and standard deviations. It will also be explored whether age, years of education, and sex influence the TMT scores, and if this is the case, then the TMT

demands fine-tuned normative scores taking into account the influence of these demographic factors (Hester et al., 2005; Ivnik et al., 1996; Siciliano et al., 2018; Tombaugh, 2004). In the vast majority of the normative TMT studies, age, education, and other related factor such as intellectual-ability effects emerged, while sex and ethnicity effects were not always present (Abe et al., 2004; Fernandez & Marcopulos, 2008; Mitrushina et al., 1999). The latter might be particularly relevant for the large variety of ethnic groups in the Indonesian population.

Traditional methods for developing normative scores require large sample sizes because demographic variables are commonly divided into discrete categories for age and education, implying that each combination of age and education requires a sufficient sample size (minimally 30, more adequate is 50). Some advocate that a cell should contain even more subjects (Bridges & Holler, 2007). Smaller and modest sample sizes limit the precision of the normative scores (Crawford & Garthwaite, 2008).

The present study aims to report psychometric properties of the TMT for subjects living in urban areas of Java island, that is, normative scores on the whole sample, the test-retest reliability, a measure of the construct validity of the TMT as an executive function test, and to investigate if and how the demographic factors age, education, and sex affect the performance scores. If so, balanced normative scores will be presented for one or more subgroups with sufficient subjects.

Methods

Participants

Java island is the most densely populated island in Indonesia; about 57% of Indonesia's population, around 267 million, live on this island. Therefore we started our research over there. 492 healthy participants living in the urban areas of West, Central and East Java island, were recruited. This group, with 60% females, had a rather extensive age range (16 to 80 years, M= 33.2), and their education varied from elementary school to postgraduate. Culturally they represent the urban Javanese population. Participants were categorized into five age-by-decade groups (Williams et al., 2009) and two other categories outside those five: (I) 20-29 years old, (II) 30-39 years old, (III) 40-49 years old, (IV) 50-59 years old, (V) 60-69 years old, (VI) 15 - 19 years old and (VII) over 70 years old. The time of education varied from 0 - 22 years (M= 13,9; SD= 2,7) and was categorized into five groups corresponding to Indonesia's education system: (I) elementary school (ES), educated for less than seven years, (II) junior high school (JHS), educated between 7-9 years, (III) senior high school (SHS or equivalent), educated between 10-12 years, (IV) undergraduate and its equivalent (UG), educated between 13-16 years, and (V) postgraduate (PG), educated for more than 17 years.

The tests were administered in the Indonesian language. The test-assistants first explained the procedures to the participants, then informed them that the data would be used only for scientific purposes. Participants gave their consent by signing the informed consent. They received a financial reward of equal to 5 USD following completion of the series of tests. In this project, ten neuropsychological tests were adapted, the TMT was one of them. Only data from the healthy

participants (with no reported history of psychiatric or neurological diseases, head trauma, drug abuse, or other illnesses that could influence the performance on the tests) were included. The research was conducted in compliance with the Helsinki Declaration, and the ethics committee of Soegijapranata Catholic University gave clearance for this research project (University Ethical Clearance number: 001B/B.7.5/FP.KEP/IV/2018). The design of the database, the transport, and storage of private, sensitive information also fulfill Indonesia's regulations as mentioned in ITE (Information and Electronic Transaction).

Table 1 provides an overview of the demographics of the subjects. Most of the participants had an undergraduate or postgraduate education (58.8%), while almost one-third had completed SHS. A small percentage finished JHS (6.7%) or ES (3.1%).

Table 1
Demographic Data of the Normative Group

Variables	Category	Frequency (N = 490)	Percent
Education	Elementary School	15	3,1
	Junior High School	33	6,7
	Senior High School	154	31,4
	Undergraduate	267	54,5
	Postgraduate	21	4,3
Age	16-19	63	12,9
	20-29	203	41,4
	30-39	73	14,9
	40-49	55	11,2
	50-59	66	13,5
	60-69	22	4,5
	70-highest	8	1,6
Sex	Female	294	60,0
	Male	196	40,0

The test-retest reliability was determined in a different sample of 50 subjects recruited in Semarang. Twenty-four males and twenty six females; their mean age was 37.5 years (range 21-64) and mean years of education 16.7 (range 9-22). The test procedure was the same as part of a larger project described aforementioned, except that some other tests (e.g., Raven and MMSE/Mini Mental State Examination) were additionally administered to examine the general cognitive functions of the participants. The interval between the two sessions was one week. These participants also reported no history of psychiatric or neurological diseases or head trauma, based on a health questionnaire for factors and causes which might have influenced the test scores. They all agreed to complete the study and permitted that the data would be used for scientific research. All participants were tested in their home situation by trained test assistants.

Measures

All individuals were assessed on both parts of the TMT. Each part began with an exercise. In Part A, there were circles with numbers from 1 – 25. Participants should draw lines to connect the numbered

circles in ascending sequence without lifting their pen or pencil from the paper. In Part B, the circles were combined of the numbered (1 – 13) and written with letters (A – L). Same as in Part A, the participants were asked to draw lines to connect the circles in ascending order again, but now with the added task of alternating between numbers and letters (i.e., 1-A-2-B-3-C, etc.). If a mistake was made, the tester would point out this to the participant and ask to correct this. The tester recorded the time spent to complete the two parts.

