INTERNATIONAL JOURNAL OF GERIATRIC PSYCHIATRY
Int J Geriatr Psychiatry 2006; **21**: 844–852.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/gps.1570

A normative study of the Trail Making Test in Korean elders

Eun Hyun Seo[1,2], Dong Young Lee[1,2], Ki Woong Kim[3], Jung Hie Lee[4], Jin Hyeong Jhoo[5], Jong Choul Youn[6], IL Han Choo[2], Jin Ha[7] and Jong Inn Woo[1,3,8]*

[1] *Interdisciplinary Program of Cognitive Science, Seoul National University, Seoul Korea*
[2] *Department of Neuropsychiatry and Clinical Research Institute, Seoul National University Hospital, Seoul, Korea*
[3] *Department of Neuropsychiatry, Seoul National University Bundang Hospital, Seongnam, Kyunggi, Korea*
[4] *Department of Neuropsychiatry, Kangwon National University Hospital, Chuncheon, Kangwon, Korea*
[5] *Department of Psychiatry, Pundang Jesaeng Hospital, Daejin Medical Center, Seongnam, Kyunggi, Korea*
[6] *Department of Neuropsychiatry, Kyunggi Provincial Hospital for the Elderly, Yongin, Kyunggi, Korea*
[7] *College of Nursing, Hanyang University, Seoul Korea*
[8] *Neuroscience Research Institute of the Medical Research Center, Seoul National University, Seoul, Korea*

SUMMARY

Objective The purpose of this study was to explore the effects of age, education and gender on the performance of the Trail Making Test (TMT) and provide normative information in Korean elders.
Methods The TMT was administered to 997 community-dwelling volunteers aged 60–90. People with serious neurological, medical and psychiatric disorders, including dementia, were excluded.
Results Education and age had significant effects on both parts of the TMT. Gender also had an effect on part A of the TMT (Trail A). Based on these results, the norms of Trail A stratified by age (four overlapping tables), education (four strata) and gender, and the norms of part B of TMT (Trail B) stratified by age (four overlapping tables) and education (three strata).
Conclusions Age and educational level had a considerable influence on both Trail A and B. Our normative information on the Trail A will be useful in the elders with poor educational attainment and can be utilized for cross-cultural comparison of the Trail A performance. The fact that a large number of elders fail to complete Trail B indicates a limited applicability of Trail B in elderly population, particularly with poor educational background. Copyright © 2006 John Wiley & Sons, Ltd.

KEY WORDS — Trail Making Test; normative data; elders; age; education; gender; Koreans

INTRODUCTION

The Trail Making Test (TMT) is a widely used neuropsychological measure that is known as a test of psychomotor speed, attention, sequencing, mental flexibility, and visual scanning (Spreen and Strauss, 1998; Mitrushina *et al*., 1999). The TMT consists of parts A and B, which are considered to assess somewhat different cognitive processes. Although specific cognitive components contributing to each part of the TMT performance are not well-defined, part A of the TMT (Trail A) is regarded to be more related to attention and part B (Trail B) is to executive function (O'Donnell *et al*., 1994; Spreen and Strauss, 1998; Stuss *et al*., 2001).

The TMT performance has been found to be related to demographic factors. The relationship between older age and declining performance of the TMT has been consistently reported, particularly in individuals above 50 years of age (Kennedy, 1981; Bornstein, 1985; Ivnik *et al*., 1996; Soukup *et al*., 1998; Spreen and Strauss, 1998; Tombaugh, 2004; Lucas *et al*., 2005). Influence of educational level has also known to be significant (Waldman *et al*., 1992; Spreen and Strauss, 1998; Stuss *et al*., 2001; Lucas *et al*., 2005), whereas gender effect on the TMT is controversial. For example, some studies reported significant gender effects for Trail A (McCurry *et al*., 2001) or Trail B (Wiederholt *et al*., 1993), whereas other studies (Ivnik *et al*., 1996; Tombaugh, 2004; Lucas *et al*., 2005) reported no gender effects for either part. In order to

*Correspondence to: J. I. Woo, Department of Neuropsychiatry, Seoul National University Hospital, 28 Yongon-dong, Chongno-gu, Seoul, 110-744, Korea. E-mail: jiwoomd@plaza.snu.ac.kr

Received 13 September 2005
Accepted 13 February 2006

make the correct interpretation, therefore, normative data of the TMT considering age, education, or gender is needed.

