Neurological Sciences (2019) 40:469–477
https://doi.org/10.1007/s10072-018-3673-y

ORIGINAL ARTICLE

Regression-based normative data and equivalent scores for Trail Making Test (TMT): an updated Italian normative study

Mattia Siciliano [1,2] · Carlo Chiorri [3] · Valeria Battini [3] · Valeria Sant'Elia [2] · Manuela Altieri [2] · Luigi Trojano [2,4] · Gabriella Santangelo [2]

Received: 31 July 2018 / Accepted: 29 November 2018 / Published online: 7 December 2018

Abstract

Objectives The Trail Making Test (TMT) is widely used to assess psychomotor speed and attentional set-shifting. Since the regression-based norms and equivalent scores (ESs) for the TMT Italian version trace back to more than 20 years ago, we aimed at providing updated normative data for basic (Part A and Part B) and derived (Score B-A and Score B/A) TMT scores collected in a larger sample with an extended age range.

Methods Three hundred fifty-five Italian volunteers stratified for sex (166 men), age decades (age range 20–90 years), and educational level (from primary school to university) completed the TMT and the Montreal Cognitive Assessment (MoCA).

Results Multiple linear regression analyses revealed that age and educational level significantly influenced performances on basic and derived TMT scores except for B/A, which was associated only with the educational level. From the derived linear equations, correction grids for basic and derived TMT raw scores were developed. Inferential cutoff scores, estimated using a non-parametric technique, and ES were computed. Basic and derived TMT scores showed a good test–retest reliability (all $r_s \geq 0.50$); Part B ($r_s = -0.48$, $p < 0.001$) and Score B-A ($r_s = -0.49$, $p < 0.001$) were moderately associated with MoCA total score.

Conclusions This study confirms the association of basic and derived TMT raw scores with sociodemographic variables and provides updated correction grids and ES for assessing the attentional/executive functions in clinical and research fields.

Keywords Trail Making Test · Norms · Executive functions · Attention · Dementia · Assessment

Introduction

Trail Making Test (TMT) was developed as a part of the "Army Individual Test Battery" to assess visual–motor and visual–conceptual abilities of soldiers belonging to the US Army in 1944, and during the following years, it was released for public use [1]. Since then, the TMT has become one of the most commonly used neuropsychological test [2, 3] and has been included in structured batteries such as the Halstead–Reitan battery [4].

The first version of the TMT was composed of two separate parts. In the first part (Part A), examinees are required to connect 25 numbered circles with direct line in ascending order (i.e., 1-2-3 and so on); time needed to complete Part A is used as a measure of psychomotor speed and visual search/attention skills [5, 6]. In the second part (Part B), examinees are required to connect numbered and lettered circles in alternated numerical and alphabetical order (i.e., one number and one letter, as in 1-A-2-B-3-C and so on); time needed to complete Part B is used as a measure of alternation/flexibility, inhibition/interference control, mental tracking, and attentional set-shifting [5–8].

In clinical practice, in addition to the completion times of Part A and Part B of the TMT (i.e., basic scores), two derived

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10072-018-3673-y) contains supplementary material, which is available to authorized users.

✉ Luigi Trojano
luigi.trojano@unicampania.it

1 Department of Medical, Surgical, Neurological, Metabolic and Aging Sciences - MRI Research Center SUN-FISM, University of Campania "Luigi Vanvitelli", Piazza Miraglia 2, 80138 Naples, Italy

2 Department of Psychology, University of Campania "Luigi Vanvitelli", Viale Ellittico 31, 81100 Caserta, Italy

3 Department of Educational Sciences, University of Genova, Genoa, Italy

4 ICS Maugeri, Scientific Institute of Telese, Via Bagni Vecchi 2, 82037 Telese, Italy

scores are often used for interpretive purposes: the B-A difference (or Score B-A) and the B/A ratio (or Score B/A). These derived scores are usually employed to remove the speed component from the test performance, thus providing a more refined measure of executive control [6, 7], strongly related to prefrontal activation [9, 10].

Different administration and scoring procedures of the TMT have been proposed [3, 11, 12]. Currently, the TMT scoring method is still being used in the form proposed by Reitan [12], which, in case of mistakes in following the sequence, requires the examiner to point out examinee's errors, and to ask for correcting them (thus increasing the time score). Several authors criticized this scoring method, which does not control for time spent for corrections of mistakes, threatening test reliability [2, 13].

It is well known that the performance on the TMT is age- [3, 12, 14–16] and education-sensitive [2, 3, 14, 15, 17]. The association with sex is still controversial, as some studies reported that men were faster than women [13, 18], whereas others reported opposite results [17], or no significant difference between sexes [14]. Beyond the sociodemographic characteristics, cultural variables (i.e., set of learned traditions and living styles shared by the members of a society) may also affect performance [19, 20]. The impact of all these variables on the TMT was addressed by means of normative studies specifically carried out in different countries, namely in Czech Republic [8], Turkey [13], Latin America [18], and Portugal [21].