Statistical Analysis

The mean, standard deviation, median, skewness, kurtosis, and percentile scores (5 and 95%) were calculated for the whole sample; they present preliminary normative data for the urban Javanese population. The intraclass coefficient (ICC) expressed the test-retest reliability, established in an independent sample of fifty subjects. The interpretation of this coefficient introduced by Koo and Li (2016) is generally employed by many studies to define the cutoff classification. Values less than 0.5 indicate poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability and values greater than 0.90 indicate excellent reliability.

Three-factor analyses of variance with age in seven categories, an education level (five levels, representing the Indonesian educational system), and sex, all as between-subjects factors followed by posthoc tests according to Bonferroni, were used to explore whether these factors significantly explained a sufficient amount of between group and combination of group variance. Partial eta squared was used as an index of effect size (Richardson, 2011). Values of η^2 of .0099, .0588, and .1379 were considered small, medial, and large respectively. Differences between groups and subgroups will have consequences for the more definite normative data.

Principal Component Analysis (PCA) was carried out on TMT and scores of four other executive function tests to test the structural validity of the TMT as an executive function test. Besides the data from the TMT (only TMT A and TMT B were used to avoid multicollinearity), data from the Digit Span Backward and Sequence, Five Point Test the number of correct unique figures and number or perseverance errors, phonemic Verbal Fluency (the letters K, T, and S were used), and Color Stroop test (Time cards 1,2,3, not the commonly used difference scores Time Card 3 minus time Card 2 for the same reason). All data were first z-transformed to facilitate comparisons between scores of the different tests. The number of components used in this study was defined using several criteria such as KMO and Bartlett (= 0.807), no multicollinearity among variables, Eigenvalues > 1., and factor loadings > .4.

As a final step, we present unidirectional cut-off scores to classify given scores as normal or abnormal. In this, we followed the procedure as described in Siciliano et al. (2018). Briefly, considering the data in our sample size of n=490, a score above 81 (5 % worse performance) was considered as the outer limit. Scores within the 67-95% range were considered borderline, that is, between 48 and 80 for the TMT A. The cut-off border was 190 for TMT B and 114 for TMT (B-A).

Results

Out of the 492 subjects, two were outliers; their scores on the TMT were extremely high (> 6 times the SD) and far away from what could be expected in a healthy population. Their scores were excluded from the analysis. Demographical data provide information that subjects from different locations (Jakarta, Semarang, and Surabaya in Java island) varied in terms of age and education. Most subjects had an undergraduate level of education, and the age category 20-29 contained the largest number of subjects. Both characteristics are rather typical for the population of three of the largest cities in Java Island. Meanwhile, the under-representation of the oldest group reflected the relatively low life expectancy in Indonesia. The youngest group (16-19 years) cannot have a complete undergraduate or postgraduate education. Therefore, some of the cells for age and education stratified groups cannot be filled. Many other cells do not contain the recommended number of subjects (30 or 50). Table 2 contains the descriptive of the three variables of the TMT for the *n*=490 sample.

Table 2
Descriptive Statistics of The Three TMT Variables of 490 Subjects

	Time TMT A	Time TMT B	Time TMT (B-A)
Mean	44.72	87.99	42,73
Std. Error of Mean	.84	2.28	1,88
Median	41.00	75.50	32,00
Mode	32	63	16,00
Std. Deviation	18.5	50.5	41,5
Skewness	1.585	2.847	2,98
Kurtosis	3.75	11.28	13,01
Minimum	9	20	51.00
Maximum	134	426	340.00
Percentiles 5 %	23	42	7
Percentile 67 %	48	88	45
Percentiles 95 % (outer limit)	81	191	114

Reliability

The test-retest analyses with ICC were estimated, and their 95% confidence intervals were calculated using SPSS statistical package version 23, based on a mean-rating (k = 2), absolute-agreement, 2-way mixed-effects model. The results revealed moderate and good reliability, namely 0.76 for the TMT A, 0.86 for the TMT B, and .74 for the TMT B-A. We also checked whether there were significant differences between the performances on the tests, and an expected practice effect was evident. The performance was better at the second administration for all three variables of the TMT. The Student- t-tests showed significant lower scores for the time to complete the TMT A (t=3.89, df 49, p<.001) and TMT B (t=3.72, df 49, p<.001), and a tendency for a decrease for the TMT (B-A) (p<.10).

The Effect of Education, Age, and Sex on TMT

The data stratified by five education and seven age groups are presented in Table 3. A three-factor analysis of variance was used to investigate the influence of the demographic variables. The outcomes of the ANOVA's showed significant education (see Table 4) and age effects (Table 5), implying that these factors affected the time to complete the TMT A, TMT B, and TMT B-A, while sex had only a minor effect (Table 6).