Although several TMT norms have been reported, most of them contained only a small number of subjects with a limited age and educational range (Soukup *et al*., 1998; Spreen and Strauss, 1998; Mitrushina *et al*., 1999). More recently published norms (Ivnik *et al*., 1996; Tombaugh, 2004; Lucas *et al*., 2005; Steinberg *et al*., 2005) reported relatively comprehensive normative data, but their subjects were mainly from well-educated and English-speaking countries such as the US or Canada. Compared with western Caucasians, a considerable number of older people from developing countries including Korea are poorly educated and are rarely exposed to neuropsychological tests. There may be performance differences between those who have experienced cognitive tests with pencil and paper and those who have never experienced that kind of test before. Pre-exposure to cognitive tests may influence the performance. In addition, when ability tests are translated and (or) adapted from an original language and culture to a new one, there can be confounders (Geisinger, 1994; Greenfield, 1997). Therefore, to the elders with poorer educational background, particularly from developing countries, it is not appropriate to apply previously published normative data of TMT.

This study aimed at investigating the effect of age, education, and gender on the performance of the TMT and to provide the initial normative information in Korean elders.

METHODS

Study population

Nine hundred and ninety-seven healthy people aged 60 to 90 were included in this study. Seven centers located in a variety of regions in Seoul, Korea (two public health centers, one welfare center for the elderly, and three dementia or memory clinics) had been conducting a service program for the early detection and management of dementia for older people in the community with the support of the Korean Association for Dementia. An announcement for this program was made through the local newspapers and posters placed on a bulletin board at each center. The study subjects were recruited from a pool of older people registered in this program from January 1997 to October 2003. All of them lived in the community. Informed consent was obtained from each participant according to the procedures approved by the institutional review board of each center.

Each subject was examined by a psychiatrist with advanced training in neuropsychiatry and dementia research according to the protocol in the Korean Version of the Consortium to Established a Registry for Alzheimer's Disease (CERAD) clinical assessment battery (Lee *et al*., 2002). The battery consisted of a standardized clinical interview on demographic information, cognitive and functional status, drug inventory, depression and medical history, a cognitive state examination including the six-item Korean version of the Short Blessed test (Lee *et al*., 1999), and a general physical and neurological examination. Reliable informants were also interviewed to acquire accurate information regarding the cognitive and functional changes and medical history of the subjects. A panel consisting of four psychiatrists with expertise in dementia research made the clinical decisions including dementia diagnosis. This panel reviewed all available raw data from the clinical evaluation.

All subjects satisfied the strict entry criteria excluding dementia and other serious medical, psychiatric, and neurological disorders that could affect mental function. A diagnosis of dementia was made according to the criteria of the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV*; American Psychiatric Association, 1994). The performance of TMT was not considered when diagnosing dementia.

All subjects possessed adequate vision and hearing, although many wore glasses and some required a hearing aid. Individuals with minor physical abnormalities (e.g. diabetes with no serious complications, essential hypertension, mild hearing loss, etc.) were included. For Trail A, people who could not read or could not recall the sequence of Arabic numerals were excluded. For Trail B, people who could not understand the test instructions, could not read or couldn't recall the sequence of the Korean alphabet were excluded.

Measurement

Trained psychologist and nurses administered the TMT at each site. The TMT test materials and guidelines used in this study are the parts of the Korean version of the CERAD neuropsychological assessment battery (Lee *et al*., 2002). The TMT was administered according to the guidelines presented by CERAD (The Consortium to Established a Registry for Alzheimer's Disease, 1994).