Four independent normative studies on TMT have been published in Italy from 1996 to date [22–25]; the only study integrally adopting the equivalent score (ES) methodology along with regression-based method [26, 27] dated from more than 20 years ago [22]. ES methodology [27] represents a landmark for most Italian neuropsychologists [26], since it allows for standardizing scores accurately and for grading and comparing performance on different neuropsychological tests within and between individuals.

The need for updated normative scores is crucial in neuropsychological evaluation for clinical purposes [28], as changes in living conditions, longer life expectancy (https://www.istat.it/en/archive/203827), and higher mean educational level modified the cultural background of the Italian general population. The aim of our study was to update regression-based norms and ES [27] for the TMT administered in the original version (as in [22]) in a large sample of healthy individuals gathered from northern and southern Italy. In addition to normative data for Part A and Part B, information is provided for the difference (Score B-A) and ratio scores (Score B/A), which are increasingly used by clinicians to assess executive functions [15]. We also evaluated test–retest reliability and investigated the association of TMT scores with general cognitive abilities.

Methods

Participants

We recruited participants in all districts of Campania (a southern Italian region including 9.6% of the entire Italian population) and Liguria (a northern Italian region including 2.6% of the entire Italian population) (https://www.istat.it/en/archive/203827).

We planned to enroll at least 222 participants on the basis of a priori power analysis for multiple regression analysis, as in previous regression-based Italian normative studies (e.g., [28]). Power analysis was performed by G*Power 3.1 with the following parameters: probability level (α) 0.05, desired statistical power ($1-\beta$) 0.80, effect size (Cohen's f^2) 0.05, and number of linear predictors 3.

A semi-structured self-report questionnaire was used to collect information about participants' medical history. In order to select only "healthy" individuals, we excluded participants who reported history of central nervous system disease, such as brain injury or stroke, epilepsy, hospitalization or protracted treatment for psychiatric illness, and/or substance abuse. In order to avoid enrolment of "supernormal" individuals, we did not exclude individuals with highly prevalent chronic medical conditions (such as mild hypertension or well-compensated type II diabetes).

To screen for cognitive impairment, we included only individuals who achieved a normal score on the Montreal Cognitive Assessment (MoCA; age- and education-adjusted score higher than or equal to 15.5 points) [29].

Three hundred fifty-five Italian volunteers, distributed across age classes (age range 20–90 years), gender and educational levels (from primary school to university) took part in the study (Table 1). Mean age of the sample was 55.02 years (*SD* 18.25; range 20–90 years), and mean formal education was 11.94 years (*SD* 4.58; range 2–28 years).

Informed consent was obtained from all participants included in the study. The study was carried out in accordance with the Declaration of Helsinki on Ethical Principles for Medical Research.

Materials and procedure

After having completed the MoCA, all participants were individually assessed on the TMT.

TMT has two parts preceded by trial runs to ensure comprehension of test instructions. In Part A, the participant is required to connect 25 circled numbers (distributed in a standardized random layout) in the correct serial order (e.g., 1-2-3 and so on). In Part B, the participant is required to alternately connect circled numbers from 1 to 13 and circled letters from A to N (e.g., 1-A-2-B-3-C and so on). In the both parts of the

Table 1 Distribution of the normative sample according to age, education level, and gender

	Age (years)							
	20–29	30–39	40–49	50–59	60–69	70–79	80–89	Total
Education level (years)								
1–5								
Men	–	–	1	3	5	2	4	15
Women	–	–	4	5	4	7	9	29
6–8								
Men	4	9	7	8	7	6	2	43
Women	2	8	7	8	4	4	3	36
9–13								
Men	10	8	11	8	9	5	5	56
Women	5	6	8	13	15	8	6	61
> 13								
Men	6	10	9	6	9	7	5	52
Women	9	8	7	8	14	11	6	63
Total								
Men	20	27	28	25	30	20	16	166
Women	16	22	26	34	37	30	24	189

test, the participant is required to connect items as quickly as possible without lifting the pencil from the paper [2].

Following Reitan [12, 30] and Giovagnoli et al. [22], we used the original forms of Part A (numbers only) and Part B (numbers and letters) and recorded the time (in seconds) the participant needed to complete either part. No time limitation was imposed. It is important to remind that the administration and scoring procedures introduced by Reitan [12, 30] require the examiner to point out errors as they occur, so that the participant could always complete the test and scoring is based on time alone. Therefore, we computed four scores: (1) time needed for completing Part A, (2) time needed for completing Part B, (3) Score B-A, and (4) Score B/A.