Table 3

Statistics (Mean and SD, Minimum and Maximum Values as Obtained For The Five Educations, Seven Age Groups and Two Sexes For The Time To Complete The TMT A, TMT B, And TMT (B-A)

Variables	Category	Frequency (N = 490)	TMT A			TMT B			TMT B – A		
			Min-Max	*M*	*SD*	Min-Max	*M*	*SD*	Min-Max	*M*	*SD*
Education	Elementary School	15	37-128	70.73	26.07	53-426	198.93	111.75	6-340	128.27	93.95
	Junior High School	33	31-134	64.36	26.06	62-327	152.39	76.15	-51-260	88.03	70.26
	Senior High School	154	9-125	44.87	17.18	31-300	90.64	43.55	-18-239	45.06	36.04
	Undergraduate	267	13-109	41.00	14.95	20-238	73.43	27.54	-21-186	31.84	23.30
	Postgraduate	21	25-85	41.48	14.72	38-127	73.29	21.83	9-71	31.81	18.08
Age	16-19	63	9-91	39.92	14.58	38-276	79.41	36.31	-14-241	39.44	34.37
	20-29	203	16-109	40.12	14.16	23-194	24.27	71.19	-21-124	30.84	20.03
	30-39	73	18-81	41.38	14.33	32-327	85.21	49.78	-14-260	42.23	43.75
	40-49	55	22-93	47.15	15.38	49-303	93.47	53.91	-3-239	46.35	48.70
	50-59	66	25-95	50.47	16.33	37-420	113.48	58.48	-2-340	62.41	51.54
	60-69	22	13-122	67.50	27.28	20-291	142.59	72.50	-18-210	72.36	57.19
	70-highest	8	59-134	103.00	27.05	67-426	209.25	111.07	-51-298	106.25	103.47
Sex	Female	294	9-134	46.31	18.81	31-426	92.36	55.76	-51-340	45.63	46.55
	Male	196	13-125	42.35	17.75	20-303	81.44	40.55	-18-227	38.39	32.24

Table 4

Mean and SD of the TMT Scores Per Education Category Outcomes of ANOVA, Effect Size and Post-Hoc Tests For Between Group Comparisons

Variable	Elementary School (ES)[a] (*N*=15)		Junior High-School (JHS)[b] (*N*=33)		Senior High-School (SHS)[c] (*N*=154)		Undergraduate (UG)[d] (*N*=267)		Postgraduate (PG)[e] (*N*=21)		*F*-value	[273?][2]
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
TMT A	70.73[cde]	26.07	64.36[cde]	26.06	44.87[ab]	17.18	41.00[ab]	14.95	41.48[ab]	14.72	23.18***	0.16
TMT B	198.93[bcde]	111.75	152.39[acde]	76.15	90.64[abd]	43.55	73.43[abc]	27.54	73.29[ab]	21.83	53.99***	0.31
TMT B-A	128.27[bcde]	93.95	88.03[acde]	70.26	45.06[abd]	36.04	31.84[abc]	23.30	31.81[ab]	18.08	40.81***	0.25

Note: superscript letter in row 3 – 5 = significantly different on post-hoc group comparisons.

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$.

Table 5

Effects of Age on TMT

Variable	16-19[a] (*N*=63)		20-29[b] (*N*=203)		30-39[c] (*N*=73)		40-49[d] (*N*=55)		50-59[e] (*N*=66)		60-69[f] (*N*=22)		70-highest[g] (*N*=8)		*F*-value	[273?]
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
TMT A	39.92[efg]	14.56	40.12[efg]	14.16	41.38[efg]	14.33	47.15[efg]	15.38	50.47[abcfg]	16.33	67.50[abcdeg]	27.28	103.00[abcdef]	27.05	32.14***	0.32
TMT B	79.41[efg]	36.31	71.19[defg]	24.27	85.21[efg]	49.78	93.47[befg]	53.91	113.48[abcfg]	58.48	142.59[abcdg]	72.50	209.25[abcdef]	111.07	24.47***	0.26
TMT B-A	39.44[efg]	34.37	30.84[defg]	20.03	42.23[efg]	43.77	46.35[bfg]	51.54	62.41[abcdg]	51.53	72.36[abcd]	57.19	106.25[abcde]	103.47	11.72***	0.15

Note: superscript letter in raw 3 – 5 = significantly different on group comparisons.

* = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$; post-hoc (Bonferroni) test $p< 0.05$

Table 6
Effects of Sex on TMT

Variable	Female (*N*=294)		Male (*N*=196)		*t*-value	Cohen's *d*
	M	*SD*	*M*	*SD*		
TMT A	46.31	18.81	42.35	17.75	2.33*	0.21
TMT B	92.36	55.76	81.44	40.55	2.36*	0.22
TMTB-A	45.63	46.55	38.39	32.24	1.90	0.18

* = $p < 0.05$; ** $p = < 0.01$; *** = $p < 0.001$

Education

Result suggests that higher educated people scored better than less long educated persons. This was clearly the case for all three TMT variables, although the size of the education effect, expressed by $\acute{\eta}^2$ and being large, was higher for the TMT B (0.31) and the TMT (B-A) (0.25) than for the TMT A (0.16). Post-hoc tests of the main effect for education showed that (a) there was no difference between the ES and JHS group on the TMT A, (b) ES and JHS tends to perform slower than SHS, UG, and PG, and (c) SHS, UG, and PG did not differ from each other. Level of education-related also associated with TMT B and the TMT (B-A). The same tendencies with previous findings were found: better-educated subjects were faster than the less well-educated groups. Also, the groups' differences were found for the TMTB and the TMT (B-A). The ES was the poorest and slower than any other groups, the JHS was slower than the three better-educated groups, and the SHS was slower than the UG. The UG and PG did not differ from each other.

Age

Older people were generally slower than the younger groups, and this can be appreciated in Table 5. The post-hoc tests following the age effects showed similar scores for the four youngest groups, a significant increase at 50-59 compared to the four younger groups, a further significant increase for the 60-69, and a significant increase above 70 years.