Trail A. This part requires the individuals to draw a line as rapidly as possible joining consecutive numbers (i.e. 1–2–3–…24–25) pseudorandomly arranged on a 10×11.6 inch page. A maximum time of 360 sec is allowed for Trail A. Once the individual commits 5 errors, the test is discontinued.

Korean version of Trail B. The original English version of Trail B required the individuals to draw a line alternating between consecutive numbers and alphabet (i.e. 1–A–2–B–3–C … –L–13) as rapidly as possible on the same size of the page. In the Korean version of Trail B, Korean alphabets were substituted for English ones based on corresponding order and arrangement. The individuals, therefore, should connect the targets alternating between consecutive numbers and Korean alphabets (i.e. 1–ㄱ[*ga*]–2–ㄴ[*na*]–3–ㄷ[*da*] … ㅌ[*ta*]–13). A maximum time of 300 sec is allowed for Trail B. Once the individual commits 5 errors, the test is discontinued.

Before each part of test trial, a practice trial of six items was administered to make sure the individuals understood each task. When the individuals made errors, the examiner called it to their attention immediately, and had them proceed from the point where the mistake occurred. The time in sec taken to complete each part of the task was measured. Errors were also counted but were not considered as a score indicating the performance level of the subject.

Test–retest reliability

To assess test–retest reliability, the TMT was administered twice at approximately a 2-month interval to the individuals, who agreed to participate in the retest session. The same examiner conducted the retest session.

Statistical analysis

A stepwise multiple linear regression analysis was performed to assess the relative contribution of age, education, and gender on each test score. Age and education were entered as continuous variables and gender was coded as 0 and 1 for woman and man, respectively.

A series of $2 \times 5 \times 2$ analyses of variance (ANOVA) were also performed to determine any main effects and interactions of age (60–74 and 75–90 years), education (0–3, 4–6, 7–9, 10–12, and $\geq$13 years), and gender (woman and man) on the test. Educational levels were divided into five groups corresponding categorical distinctions in Korean public education system. Six years of elementary school were, however, divided into two groups (i.e. 0–3 and 4–6 years) because considerable number of participants ($n = 222$) were finished only less than 4 years of education, and their test performance was significantly different from those who finished 4 to 6 years of education. Age was divided into two groups because the performance between ages 60–74 and 75–90 groups were significantly distinct. *Post hoc* contrasts with Tukey's method were conducted when any main effect of education was determined to be significant by ANOVA at the $p < 0.05$ levels.

Test–retest reliability was tested by Pearson correlation analysis and paired sample *t*-test.

RESULTS

Demographic characteristics

The demographic characteristics of the subjects who completed Trails A and B are presented in Table 1. Sixty-three percent of the subjects were women, and the mean age and education of the male group were significantly higher than those of the female group ($t = 8.92$, $p < 0.001$ for age; $t = 11.58$, $p < 0.001$ for education).

Trail A was administered to 997 subjects and all completed it. Trail B was initially administered to 773 individuals, but 196 subjects among them were excluded before the administration of Trail B because they could not understand test instructions ($n = 88$) or were unable to recall the sequence of the Korean alphabet correctly ($n = 108$). Among 577 individuals to whom Trail B was actually administered, 181 did not complete it because they could not finish the test within 300 sec ($n = 136$) or made 5 errors ($n = 45$). Ultimately, only 396 (51.2%) among 773 initial candidate subjects for Trail B completed it. Failure of completion was more frequent in women than in men [$\chi^2(1, n = 773) = 31.00$, $p < 0.001$], in lower educational levels (i.e. 0–3, and 4–6 years of education) than in higher ones (i.e. 7–9, 10–12, and $\geq$13 years of education) [$\chi^2(5, n = 773) = 110.46$, $p < 0.001$], and in the older age group (i.e. ages 75–90) than the younger one (i.e. ages 60–74) [$\chi^2(1, n = 773) = 29.86$, $p < 0.001$].