Statistical analysis

The raw basic (i.e., Part A and Part B) and derived (Score B-A and Score B/A) scores achieved by participants on the TMT were analyzed by means of simultaneous multiple regression analyses to test the influence of age, educational level, and sex. Before running this analysis, the effects of age and education level (expressed as years of schooling) were explored after several mathematical transformations (e.g., quadratic, cubic, logarithmic, reciprocal) by means of individual regression analyses to determine which was the most effective in reducing residual variance of the TMT scores.

The variables that resulted significant at this stage were included in simultaneous regression analyses, in which the effect of each variable was tested partialling out the effect common with the other terms of the model. At this step, the Bonferroni correction for multiple comparisons was applied to reduce the inflation of the type I error, and variables were included into the model only when the significance level related to each of them was lower than or equal to 0.017.

Then, for basic and derived TMT scores, we evaluated the best fitting linear regression model that could be used to adjust raw scores according to the demographic variables. To this aim, we estimated a linear model that provided the expected score for a given subject on the basis of his/her demographic features. Taking this model as a baseline, we calculated from the raw scores an adjusted score, by adding or subtracting the contribution given by each significant concomitant variable in the final correction model. Following this approach, scores can be directly compared across subjects of different demographic classes. After correcting all the raw scores, we considered a non-parametric procedure to evaluate unidirectional tolerance limits that can classify a given score as normal or abnormal with a confidence level set at 95% [31].

According to the procedure described by Ackermann [32], we computed the outer and inner tolerance limits separately, with the scores falling between them defined as "borderline scores," because inferentially controlled judgment cannot be expressed. The outer tolerance limit is commonly recognized as the cutoff value and is defined as the score at which or above which the probability that an individual belongs to the normal population is less than 0.05 [32].

To allow the adjustment of the raw scores of newly tested individuals according to demographic variables, a correction grid was built for any combination of age level (by 10-year steps) and educational level (according to the Italian schooling system). Since the use of adjusted scores is more informative when it is standardized, we have converted adjusted scores into a five-point ordinal scale, thus obtaining ES with the following meaning: 0 = scores equal or higher than the outer tolerance limit (5%); 4 = scores lower than the median value of the whole sample; 1, 2, and 3 were obtained by dividing into three equal parts the area of distribution between 0 and 4 [33].

Table 2 Descriptive statistics for raw scores on Trail Making Test (TMT) and Montreal Cognitive Assessment (MoCA)

	M	*SD*	*Mdn*	$\text{Range}_{\text{(min–max)}}$
Trail Making Test				
Basic scores:				
Part A	45.64	34.13	35.00	12.00–268.00
Part B	110.14	82.50	82.00	24.00–512.00
Derived scores:				
Score B-A	64.96	59.32	46.00	5.00–435.00
Score B/A	2.52	0.97	2.31	0.86–6.65
Montreal Cognitive Assessment				
Raw total score	23.53	4.09	24.00	11.00–30.00
Adjusted total score	22.96	3.13	23.34	15.53–30.25
MoCA domains:				
Executive functions	2.78	1.20	3.00	0.00–4.00
Visuospatial abilities	2.90	1.03	3.00	0.00–4.00
Language	5.32	0.90	6.00	1.00–6.00
Memory	2.37	1.74	2.00	0.00–5.00
Attention, concentration, and working memory	5.34	0.96	6.00	1.00–6.00
Orientation	5.78	0.57	6.00	1.00–6.00

M mean, *SD* standard deviation, *Mdn* median, *MoCA* Montreal Cognitive Assessment

Spearman's correlation coefficient (r_s) was used to explore the (i) degree of association between basic and derived TMT scores; (ii) correlation between basic and derived TMT scores and total and sub-domain MoCA scores; (iii) test–retest reliability (assessed on 40 subjects 3 weeks after the first TMT administration). Test–retest reliability was investigated in a subsample of healthy individuals (10% of the overall sample) that underwent TMT at two time points 1 month apart. Effect size for the correlation coefficient was defined by the following criteria: $r_s < 0.10$, negligible; $0.10 \le r_s < 0.30$, weak; $0.30 \le r_s < 0.50$, moderate; $r_s \ge 0.50$, strong [34].

Finally, we conducted pairwise *t* tests to determine the "practice effect" between the two repetition trials of TMT, and used the Cohen's *d* statistic to estimate effect size (with values < 0.20 indicating small effects, values around 0.50 indicating moderate effects, and values about 0.80 indicating large effects [34]).

The Benjamini–Hochberg [35] procedure was used to control for false discovery rate at the 0.05 level.

Results

Descriptive statistics of raw scores achieved on TMT and MoCA scores are reported in Table 2 (mean and standard deviation of basic and derived TMT raw scores stratified by age and education ranges were reported in Supplementary Table 1).