More age-related differences were found for the TMT B, including an earlier age effect that was found for the TMT (B-A). There were no differences between the three youngest groups, significantly higher scores for the 40-49 group compared to the 20-29, higher scores for the 50-59 compared to 16-39, and higher scores for 60-69 compared to 16-49. The oldest group had higher scores than all other groups. The age-related increase of the TMT (B-A) showed an almost similar age effect as the TMT B: that is no differences between the three youngest group and an age-related increase in time for the 50-59 compared to the three youngest groups, higher scores for the 60-69 compared to the four youngest groups and again higher scores for the > 70 group compared to the five youngest groups.

Sex

Significant sex effects were found in this sample for the TMT A and the TMT B. The results are presented in Table 6. The males were generally a bit faster than females on all three dependent variables. However, the sex effects were small, as indicated by Cohen's *d*. This finding suggests that the normative scores should be developed based on age and education only.

The Combination of Groups.

We tested the interaction effects between the level of education and age on the TMT measurements. The two-way ANOVA was performed and the result showed significant interaction effects on the TMT A (*F*

(19.490) = 1.671. $p < 0.05$. $\acute{\eta}^2 = 0.07$). This result suggests that the difference in the TMT A between the different educated groups varies with age, and from the description, it becomes clear that the education effect becomes larger when people are getting older. Interaction effects were also obtained for the TMT B (F (19. 490) = 4.451. $p < 0.001$. $\acute{\eta}^2 = 0.16$) and TMT (B–A) (F (19.490) = 4.570. $p < 0.001$. $\acute{\eta}^2 = 0.16$). A series of post-hoc tests to disentangle the interaction effects regarding the presence of age effects per level of education showed no significant differences between the three youngest groups (16-39) which had either SHS or UG as well levels of education on all three TMT scores. Next, the post-hoc tests per age group showed that there were no differences between the three 16-39 years groups having an SHS and UG education on all three TMT variables. Therefore, given that the number of subjects in this subgroup was large (N = 311), it was decided to combine the normative data for 16-39 years groups with an SHS and UG level of education. The results are presented in Table 7.

Table 7
Descriptive Statistics of The Three TMT Variables of 311 Subjects With SHS or UG Level of Education and Age Between 16 and 39 Years

	Time TMT A	Time TMT B	Time TMT (B-A)
Mean	39.57	72.15	32.054
Std. Error of Mean	.773	1.386	1.209
Median	37.00	68.00	29.00
Mode	32	63	16.00
Std. Deviation	13.628	24.440	21.326
Kurtosis	2.803	2.845	2.482
Std. Error of Kurtosis	.276	.276	.276
Range	100	171	149
Minimum	9	23	-21.00
Maximum	109	194	128
Percentiles 1 %	17.12	32.00	-13.640
Percentiles 5 %	22.00	40.00	4.00
Percentiles 10 %	24.00	44.00	10.00
Percentiles 33 %	32.00	60.00	21.00
Percentiles 50 %	37.00	68.00	29.00
Percentiles 67 %	43.00	78.00	37.04
Percentiles 90 %	57.00	104.00	58.00
Percentiles 95 %	66.40	120.40	74.00
Percentiles 99 %	86.52	163.32	105.800

PCA Analyses For The Validity of The TMT

The construct validity of the TMT for Indonesia was tested by applying Principal Component Analysis on the z-transformed TMT scores and scores of four additional executive function tasks that were administered to the same subjects: the Five Point Test (measuring cognitive flexibility, Stroop Test (attention, inhibition and sensitivity for interference, Phonemic Verbal Fluency Test (flexibility in word production), and Digit Span Backward and Sequence (working memory). The prerequisite tests showed that both the Kaiser-Meyer-Olkin (KMO) and Bartlett's test of sphericity fulfilled the demands if the TMT (B-A) and Stroop card 3-card 2 scores were excluded since it caused the correlation matrix to be no longer positive definite. Eigenvalues of > 1 and factor loadings > .4 were used to decide upon the number of factors. The main results are presented in Table 8.

A four-factor model was obtained, which explained 78.65 % of the total variance. The TMT-B and the TMT-A scores loaded -.809 and -746 respectively on the same factor, which was also characterized

by performance on two other executive function tests, Digit Span Backward and Forward and the Five Point Test. The combination of high factor loading of different executive function tests on a single construct demonstrates that the TMT A and the TMT B have convergent validity. The TMT did not load on the other constructs that represent and reflect the unique components such as Stroop Test, the Verbal Fluency Test, and the number or perseverance errors of the Five Point Test. Therefore, it can be concluded that the TMT also has discriminant validity. The data suggest that the cognitive factors as measured with the TMT-A and the TMT-B, measure shared and unique aspects of executive functions. The scores of the TMT-A and TMT-B can be interpreted and encompass motor and visual-spatial scanning speed, attentional set-shifting, and motor speed skills, while the first construct as found with the PCA, including the TMT scores, also incorporates working memory, and visual-spatial creativity or strategy use.