Test–retest reliability

The TMT was administered twice at a 2-month interval to 27 individuals for Trail A and 19 for Trail B. Age and educational levels of retested sample were not significantly different from those of the whole study

Table 1. Demographic characteristics of the subjects

	Men	Women	Total
Number	369 (193)	628 (203)	997 (396)
Age (yr)	72.1 ± 6.2[a] (70.6 ± 5.5)	68.9 ± 5.1 (67.6 ± 4.2)	70.1 ± 5.7 (69.0 ± 5.1)
60–64	35[b] (17)	130 (46)	165 (63)
65–69	102 (71)	247 (95)	349 (166)
70–74	112 (69)	154 (49)	266 (118)
75–79	71 (23)	78 (10)	149 (33)
80–84	38 (7)	17 (3)	55 (10)
85–89	11 (6)	2 (0)	13 (6)
Education	9.2 ± 5.6 (10.6 ± 5.9)	5.9 ± 4.5 (8.2 ± 3.9)	7.2 ± 5.2 (9.4 ± 5.1)
0	33 (6)	96 (8)	129 (14)
1–3	15 (5)	78 (10)	93 (15)
4–6	95 (44)	275 (79)	370 (123)
7–9	64 (35)	80 (39)	144 (74)
10–12	80 (46)	74 (50)	154 (96)
≥13	82 (57)	25 (17)	107 (74)

Note. The number and mean of subjects who completed Trail B is in parenthesis.
[a]Mean ± *SD*.
[b]Number.

subjects, but ratio of female subjects were slightly higher in the retested sample ($p = 0.044$).

As shown in Table 2, correlations between test and retest score were significant for both parts ($r = 0.79$ for Trail A; $r = 0.82$ for Trail B). The performances for both parts were slightly improved at the second assessment, however, the change in performance was not significant.

Effects of age, education and gender on the test

Stepwise multiple regression analysis revealed that age, education and gender affected the test scores significantly. Among the demographic variables, education accounted for the greatest proportion of the score variance of both Trails A and B (see Table 3). Education accounted for 20.60% and 13.50% of the variance for Trails A and B, respectively. Age and gender also accounted for the variance of each Trail significantly, though the size of variance was less than 2% for both Trails.

Three-way ANOVA was also performed to determine any main effects or interactions among demographic variables. Main effects of age [$F(1, 977) = 16.31$, $p < 0.001$], education [$F\ (4, 977) = 36.74$, $p < 0.001$], and gender [$F\ (1, 977) = 6.89$, $p < 0.01$] were significant for Trail A. Main effects of age [$F(1, 377) = 4.16$, $p < 0.05$] and education ($F(4, 377) = 4.22$, $p < 0.01$) were significant, but there was no significant gender difference for Trail B. No interactions among demographic variables were found for both Trails. Because there was a main effect of education, educational groups (i.e. 0–3, 4–6, 7–9, 10–12, and ≥13 years) were compared using the *post hoc* contrasts. For Trail A, the significant difference was observed between any two among 0–3, 4–6, 7–9, and 10–12 years of educational groups, whereas there was no significant difference

Table 2. Test-retest reliabilities of Trails A and B

	T1	T2	*R*
Trail A ($n = 27$)	99.59 ± 40.20[a]	89.78 ± 38.36	0.79***
Trail B ($n = 19$)	219.32 ± 77.50	208.68 ± 70.94	0.82***

Note. T1 = 1st assessment, T2 = 2nd assessment, r = Pearson's correlation coefficient.
[a]Mean ± *SD*.
$^{***}p < 0.001$.