The distribution of basic and derived TMT scores was skewed, with a longer right tail. The individual regression analyses showed that the square root of education (in years) and the logarithmic transformation of age [log10 (100 − age)] were the most effective in reducing residual variance for all measurements.

The simultaneous regression analyses showed that the influence of age and educational level was always statistically significant for all TMT scores but for B/A, which was only associated with the educational level. Further, the linear effect of the sex (with men outperforming women) was statistically significant for the Part B and Score B-A in the individual regression analyses, but not in the simultaneous regression analyses, when the Bonferroni's correction for multiple comparisons was used (Supplementary Table 2).

The summary of multiple regression analyses (Enter Method) considering the sociodemographic factors significantly associated with basic and derived TMT scores in the previous steps (i.e., simultaneous regression analyses) is reported in Supplementary Table 3.

On these premises, we specified a linear model to estimate the score expected for a given subject on the basis of his/her age and educational level (Supplementary Table 4).

Taking this model as a reference, we built the formulae for exact direct calculation of adjusted TMT scores (see footnotes of the Table 3). We computed the correction grid for any combination of age (by 10-year steps) and educational level (according to the Italian schooling system) to allow adjustment of raw scores of newly tested individuals (Table 3). For individuals with demographic characteristics not included in the correction grid, it is possible to use the formulae for exact direct calculation of adjusted basic and derived TMT scores shown in footnotes of Table 3, but this procedure should be considered with caution in clinical practice.

For a sample of 355 individuals and using a non-parametric procedure, outer and inner tolerance limits were defined by values corresponding to the 11th and 25th worst observations (Table 4). Adjusted TMT scores higher than or equal to outer tolerance limit (or cutoff point) can be considered abnormal, values lower than inner tolerance limit indicate a normal performance while intermediate scores indicate a “borderline” performance, which in our study was obtained by 3.94% of the sample.

The score interval corresponding to each ES, the density of observations and the cumulative frequency of each ES are shown in Table 5.

We found strong correlations between Part A and Part B, and between Part B and Score B-A, whereas there was a moderate correlation between Part A and Score B-A; Score B/A was weakly correlated with Part B and Part A and strongly correlated with Score B-A. Moreover, our results showed weak to moderate negative correlations between total and sub-domain MoCA scores (except for the orientation domain score) and basic and derived TMT scores (Table 6).

Finally, we found a high test–retest reliability for basic and derived TMT scores and a small-moderate practice effect for Part A and Part B, whereas Scores B-A and B/A did not change across examinations (Table 7).

Discussion

We provided here updated normative values for one of the most used attentional/executive tests, the TMT, in a wide sample of Italian healthy individuals, stratified by age and education. To date, four independent normative studies on TMT have been published in Italy [22–25], but only one of these [22] integrally used the ES methodology [27] shared by Italian clinical neuropsychologists engaged in clinical, rehabilitative, and forensic settings [26]. However, this study was published more than 20 years ago [22] and did not enroll subjects aged 79 years and over; moreover, it did not assess the

Table 3 Correction grid for raw basic and derived Trail Making Test (TMT) scores

Education (years)	Age (years)						
	20–29	30–39	40–49	50–59	60–69	70–79	80–89
Basic scores:							
Part A							
1–5	− 10*	− 15*	− 20	− 27	− 36	− 47	− 64
6–8	10	5	0	− 7	− 15	− 27	− 44
9–13	20	15	9	2	− 6	− 17	− 34
> 13	29	24	18	11	3	− 8	− 26
Part B							
1–5	− 56*	− 67*	− 81	− 96	− 116	− 142	− 183
6–8	13	1	− 12	− 27	− 47	− 74	− 114
9–13	45	34	21	5	− 15	− 41	− 82
> 13	75	64	51	35	15	− 11	− 52
Derived scores:							
Score B-A							
1–5	− 45*	− 51*	− 59	− 68	− 80	− 95	− 119
6–8	3	− 3	− 11	− 20	− 32	− 47	− 71
9–13	26	19	12	2	− 9	− 24	− 48
> 13	47	40	32	23	12	− 4	− 27
Score B/A							
1–5	− 0.70	− 0.70	− 0.70	− 0.70	− 0.70	− 0.70	− 0.70
6–8	− 0.24	− 0.24	− 0.24	− 0.24	− 0.24	− 0.24	− 0.24
9–13	− 0.02	− 0.02	− 0.02	− 0.02	− 0.02	− 0.02	− 0.02
> 13	0.18	0.18	0.18	0.18	0.18	0.18	0.18