Table 8
The Result of PCA of Five Executive Function Tests (n=490)

Variables	Components			
	1	2	3	4
TMT Time B	-.809			
TMT Time A	-.746			
Digit Span Sequence	.632			
Digit Span Backward	.630			
Five Point Unique Number	.565			
Verbal Fluency score K		.862		
Verbal Fluency score T		.838		
Verbal Fluency score S		.816		
Stroop Card 2 Time			.902	
Stroop Card 3 Time			.795	
Stroop Card 1 time			.774	
Five Point Number of repetitions				.855

Note. All variables already standardized using *Z* Score Cut-off scores for the whole (n = 490) and subsample (n=311) of 16-39 year old subjects with SHS or UG level of education fot the three TMT variables. Note the better scores in this Table between the normal scores (below inner limit) and pathological scores (above outer limit) of the whole and subsample

Table 9
Score Comparisons Among TMT

	TMT A	TMT B	TMT (B-A)
Outer limit score complete sample	81	191	114
Threshold scores complete sample	80-48	190-88	113-45
Inner limit score complete sample	<48	<88	<45
Outer limit (n = 311)	67	121	74
Threshold scores (n=311)	43-66	78-120	73-37
Inner limit score (n = 311)	< 43	< 78	< 37

Note. Normal (below inner limit score) and pathological scores (> outer limit score) using percentile 95; the threshold is a score between 67 and 95 percentiles given for the whole sample, and the subsample of n=311.

Cut-off scores : Cut-off scores for the whole (n = 490) and subsample (n=311) of 16-39 year old subjects with SHS or UG level of education fot the three TMT variables. Note the better scores in this Table between the normal scores (below inner limit) and pathological scores (above outer limit) of the whole and subsample.

Discussion

Lack of adequate recently collected normative data is a frequent and large reported concern of neuropsychologists (Rabin et al., 2016), certainly in Indonesia. We started with collecting normative scores in urban areas of Java for some neuropsychological tests, here we report psychometric data on the TMT. We reported firstly normative data for the whole sample. Considering that age and education effects were found between some subgroups and not between others, we also presented normative data on a single large subgroup. Next, the test-retest reliability was established. The validity of the TMT as an executive function test was demonstrated by the outcomes of the Principal Component Analysis.

The obtained measures for the TMT-A (mean 44.72) fell in the middle of what is internationally reported (Fernandez & Marcopulos, 2008), scores as low as 19 for a group of 25-34 years USA male citizens, and as high as 88 for a 65-75 years UK sample. Our mean TMT-B score was 87.99, and the fastest was the 25-34 years US citizens (49.50). The slowest was an Italian group of 70–79-year-old persons, and their mean score was 336. Our 95% cut-off score for the TMT A and the TMT B are lower than reported by Siciliano et al. (2018). This might be due to the large range in their TMT A and TMT B scores: their highest score was 268 for the TMT A, while in our sample this was 134, for the TMT B this was 512, and in ours, it was 412. Also, the mean scores were lower in our sample compared to the Italian sample. A detailed comparison with normative data from other countries is less meaningful considering the existence of sample bias, which occurs when samples are not comparable regarding demographic factors and demographics have a significant effect on the test outcomes and that the composition of the samples is not identical between the two studies. Another reason why comparisons with scores from other countries or cohorts are troublesome is that there is also administration bias, implying that different administration systems were followed: in one system, the subject is stopped as he/she makes mistakes and is asked to correct them, while other procedures allow the subject to perform with errors. Moreover, and most importantly, each country and culture deserve recently collected normative scores collected in a strict, standardized protocol, representing the averaged healthy population.

In the past, mean, dispersion and absolute cut-off scores of the TMT based on all probands were initially used to identify organic impairment (Matarazzo et al., 1974). However, this practice is no longer in vogue since it was demonstrated that age, level of education, and intelligence affected the TMT's performance and some other neuropsychological tests (Strauss et al., 2006). Therefore, we established the most relevant demographic factors (education, age, and sex) on the TMT for the Indonesian sample. Age, in general, decreased the performance in our sample with the notion that the age-dependent decline was most prominent for the TMT-A from 60 years onwards and on the TMT-B and the TMT (B-A) from 50 onwards. More years of education increased the performance; this was most prevalent for the TMT-B and the TMT (B-A), and to a lesser degree, on the TMT-A. Especially persons with ES and JHS performed relatively poorly. This result also seems in agreement with the notion that the performance on the TMT score correlates high and positive measures of intelligence (Bowie & Harvey, 2006), considering that intelligent people are often better educated. Different normative data for older people are essential since their increased chances of cognitive decline and dementia. The relatively large age effects on the TMT-A and the TMT-B and significant education effects on the TMT-B and the TMT (B-A) are in good agreement with the international literature (Fernandez & Marcopulos, 2008; Siciliano et al., 2018; Strauss et al., 2006) since we confirmed the age and education effects. Also, the small sex effect was more often found. Sex effects and their interactions were minor and were therefore ignored by us; however, stronger sex effects and interactions with sex

might emerge in different cultures or ethnic groups and remain an everlasting topic of interest.

The implications of the significant effects of education and age as well as their interactions are twofold: First, it is necessary to create and adapt normative data for different age and education groups. However, as revealed by post-hoc tests, not all age and education categories and combinations of groups differed from each other. The post-hoc tests showed no differences between the three youngest groups with SHS or UG level of education for all three dependent variables, and therefore, the normative data of these six groups were combined. The other groups did not have enough subjects for reliable normative scores as yet.