Table 3. Stepwise multiple linear regression of age, education and gender on Trails A and B

Variables	*B*	*SE(B)*	β	ΔR^2
Trail A				
Education	−4.00	0.31	−0.38***	20.60
Age	1.76	0.28	0.19***	1.80
Gender	−17.01	3.49	−0.15***	1.80
Trail B				
Education	−4.00	0.64	−0.31***	13.50
Age	1.45	0.55	0.13**	1.50
Gender	−16.24	5.70	−0.15**	0.90

Note. B = regression coefficient; $SE(B)$ = standard error of B; β = standardized regression coefficient; ΔR^2 = percent variance explained by the each variable, Age and education were entered as a continuous variable, and gender was coded as 0 and 1 for woman and man, respectively.
$^{**}p < 0.01$;
$^{***}p < 0.001$.

between 10–12 and ≥13 years of educational groups. For Trail B, significant differences were found between 0–3 and all of the ≥7 years of educational groups (i.e. 7–9, 10–12, and ≥13 years), between 4–6 and 10–12, and between 4–6 and ≥13 years of educational groups, whereas no significant difference was observed between 0–3 and 4–6, and between 10–12 and ≥13 years of educational groups.

Normative data

On the basis of the analysis for the effect of demographic variables, we decided to stratify the norm of Trail A by age, education, and gender (Table 4 and 5 and the norm of Trail B only by age and education (Table 6).

For both Trails A and B, the age groups were divided into four overlapping stratifications following the procedures described by Lee *et al*. (2004). The overlapping strata have been initially suggested as a way to maximize the clinical usefulness of normative test data by Pauker (1988). Tables composed of overlapping age strata with midpoint ages occurring at five-year intervals (i.e. 67, 72, 77, and 82 years of age, respectively for four strata) were developed as shown in Tables 4–6. In each table, the user should select the age strata with the closest midpoint from the subject's age to interpret the TMT performance. As a result, normative data from ages 60–74, 65–79, 70–84, and 75–90 years were used for persons whose ages range from 60–69, 70–74, 75–79, and 80–90 years, respectively. Each of these ranges was derived to encompass all ages closer to one midpoint age than to the other adjacent midpoints.

The educational groups were divided into four strata (i.e. 0–3, 4–6, 7–9, and ≥10 years of education) for Trail A, and 3 strata (i.e. 0–6, 7–9, and ≥10 years of education) for Trail B. These strata of education were determined considering the results from the *post hoc* contrasts between the five educational groups, and the number of subjects within each cell.

The performance scores of the TMT stratified by demographic variables were shown in the forms of a mean, a standard deviation, a median and a range from the 5th to the 95th percentile (Tables 4–6). Since there were a considerable number of failures for Trail B, the

Table 4. Normative data of Trail A for men: mean, standard deviation, median and range from the 5th to the 95th percentile

Education (years)		0–3	4–6	7–9	≥10
Age (years)					
60–69[a]	*n*	20	57	46	126
	Mean	119.20	84.39	75.89	58.79
	SD	69.98	39.72	46.48	22.32
	5th percentile	50.45	35.90	34.00	31.70
	Median	90.50	73.00	61.50	55.00
	95th Percentile	288.75	165.70	181.95	103.90
70–74[b]	*n*	30	76	49	130
	Mean	137.80	88.24	79.16	61.64
	SD	75.22	40.07	43.99	22.98
	5th percentile	54.95	36.85	34.50	32.55
	Median	106.50	75.00	66.00	57.00
	95th Percentile	294.50	164.20	181.50	106.90
75–79[c]	*n*	33	68	39	81
	Mean	140.49	95.28	80.51	67.90
	SD	74.95	40.47	34.69	29.82
	5th percentile	50.00	39.25	38.00	31.10
	Median	123.00	92.50	74.00	63.00
	95th Percentile	290.90	167.40	183.00	120.80
80–90[d]	*n*	28	38	18	36
	Mean	138.68	101.95	85.06	79.44
	SD	73.52	43.80	34.77	36.85
	5th percentile	52.70	43.90	38.00	43.05
	Median	116.50	94.00	79.00	74.50
	95th Percentile	294.15	188.40	187.00	160.15

[a]Normative data from the people whose ages range from 60 to 74 years;
[b]Normative data from the people whose ages range from 65 to 79 years;
[c]Normative data from the people whose ages range from 70 to 84 years;
[d]Normative data from the people whose ages range from 75 to 90 years.