Values marked by the asterisk (*) should be taken cautiously because they were obtained by extrapolation from the formulas reported below. Formulas for exact direct calculation of adjusted score: Part A = raw Part A score + $78.04x\ [\log_{10}(100 - \text{age}) - 1.61] + 14.18x\ (\sqrt{\text{education}} - 3.38)$; Part B = raw Part B score + $182.76x\ [\log_{10}(100 - \text{age}) - 1.61] + 48.16x\ (\sqrt{\text{education}} - 3.38)$; Score B-A = raw Score B-A + $106.63x\ [\log_{10}(100 - \text{age}) - 1.61] + 33.54x\ (\sqrt{\text{education}} - 3.38)$; Score B/A = raw Score B/A + $0.32x\ (\sqrt{\text{education}} - 3.38)$

Table 4 One-tailed non-parametric tolerance limits for the upper 5% (worse performance) of the adjusted times with 95% confidence. Individuals with scores higher than the outer limits are pathological; those with scores that are lower than the inner limits are normal

	Trail Making Test			
	Part A	Part B	Score B-A	Score B/A
Outer limit	≥ 127	≥ 294	≥ 163	≥ 4.96
Borderline scores	126–76	293–197	162–136	4.95–4.12
Inner limit	≤ 75	≤ 196	≤ 135	≤ 4.11

psychometric properties of the Score B/A, which, along with Score B-A, is increasingly used as a measure of executive functions [15].

Two subsequent studies [23, 25] provided the regression equations and the standard deviations (SDs) for the Part A, Part B, and Score B-A in order to convert TMT raw scores in Z-scores adjusted for the main sociodemographic variables (i.e., age, education, sex, and job). Conversely, Mondini et al. [24] used the percentile method to determine the cutoffs separately for individuals with low (< 8 years of schooling) versus high (> 8 years of schooling) educational level.

Table 5 Equivalent scores (ES) for adjusted basic and derived Trail Making Test (TMT) scores

ES	Interval	Cumulative frequency	Density
Basic scores:			
Part A			
0	≥ 127	11	11
1	126–68	38	27
2	67–54	95	57
3	53–42	178	83
4	< 42	355	177
Part B			
0	≥ 294	11	11
1	293–170	38	27
2	169–128	95	57
3	127–103	178	83
4	< 103	355	177
Derived scores:			
Score B-A			
0	≥ 163	11	11
1	162–109	38	27
2	108–77	95	57
3	76–59	178	83
4	< 59	355	177
Score B/A			
0	≥ 4.96	11	11
1	4.95–3.75	38	27
2	3.74–2.93	95	57
3	2.92–2.34	178	83
4	< 2.34	355	177

Our study showed that time of execution significantly increased with age and that education independently affected Part A and Part B, as well as the Score B-A. We found a marginal advantage of males for Part B and Score B-A, but this effect did not survive to more stringent statistical criteria (i.e., Bonferroni's method for multiple comparisons) used to select the predictors to be included in the final regression models.

The age-related decline in TMT performance was consistent with previous Italian [22–25] and non-Italian [8, 13–15, 17, 18, 21, 36, 37] normative studies. Similarly, the effect of education was in keeping with previous Italian normative studies [22–25] as well as with other international studies [8, 13, 14, 18, 21, 36]. The lack effect of sex was in agreement with several previous Italian and international normative studies [21–25], whereas others reported sex differences in TMT performances [13, 17, 18, 21, 22, 37]. These inconsistent results may be partially ascribed to uncontrolled differences across normative studies in educational background between males and females [28] and to culture-specific differences [20].

As expected, we found that the influence of sociodemographic factors varied considerably across the basic and derived scores of TMT [38–40]. Indeed, the effects of age and education were more marked for Score B-A than for Score B/A, which was less sensitive to demographic characteristics [38, 39]. This variability highlighted the need of interpreting TMT performances with reference to the subject's demographic characteristics.

The very strong correlation between Part B and Score B-A suggested that two scores assess very similar aspects in healthy individuals, as in the previous normative study [22], but it remains possible that Score B-A could provide useful information in patients with pathological brain conditions [36]. Moreover, our results showed a close association between Score B-A and high-order cognitive abilities (e.g., executive functions, language) as assessed by MoCA, as well as moderate associations of Part B and Score B-A with MoCA total score. These results seem to substantiate Reitan's claims [12] that TMT, and particularly its Part B, taps alertness and concentrated attention but also language, executive functions, and visuospatial abilities during its execution.

Although it has been suggested that the two derived scores of the TMT (Score B-A and Score B/A) could both represent a measure of executive functions independent from motor and perceptual confounding factors [7], our results might foster the use of Score B-A compared with Score B/A, as the latter was weakly correlated with the two basic TMT scores and with measures of global cognition. However, further research is needed to gain better insight into the diagnostic utility of the two derived TMT measures.