Another psychometric property of the TMT, such as the test-retest reliability, was additionally determined. From the literature, it can be inferred that the TMT has an adequate test-retest reliability for both Part A and Part B in a healthy control group (r=0.46 and 0.44, respectively). An excellent and adequate test-retest reliability was found for Part A and Part B of the TMT, respectively (r=0.78 and 0.67) among participants with diffuse cerebrovascular disease. The test-retest reliability of the TMT as a part of the Halstead-Reitan battery in a sample of 150 neuropsychiatric patients, including patients with stroke, showed excellent test-retest reliability for both Part A and Part B were found (0.94 and 0.86 respectively) in the sub-group of patients with stroke; and adequate reliability for the entire participant sample (0.69 and 0.66 respectively. Siciliano's group showed .80, .81 and .70 for the TMT-B, TMT-A, and TMT (B-A). Our present outcomes showed similarly acceptable to good reliabilities for the TMT-A and the TMT (B-A), excellent for the TMT B. Also, the practice effect found by us for the TMT-A and the TMT B and the smaller effect for the TMT (B-A) was earlier reported (Siciliano et al., 2018).

The validity of the TMT as an executive function test for Indonesians showed its strong association with some other executive function tasks, not with all others. This is in line with the general accepted idea that executive functions are an umbrella term covering different cognitive abilities and that executive function does not refer to a single, undifferentiated cognitive process. Neuropsychological assessment studies both in healthy subjects as well as in brain-injured people show the presence of multiple distinct factors in scores derived from batteries of executive function tests (Gilbert & Burgess, 2008). We identified four constructs with high factors loadings in our small battery of only five executive functions tests. Meanwhile, Miyake et al. (2000) found only three factor loadings for their executive function test. Next, the two subscales of the TMT loaded on the same construct as two other executive function tests and not on the two others. This result demonstrates that the TMT-A and the TMT-B have construct validity, including convergent and discriminant validity (Mitrushina et al., 1999). Together with the sensitivity of the TMT for the demographic factors education and age, the satisfying test-reliability scores, their sensitivity for repeated testing, and the outcomes of the PCA analyses, it is concluded that the TMT is a valid, reliable, and sensitive instrument for the assessment of performance in the executive function domain for Indonesians.

Normative Scores

Although data were collected from almost 500 subjects and normative data for this sample are now available, the ANOVA showed age, education, and interaction effects, making it necessary to develop normative data for subgroups. However, our post-hoc analyses showed that not all combinations of the five chosen education and seven age categories need their normative scores. More specifically, the combined 16–39-year-old subjects with SHS and UG showed no differences on TMT A, TMT B, and TMT (B-A), and therefore we propose similar normative scores. This group is relatively large (> 300), and they represent the urban Javanese until 40 years old population. The obtained scores of all TMT variables in this subsample were better than the whole sample, considering that older and less

well-educated groups did not fall in this category. Most of our other cells did not have a sufficient number of subjects for solid normative scores. Unfortunately, normative data stratifying by relevant demographics needs quite a number of subjects per cell and insufficient access to resources needed to collect in a standardized way data from an even much larger sample is a major challenge (Fellows & Schmitter-Edgecombe, 2019) and certainly in middle-income countries, including Indonesia. We have stratified our current sample to West, Central, and East Java, considering that 40% of the Indonesian population lives in Java; however, many combinations of age and education level did not fulfill the n=50 sample size demand and await more subjects (Oosterhuis et al., 2015). In the future, regression-based normative data may be obtained from relatively small groups of healthy controls in case the sample is representative, including sufficient less educated, older and people living in rural areas (Berrigan et al., 2014; Burggraaff et al., 2017; Parmenter et al., 2009) and also on other islands than Java.

The cut-off scores for the whole sample and the subgroup of 311 subjects are based on the more commonly chosen 95% criteria, and from Table 9, the better performances and the lower cut-off scores of this group can be appreciated. The inner threshold represents the 67-percentile score.

It is well accepted that culture exerts a strong influence on cognition since cognitive functions can develop in culturally distinctive ways (Henrich et al., 2010). Next, ethnicity and culture-bias exist in (neuro)psychological or cognitive assessments (van de Vijver & Tanzer, 2004), and this can lead to over-or under-diagnosing cognitive impairment for specific groups (Wong et al., 2000). A comparison between normative scores of several different countries and cultures demonstrated large differences (Fernandez & Marcopulos, 2008). Therefore, it is necessary to collect data in the very near future from other parts of Indonesia and establish whether cultural differences exist.

Conclusions

It can be concluded that the TMT is a valid and reliable test that can be used as an EF test measuring planning, mental flexibility, inhibition/interference control, and attentional set-shifting as well as psychomotor speed and visual motor skills. The TMT measures shared and unique aspects of the broad domain of EF. Normative scores for 16-39 year persons with senior high school and undergraduate level of education living in Java are presented, as well as cutoff scores for presumed pathology for the same education and age group. This implies that the TMT with its present normative scores for this group can be safely used. From the current results and in particular the differences in TMT scores between the n=311 and the whole sample of n=490, the outcomes of the various post-hoc tests, and the international literature, it is more than likely that older Indonesian persons or with a different type of education, need different normative scores. The normative scores for other age and education groups can be based on a powerful multifactorial curvilinear regression approach.

Recommendations

In case cultural effects will be found, this may also lead to further adaptation and or extension of the presently proposed normative scores, which are based on the urban Javanese population.

Declarations

Acknowledgements

The authors express the contribution of the test-assistants, who collected the data, and the other members of the Indonesian Consortium of Neuropsychologists who were instrumental for the present work.

Funding

The work presented here is supported by DIKTI grant 010/I.6/AK/SP2H.1/PENELITIAN/2019.