Table 5. Normative data of Trail A for women: mean, standard deviation, median and range from the 5th to the 95th percentile

Education (years)		0–3	4–6	7–9	≥10
Age(years)					
60–69[a]	*n*	131	243	69	88
	Mean	129.83	102.50	81.55	63.99
	SD	60.49	43.62	40.66	24.02
	5th percentile	48.20	43.20	42.50	33.45
	Median	117.00	94.00	76.00	60.00
	95th Percentile	241.20	188.00	134.50	110.10
70–74[b]	*n*	142	211	56	70
	Mean	143.35	105.70	81.54	69.84
	SD	67.01	44.84	25.07	27.46
	5th percentile	49.60	43.60	46.00	33.50
	Median	131.50	100.00	77.50	64.00
	95th Percentile	296.40	188.80	141.75	129.30
75–79[c]	*n*	84	104	30	31
	Mean	161.17	107.50	88.00	80.71
	SD	71.82	50.98	25.50	33.28
	5th percentile	57.50	43.00	48.75	28.40
	Median	149.50	98.50	82.50	79.00
	95th Percentile	312.75	200.50	147.80	145.40
80–90[d]	*n*	43	32	11	11
	Mean	181.12	118.84	99.36	92.73
	SD	77.78	59.72	35.54	32.94
	5th percentile	65.80	57.25	46.00	30.00
	Median	185.00	105.00	105.00	92.00
	95th Percentile	337.00	291.65	150.00	149.00

[a]Normative data from the people whose ages range from 60 to 74 years;
[b]Normative data from the people whose ages range from 65 to 79 years;
[c]Normative data from the people whose ages range from 70 to 84 years;
[d]Normative data from the people whose ages range from 75 to 90 years.

percentage of completion within each cell is presented in Table 6. On the other hand, because there were significant differences on the performance of Trails A between the whole sample and the individuals who completed both Trails A and B (age, education, and gender controlled ANOVA, $F\ (1,974) = 17.16$, $p < 0.05$), the normative data of Trail A in the latter group ($n = 396$) is also presented in Table 6.

DISCUSSION

The present investigation was conducted to examine the effect of demographic factors on the performance of the two parts of the TMT and to provide normative information to assist clinicians and researchers in their interpretation of the test. The results of this study indicated that an older age and lower educational level were associated with lower performance in both Trails A and B. These findings are consistent with those reported in other normative studies (Bornstein, 1985; Ivnik *et al*., 1996; Soukup *et al*., 1998; Spreen and Strauss, 1998; Tombaugh, 2004; Lucas *et al*., 2005). However, the relative contribution of age and education to test performance found in our study was different from that in previous studies. Age has been reported to have a greater effect than education in many previous studies (Bornstein, 1985; Ivnik *et al*., 1996; Spreen and Strauss, 1998; Tombaugh, 2004), whereas education had a greater effect in our study. This discrepancy may reflect the different distribution of educational attainment between the subjects of other studies and ours. In our study, a large number of participants (59.5%) were poorly educated (i.e. less than seven years of education) and the range of educational level was very wide, from zero to 22 years. In contrast, the subjects of the studies of Ivnik *et al*. (1996) and Tombaugh (2004) had relatively homogeneous educational attainment and nearly all of them were highly educated (i.e. more than seven years of education).

The influence of gender was not identical between Trails A and B. Men performed better than women on Trail A, while there was no gender effect found for Trail B. However, it should be noted that for Trail B there were many more failures for women than for men. A significant gender effect for Trail B seems to

Table 6. Normative data of Trails A and B: mean, standard deviation, median and range from the 5th to the 95th percentile