Table 6 Spearman's correlations (r_s) among basic and derived Trail Making Test (TMT) scores, and between basic and derived TMT scores and total and subdomains Montreal Cognitive Assessment (MoCA) scores

	Trail Making Test			
	Part A	Part B	Score B-A	Score B/A
Trail Making Test				
Basic scores:				
Part A	–	–	–	–
Part B	0.78*	–	–	–
Derived scores:				
Score B-A	0.53*	0.92*	–	–
Score B/A	− 0.24*	0.37*	0.63*	–
Montreal Cognitive Assessment				
Raw total score	− 0.41*	− 0.52*	− 0.53*	− 0.27*
MoCA domains:				
Executive functions	− 0.36*	− 0.44*	− 0.43*	− 0.21*
Visuospatial abilities	− 0.17*	− 0.29*	− 0.32*	− 0.21*
Language	− 0.36*	− 0.43*	− 0.40*	− 0.17*
Memory	− 0.27*	− 0.33*	− 0.32*	− 0.16*
Attention, concentration, and working memory	− 0.21*	− 0.32*	− 0.36*	− 0.22*
Orientation	− 0.00	− 0.03	− 0.05	− 0.02

*$p < 0.001$ according to Benjamini–Hochberg adjustment

A last remark is about the high test–retest reliability of basic and derived TMT scores, as in the previous Italian normative study [22]. This finding would suggest that the administration and scoring procedures proposed by Reitan [12] for TMT are reliable and are not as strongly sensitive to practice effect as other tests for executive functions, particularly if one considers Scores B-A and B/A. These observations foster the use of this test in longitudinal research designs and clinical follow-up examinations.

It is important to acknowledge that the present study has limitations. First, we enrolled participants with normal MoCA scores, but this does not imply "normal" cognitive functioning. Moreover, we did not exclude individuals affected by chronic vascular or metabolic illnesses, such as hypertension or type 2 diabetes, which are very frequent in the aged people. These choices might have contributed to age-related decrease in TMT performances. However, many Italian normative studies followed these methodological procedures (e.g., [41–43]), and we conformed to this practice, also for the sake of consistency among Italian normative data. Second, our power analysis required enrollment of at least 222 participants, and we planned to enroll at least 32 individuals for each decade between 20 and 89 years, with participants evenly distributed across gender and education levels. This goal was not fully achieved, and this might have inflated observed statistical power. However, the low proportion of young adults with low schooling was related to the current Italian legislation requiring a minimum of 8 years of education since 1962 (law 1859/62), and the low proportion of elderly Italian people with education level greater than 8 years is consistent with Italian statistical data (https://www.istat.it/en/archive/203827). The same difficulties have been

Table 7 Test-retest reliability and practice effect of basic and derived Trail Making Test (TMT) scores

Trail Making Test	Test	Retest	Test–retest reliability	Practice effect			
	M (*SD*)	*M* (*SD*)	r_s	*t* test	*p*	Adj-*p*	Cohen's *d*
Basic scores:							
Part A	26.00 (9.27)	22.93 (8.20)	0.80	2.69	0.018	0.036	0.35
Part B	74.50 (29.87)	56.21 (14.07)	0.81	2.86	0.013	0.036	0.56
Derived scores:							
Score B-A	49.41 (28.99)	40.00 (21.75)	0.70	1.48	0.158	0.113	0.34
Score B/A	3.10 (1.41)	2.60 (0.73)	0.74	1.86	0.085	0.158	0.36

Adj-*p* represents *p* value corrected for Benjamini–Hochberg procedure encountered in other Italian normative studies as well (e.g., [28, 44, 45]).

In conclusion, we collected an updated set of norms for TMT in healthy subjects from 20 to 90 years old, balanced for age, education, and sex. Basic and derived scores have been analyzed and cutoff values estimated. The TMT proved to be a reliable and valid instrument for measuring attentional/executive cognitive functioning, so the present normal dataset paves the way to further application of TMT in Italian experimental and research contexts. In particular, as TMT is broadly used in neuro-geriatric settings for clinical and research purposes, our normative study including a sample of very old individuals not well represented in previous studies [22] could promote application of TMT to elderly Italian population.

Acknowledgments The authors thank Dr. Paola Falanga, Dr. Ilaria Fornaro, Dr. Maria Rosaria Figliola, Dr. Domenico Vivarelli, Dr. Anita Sportiello, and UNITRE for their contribution in collecting data.

Compliance with ethical standards

The study was carried out in accordance with the Declaration of Helsinki on Ethical Principles for Medical Research.