Authors' contributions

LTW conceived the study, wrote the manuscript, organized the data collection, and analyzed the data. GvL conceived the study, wrote the manuscript, and analyzed the data. AOS organized the data collection and reviewed the manucript draft and endorsed the final version of the manuscript. YRH reviewed the manuscript draft and endorsed the final version of the manuscript. AS supervised the statistical analysis process, endorsed the result, and reviewed the manuscript draft and endorsed the final version of the manuscript.

Competing Interest

The authors declare not to have any competing interests related to this work.

Orcid ID

Lucia Trisni Widhianingtanti https://orcid.org/0000-0002-1706-0662
Gilles van Luijtelaar https://orcid.org/0000-0002-0710-3403
Augustina Sulastri https://orcid.org/0000-0002-0107-7590
Yohana Ratrin Hestyanti https://orcid.org/0000-0001-9091-7944
Angela Oktavia Suryani https://orcid.org/0000-0001-5016-4802

References

Abe, M., Suzuki, K., Okada, K., Miura, R., Fujii, T., Etsurou, M., & Yamadori, A. (2004). Normative data on tests for frontal lobe functions: Trail making test, verbal fluency, wisconsin card sorting test (keio version). *Rain and Nerve*, *56*(7), 567–574.

Arbuthnott, K., & Frank, J. (2000). Trail making test, part b as a measure of executive control: Validation using a set-switching paradigm. *Journal of Clinical and Experimental Neuropsychology*, *22*(4), 518–528.

Ashendorf, L., Jefferson, A. L., Connor, M. K., Chaisson, C., Green, R. C., & Stern, R. A. (2008). Trail making test errors in normal aging, mild cognitive impairment, and dementia. *Archives of Clinical Neuropsychology*, *23*, 129–137.

Berrigan, L. I., Fisk, J. D., Walker, L. A. S., Wojtowicz, M., Rees, L. M., Freedman, M. S., & Marrie, R. A. (2014). Reliability of regression-based normative data for the oral symbol digit modalities test: An evaluation of demographic influences, construct validity, and impairment classification rates in multiple sclerosis samples. *The Clinical Neuropsychologist*, *28*(2), 281–299. https://doi.org/10.1080/13854046.2013.871337

Bezdicek, O., Motak, L., Axelrod, B. N., Preiss, M., Nikolai, T., Vyhnalek, M., Poreh, A., & Ruzicka, E. (2012). Czech version of the trail making test: Normative data and clinical utility. *Archives of Clinical Neuropsychology*, *27*(8), 906–914. https://doi.org/10.1093/arclin/acs084

Bouattour, N., Farhat, N., Hadjkacem, H., Hdiji, O., Dammak, M., & Mhiri, C. (2017). Trail making test (TMT): Tunisian normative values from 339 normal adult controls. *Journal of the Neurological Sciences*, *381*, 846. https://doi.org/10.1016/j.jns.2017.08.2381

Bowie, C. R., & Harvey, P. D. (2006). Administration and interpretation of the trail making test. *Nature Protocols*, *1*(5), 2277–2281. https://doi.org/10.1038/nprot.2006.390

Bridges, A. J., & Holler, K. A. (2007). How many is enough? determining optimal sample sizes for normative studies in pediatric neuropsychology. *Child Neuropsychology*, *13*(6), 528–538. https://doi.org/10.1080/09297040701233875

Burggraaff, J., Knol, D. L., & Uitdehaag, B. M. (2017). Regression-based norms for the symbol digit modalities test in the dutch population: Improving detection of cognitive impairment in multiple sclerosis. *European Neurology*, *77*(5-6), 246–252. https://doi.org/10.1159/000464405

Cavaco, S., Gonçalves, A., Pinto, C., Almeida, E., Gomes, F., Moreira, I., Fernandes, J., & Teixeira-Pinto, A. (2013). Trail Making Test: Regression-based norms for the portuguese population. *Archives of Clinical Neuropsychology*, *28*(2), 189–198. https://doi.org/10.1093/arclin/acs115

Crawford, J. R., & Garthwaite, P. H. (2008). On the optimal size for normative samples in neuropsychology: Capturing the uncertainty when normative data are used to quantify the standing of a neuropsychological test score. *Child Neuropsychology*, *14*(2), 99–117. https://doi.org/10.1080/09297040801894709

Demakis, G. J. (2004). Frontal lobe damage and tests of executive processing: A meta-analysis of the category test, stroop test, and trail-making test. *Journal of Clinical and Experimental Neuropsychology*, *26*(3), 441–450. https://doi.org/10.1080/13803390490510149

Fellows, R. P., & Schmitter-Edgecombe, M. (2019). Symbol digit modalities test: Regression-based normative data and clinical utility. *Archives of Clinical Neuropsychology*, *35*(1), 105–115. https://doi.org/10.1093/arclin/acz020

Fernandez, A. L., & Marcopulos, B. A. (2008). A comparison of normative data for the trail making test from several countries: Equivalence of norms and considerations for interpretation. *Scandinavian Journal of Psychology*, *49*(3), 239–246. https://doi.org/10.1111/j.1467-9450.2008.00637.x

Gilbert, S. J., & Burgess, P. W. (2008). Executive function. *Current Biology*, *18*(3), R110–R114. https://doi.org/10.1016/j.cub.2007.12.014

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83. https://doi.org/10.1017/s0140525x0999152x

Hester, R. L., Kinsella, G. J., Ong, B., & McGregor, J. (2005). Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults. *The Clinical Neuropsychologist*, *19*(1), 45–54. https://doi.org/10.1080/13854040490524137

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., & Petersen, R. C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist*, *10*(3), 262–278. https://doi.org/10.1080/13854049608406689

Jarros, R. B., Salum, G. A., da Silva, C. T. B., Toazza, R., Becker, N., Agranonik, M., de Salles, J. F., & Manfro, G. G. (2017). Attention, memory, visuoconstructive, and executive task performance

in adolescents with anxiety disorders: A case-control community study. *Trends in Psychiatry and Psychotherapy*, *39*(1), 5–11. https://doi.org/10.1590/2237-6089-2016-0032

Koo, T. K., & Li, M. Y. (2016). A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Lezak, M. D. (1995). *Neuropsychological assessment*. Oxford University Press.