		Trail A[e]			Trail B		
Education (years)		0–6	7–9	≥10	0–6	7–9	≥10
Age(years)							
60–69[a]	*n*	130	65	154	130 (39.40)[f]	65 (67.70)	154 (79.20)
	Mean	77.75	68.32	56.18	201.85	185.03	159.76
	SD	29.57	20.57	19.72	55.10	54.88	50.01
	5th percentile	39.00	34.30	31.65	108.45	98.20	77.70
	Median	73.50	69.00	52.50	198.50	188.00	153.50
	95th percentile	143.15	105.10	94.10	291.80	288.20	243.25
70–74[b]	*n*	126	57	134	126 (37.40)	57 (64.80)	134 (76.60)
	Mean	79.06	67.93	58.01	202.34	190.47	163.85
	SD	29.78	19.96	20.35	55.29	49.10	52.40
	5th percentile	39.00	34.90	30.75	104.65	113.90	77.00
	Median	74.00	69.00	55.00	197.00	189.00	155.00
	95th percentile	146.50	103.30	99.00	290.65	290.00	254.75
75–79[c]	*n*	69	31	61	69 (33.00)	31 (53.40)	61 (64.20)
	Mean	76.35	67.35	59.18	201.43	188.81	170.02
	SD	27.39	17.60	20.71	55.54	44.36	54.39
	5th percentile	40.00	36.40	28.20	102.50	106.20	78.30
	Median	69.00	69.00	58.00	196.00	192.00	166.00
	95th percentile	125.50	100.60	101.60	290.50	270.80	270.20
80–90[d]	*n*	22	9	18	22 (23.40)	9 (39.10)	18 (47.40)
	Mean	84.05	65.44	69.50	217.27	195.45	197.28
	SD	27.06	17.10	19.23	53.97	39.72	45.75
	5th percentile	50.00	38.00	30.00	99.35	137.00	112.00
	Median	83.50	69.00	67.50	220.50	196.00	195.00
	95th percentile	156.60	98.00	102.00	290.85	257.00	290.00

[a]Normative data from the people whose ages range from 60 to 74 years;
[b]Normative data from the people whose ages range from 65 to 79 years;
[c]Normative data from the people whose ages range from 70 to 84 years;
[d]Normative data from the people whose ages range from 75 to 90 years;
[e]Normative data from the people who completed both Trails A and B;
[f]Percentage of completed Trail B is in parenthesis.

have been undetected because a considerable number of women with poorer cognitive performance could not complete the test, thus obscuring the actual gender difference. A gender effect on the TMT is still controversial in the older population. The significant gender effect seen in our study may also be explained by a greater proportion of the poorly educated among our subjects with different social roles between man and woman. Older men, even those with little formal education, had more opportunity to obtain informal learning during their outside occupational activities. On the other hand, the poorly educated women, who were usually devoted to housework, had fewer opportunities for outside intellectual stimulation. However, this gender effect tends to be less important in those who received at least seven years of education (Tables 4 and 5), although no significant interaction effect of education and gender was found.

As noted in previous studies, the performance on Trail B was poorer than that of Trail A in this study. Trail B has a more complex physical layout, increased task demands, and it is more difficult to process and complete than Trail A (Fossum *et al*., 1992; Rossini and Karl, 1994; Arnett and Labovitz, 1995; Gaudino *et al*., 1995). Nearly half of the initial candidates for Trail B failed to finish the test due to several reasons, whereas no such failure occurred for Trail A. Such large numbers of failures in completing Trail B have not been previously reported in other normative studies on the TMT. They were partly related to the method of test administration we adopted. A maximum of 300 sec and five errors was allowed for Trail B in our study. Following the recommendations of CERAD (The Consortium to Established a Registry for Morris *et al*., 1989; Alzheimer's Disease, 1994), we applied this limitation for time and number of errors so that our participants who had considerable difficulty with this task would have minimal discomfort. Nearly half of the overall failures (181 subjects) originated from this limitation. Additionally, many candidate individuals were excluded before the test was administered because they did not understand the test instructions or recall the sequence of the Korean alphabet correctly. As indicated in the discussion on gender effect, the failures as a whole seem to be related to poor educational attainment. There were actually many more frequent failures in individuals with lower educational attainment than in higher attainment (Table 6). These findings indicate that Trail B of the Korean version of TMT has considerable limitations in assessing the cognitive function of older people, in particular, poorly-educated individuals. Because our normative data for Trail B (Table 6) are based only on the scores of the subjects completing the test, the users of this norm should initially consider the completion rate shown in each cell and thereafter cautiously interpret a performance score. For the sake of users' convenience in comparing the performance for Trail B with that of Trail A, we provide the normative information for Trail A obtained only from the Trail B completers ($n = 396$, Table 6), as well as that to the whole group ($n = 997$, Tables 4 and 5).