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. US War Department AGsO (1944) The new Army Individual Test of General Mental Ability. Psychol Bull 41:532–538
2. Lezak MD (1995) Neuropsychological assessment, 3rd edn. Oxford University Press, New York
3. Spreen O, Strauss E (1998) A compendium of neuropsychological tests: administration, norms, and commentary, 3rd edn. Oxford University Press, New York
4. Reitan R, Wolfson D (1993) The Halstead-Reitan neuropsychological test battery: theory and clinical interpretation. Neuropsychology Press, Tucson
5. Crowe SF (1998) The differential contribution of mental tracking, cognitive flexibility, visual search, and motor speed to performance on parts A and B of the Trail Making Test. J Clin Psychol 54:585–591
6. Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, Rodríguez-Sánchez JM, Ríos-Lago M, Tirapu J, Barceló F (2009) Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. J Int Neuropsychol Soc 15:438–450. https://doi.org/10.1017/S1355617709090626
7. Arbuthnott K, Frank J (2000) Trail making test, part B as a measure of executive control: validation using a set-switching paradigm. J Clin Exp Neuropsychol 22:518–528
8. Bezdicek O, Motak L, Axelrod BN, Preiss M, Nikolai T, Vyhnalek M, Poreh A, Ruzicka E (2012) Czech version of the Trail Making Test: normative data and clinical utility. Arch Clin Neuropsychol 27:906–914. https://doi.org/10.1093/arclin/acs084
9. Moll J, de Oliveira-Souza R, Moll FT, Bramati IE, Andreiuolo PA (2002) The cerebral correlates of set-shifting: an fMRI study of the trail making test. Arq Neuropsiquiatr 60:900–905
10. Jacobson SC, Blanchard M, Connolly CC, Cannon M, Garavan H (2011) An fMRI investigation of a novel analogue to the Trail-Making Test. Brain Cogn 77:60–70. https://doi.org/10.1016/j.bandc.2011.06.001
11. Armitage SG (1946) An analysis of certain psychological tests used for the evaluation of brain damage. J Gerontol 60:30–34
12. Reitan RM (1958) Validity of the Trail Making Test as an indicator of organic brain damage. Percept Mot Skills 8:271–276
13. Cangoz B, Karakoc E, Selekler K (2009) Trail Making Test: normative data for Turkish elderly population by age, sex and education. J Neurol Sci 283:73–78. https://doi.org/10.1016/j.jns.2009.02.313
14. Tombaugh TN (2004) Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 19:203–214
15. Hester RL, Kinsella GJ, Ong B, McGregor J (2005) Demographic influences on baseline and derived scores from the trail making test in healthy older Australian adults. Clin Neuropsychol 19:45–54
16. Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ (2005) Mayo's older Americans normative studies: age- and IQ-adjusted norms for the Trail-Making Test, the Stroop Test, and MAE Controlled Oral Word Association Test. Clin Neuropsychol 19:329–377
17. Bornstein RA (1985) Normative data on selected neuropsychological measures from a nonclinical sample. J Clin Psychol 41:651–659
18. Arango-Lasprilla JC, Rivera D, Aguayo A, Rodríguez W, Garza MT, Saracho CP, Rodríguez-Agudelo Y, Aliaga A, Weiler G, Luna M, Longoni M, Ocampo-Barba N, Galarza-Del-Angel J, Panyavin I, Guerra A, Esenarro L, García de la Cadena P, Martínez C, Perrin PB (2015) Trail Making Test: normative data for the Latin American Spanish speaking adult population. NeuroRehabilitation 37:639–661. https://doi.org/10.3233/NRE-151284
19. Ardila A, Moreno S (2001) Neuropsychological test performance in Aruaco Indians: an exploratory study. J Int Neuropsychol Soc 7:510–515
20. Fernández AL, Marcopulos BA (2008) A comparison of normative data for the Trail Making Test from several countries: equivalence of norms and considerations for interpretation. Scand J Psychol 49:239–246. https://doi.org/10.1111/j.14679450.2008.00637.x
21. 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. Arch Clin Neuropsychol 28:189–198. https://doi.org/10.1093/arclin/acs115
22. Giovagnoli AR, Del Pesce M, Mascheroni S, Simoncelli M, Laiacona M, Capitani E (1996) Trail making test: normative values from 287 normal adult controls. Ital J Neurol Sci 17:305–309
23. Amodio P, Wenin H, Del Piccolo F, Mapelli D, Montagnese S, Pellegrini A, Musto C, Gatta A, Umiltà C (2002) Variability of trail making test, symbol digit test and line trait test in normal people. A normative study taking into account age-dependent decline and sociobiological variables. Aging Clin Exp Res 14:117–131
24. Mondini S, Mapelli D, Vestri A, Bisiacchi P (2003) Esame neuropsicologico breve [Brief neuropsychological exam]. Raffaello Cortina Editore, Milano
25. Amodio P, Campagna F, Olianas S, Iannizzi P, Mapelli D, Penzo M, Angeli P, Gatta A (2008) Detection of minimal hepatic encephalopathy: normalization and optimization of the Psychometric Hepatic Encephalopathy Score. A neuropsychological and quantified EEG study. J Hepatol 49:346–353. https://doi.org/10.1016/j.jhep.2008.04.022