Matarazzo, J., arthur n. Wiens, ruth g. Matarazzo, & steven g. Goldstein. (1974). Psychometric and clinical test-retest reliability of the halstead impairment index in a sample of healthy, young, normal men. *The Journal of Nervous and Mental Disease*, *158*(1), 37–49. https://doi.org/10.1097/00005053-197401000-00006

Mitrushina, M. N., Boone, K. L., & D'Elia, L. (1999). *Handbook of normative data for neuropsychological assessment*. Oxford University Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

O'Bryant, S. E., Edwards, M., Johnson, L., Hall, J., Gamboa, A., & O'jile, J. (2017). Texas mexican american adult normative studies: Normative data for commonly used clinical neuropsychological measures for english- and spanish-speakers. *Developmental Neuropsychology*, *43*(1), 1–26. https://doi.org/10.1080/87565641.2017.1401628

Ojeda, N., Aretouli, E., Peña, J., & Schretlen, D. J. (2014). Age differences in cognitive performance: A study of cultural differences in historical context. *Journal of Neuropsychology*, *10*(1), 104–115. https://doi.org/10.1111/jnp.12059

Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, K. (2015). Sample size requirements for traditional and regression-based norms. *Assessment*, *23*(2), 191–202. https://doi.org/10.1177/1073191115580638

Parmenter, B. A., Testa, S. M., Schretlen, D. J., Weinstock-Guttman, B., & Benedict, R. H. B. (2009). The utility of regression-based norms in interpreting the minimal assessment of cognitive function in multiple sclerosis (MACFIMS). *Journal of the International Neuropsychological Society*, *16*(1), 6–16. https://doi.org/10.1017/s1355617709990750

Periáñez, J. A., Ríos-Lago, M., Rodríguez-Sánchez, J. M., Adrover-Roig, D., Sánchez-Cubillo, I., Crespo-Facorro, B., Quemada, J. I., & Barceló, F. (2007). Trail making test in traumatic brain injury, schizophrenia, and normal ageing: Sample comparisons and normative data. *Archives of Clinical Neuropsychology*, *22*(4), 433–447. https://doi.org/10.1016/j.acn.2007.01.022

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-Year period: A Follow-up survey of INS and NAN members. *Archives of Clinical Neuropsychology*, *31*(3), 206–230. https://doi.org/10.1093/arclin/acw007

Reitan, R. M., & Wolfson, D. (1985). *The halstead-reitan neuropsychological test battery: Theory and clinical interpretation* (Vol. 4). Reitan Neuropsychology.

Reitan, R. M., & Wolfson, D. (2004). The trail making test as an initial screening procedure for neuropsychological impairment in older children. *Archives of Clinical Neuropsychology*, *19*(2), 281–288. https://doi.org/10.1016/s0887-6177(03)00042-8

Richardson, J. T. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review*, *6*(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Rivera, D., & Arango-Lasprilla, J. (2017). Methodology for the development of normative data for spanish-speaking pediatric populations (J. C. Arango-Lasprilla & D. Rivera, Eds.). *NeuroRehabilitation*, *41*(3), 581–592. https://doi.org/10.3233/nre-172275

Seo, E. H., Lee, D. Y., Kim, K. W., Lee, J. H., Jhoo, J. H., Youn, J. C., Choo, I. H., Ha, J., & Woo, J. I. (2006). A normative study of the trail making test in korean elders. *International Journal of Geriatric Psychiatry*, *21*(9), 844–852. https://doi.org/10.1002/gps.1570

Siciliano, M., Chiorri, C., Battini, V., Sant'Elia, V., Altieri, M., Trojano, L., & Santangelo, G. (2018). Regression-based normative data and equivalent scores for Trail Making Test (TMT): An Updated Italian normative study. *Neurological Sciences*, *40*(3), 469–477. https://doi.org/10.1007/s10072-018-3673-y

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary, 3rd edn*. Oxford University Press.

Tombaugh, T. (2004). Trail making test a and b: Normative data stratified by age and education. *Archives of Clinical Neuropsychology*, *19*(2), 203–214. https://doi.org/10.1016/s0887-6177(03)00039-8

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004

Varjacic, A., Mantini, D., Demeyere, N., & Gillebert, C. R. (2018). Neural signatures of Trail Making Test performance: Evidence from lesion-mapping and neuroimaging studies. *Neuropsychologia*, *115*, 78–87. https://doi.org/10.1016/j.neuropsychologia.2018.03.031

Wong, T. M., Strickland, T. L., Fletcher-Janzen, E., Ardila, A., & Reynolds, C. R. (2000). Theoretical and practical issues in the neuropsychological assessment and treatment of culturally dissimilar patients. In *Critical issues in neuropsychology* (pp. 3–18). Springer US. https://doi.org/10.1007/978-1-4615-4219-3_1