26. Bianchi A, Dai Prà M (2008) Twenty years after Spinnler and Tognoni: new instruments in the Italian neuropsychologist's toolbox. Neurol Sci 29:209–217. https://doi.org/10.1007/s10072-008-0970-x
27. Capitani E (1997) Normative data and neuropsychological assessment. common problems in clinical practice and research. Neuropsychol Rehabil 7:295–310. https://doi.org/10.1080/713755543
28. Brugnolo A, De Carli F, Accardo J, Amore M, Bosia LE, Bruzzaniti C, Cappa SF, Cocito L, Colazzo G, Ferrara M, Ghio L, Magi E, Mancardi GL, Nobili F, Pardini M, Rissotto R, Serrati C, Girtler N (2016) An updated Italian normative dataset for the Stroop color word test (SCWT). Neurol Sci 37:365–372. https://doi.org/10.1007/s10072-015-2428-2
29. Santangelo G, Siciliano M, Pedone R, Vitale C, Falco F, Bisogno R, Siano P, Barone P, Grossi D, Santangelo F, Trojano L (2015) Normative data for the Montreal Cognitive Assessment in an Italian population sample. Neurol Sci 36:585–591. https://doi.org/10.1007/s10072-014-1995-y
30. Reitan RM (1955) The relation of the Trail Making Test to organic brain damage. J Consult Psychol 19:393–394
31. Wilks SS (1941) Determination of samples size for setting tolerance limits. Ann Math Stat 12:91–96
32. Ackermann H (1985) Mehrdimensionale nicht-parametrische Normbereiche. Methodologische und medizinische. Aspekte. Springer-Verlag, Berlin
33. Spinnler H, Tognoni G (1987) Standardizzazione e taratura italiana di test neuropsicologici. [Italian standardization and adjustment of neuropsychological tests]. Ital J Neurol Sci 6:8–20
34. Cohen JW (1988) Statistical power analysis for the behavioral sciences, 2nd edn. Lawrence Erlbaum Associates Publishers, Hillsdale
35. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B Stat Methodol 57:289–300. https://doi.org/10.2307/2346101
36. Bezdicek O, Stepankova H, Axelrod BN, Nikolai T, Sulc Z, Jech R, Růžička E, Kopecek M (2017) Clinimetric validity of the Trail Making Test Czech version in Parkinson's disease and normative data for older adults. Clin Neuropsychol 31:42–60. https://doi.org/10.1080/13854046.2017.1324045
37. Knight RG, McMahon J, Green TJ, Skeaff CM (2006) Regression equations for predicting scores of persons over 65 on the rey auditory verbal learning test, the mini-mental state examination, the trail making test and semantic fluency measures. Br J Clin Psychol 45:393–402
38. Corrigan JD, Hinkeldey NS (1987) Relationships between parts A and B of the Trail Making Test. J Clin Psychol 43:402–409
39. Lamberty GJP, Chatel SH, Bieliauskas DM, Linas A (1994) Derived Trail Making Test indices: a preliminary report. Neuropsy Neuropsy Behav 7:230–234
40. Strauss E, Sherman EMS, Spreen O (2006) A compendium of neuropsychological tests: administration, norms, and commentary, 3rd edn. Oxford University Press, New York
41. Caffarra P, Vezzadini G, Dieci F, Zonato F, Venneri A (2002) Rey-Osterrieth complex figure: normative values in an Italian population sample. Neurol Sci 22:443–447
42. Pigliautile M, Chiesi F, Rossetti S, Conestabile Della Staffa M, Ricci M, Federici S, Chiloiro D, Primi C, Mecocci P (2015) Normative data for the ACE-R in an Italian population sample. Neurol Sci 36:2185–2190
43. Catricalà E, Gobbi E, Battista P, Miozzo A, Polito C, Boschi V, Esposito V, Cuoco S, Barone P, Sorbi S, Cappa SF, Garrard P (2017) SAND: a screening for aphasia in nurodegeneration. Development and normative data. Neurol Sci 38:1469–1483
44. Basagni B, Luzzatti C, Navarrete E, Caputo M, Scrocco G, Damora A, Giunchi L, Gemignani P, Caiazzo A, Gambini MG, Avesani R, Mancuso M, Trojano L, De Tanti A (2017) VRT (verbal reasoning test): a new test for assessment of verbal reasoning. Test realization and Italian normative data from a multicentric study. Neurol Sci 38:643–650
45. Appollonio I, Leone M, Isella V, Piamarta F, Consoli T, Villa ML, Forapani E, Russo A, Nichelli P (2005) The Frontal Assessment Battery (FAB): normative values in an Italian population sample. Neurol Sci 26:108–116