Soukup *et al*. (1998), by reviewing 26 published reports of normative data for the TMT, reported the broad range of variability associated with 'normative performance' among various samples, particularly on the Trail B and among elders. When we roughly compared the mean scores of Trail B among various elderly samples with similar educational level, the actual differences were suspected. For example, there was a big difference between Canadian (the mean score of part B was 74.55 ± 19.55; Tombaugh, 2004) and Italian samples (the mean score was 164.54 ± 97.41; Giovagnoli *et al*., 1996). Difference was also found between Japanese (the mean scores of part B was 92.7 ± 33.0; Abe *et al*., 2004) and our Korean samples (the mean score was 163.66 ± 47.40). Besides age or gender differences of populations, the different alphabets probably contributed to the discrepancies of normative performances of old people between various countries.

We adopted the overlapping age stratification method proposed by Pauker (1988) and were able to present more stratified and accurate normative data with adequate numbers of subjects for most normative cells. However, the cells of the oldest age group with a higher educational level still have limited sample size (e.g. 11 in the cell for 75–90 age group with educational level of 7–9 and $\geq$10 years in Table 5 and 9 in the cell for 75–90 age group with educational level of 7–9 years in Table 6). Therefore, users should exercise greater caution when the test scores of individuals within such cells are interpreted.

Test–retest reliabilities in our study were higher than those in other studies. For example, Snow *et al.* (1988) reported reliabilities of 0.64 and 0.72 for Trails A and B, respectively. Mitrushina and Satz (1991) also reported reliabilities of 0.53 and 0.67 for Trails A and B, respectively. This may be related to the test interval and subject recruitment. The test interval between test and retest was two months and the individuals ($n = 27$ for Trail A; $n = 19$ for Trail B) who agreed to the retest session were consecutively recruited in our study, whereas the test interval in two previous studies was one year and all participants were retested. It is possible that a longer test interval may lower the reliability coefficient.

In conclusion, age and educational level had a considerable influence on both Trails A and B. Gender also had a significant effect on the performance of Trail A. Because previously published norms were mostly based on the well-educated elders, our normative information on the Trail A will be useful in the elders with poor educational attainment and can be utilized for cross-cultural comparison of the Trail A performance. The fact that a large number of elders fail to complete Trail B indicates a limited applicability of Trail B in elderly populations, particularly those with poor educational backgrounds.

ACKNOWLEDGEMENTS

This work was supported by the Grant No. 04-2000-045 from the Seoul National University Hospital and Biotech 2000 (Grants 97-N1-02-03-A-12 and 98-N1-02-03-A-12) of the Ministry of Science and Technology of Korea.

REFERENCES

Abe M, Suzuki K, Okada K, *et al*. 2004. Normative data on tests for frontal lobe function: Trail Making Test, Verbal fluency, Wisconsin Card Sorting Test (Keio version). *No To Shinkei* **56**: 567–574.

Ameican Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association Press: Washington, DC.

Arnett JA, Labovitz S. 1995. Effect of physical layout in performance of the Trail Making Test. *Psychol Assess* **7**: 220–221.

Bornstein RA. 1985. Normative data on selected neuropsychological measures from a nonclinical sample. *J Clin Psychol* **41**: 651–659.

Fossum B, Holmberg H, Reinvang I. 1992. Spatial and symbolic factors in performance on the Trail Making Test. *Neuropsychology* **6**: 71–75.

Gaudino EA, Geisler MW, Squires NK. 1995. Construct validity of the Trail Making Test: what makes part B harder? *J Clin Exp Neuropsychol* **17**: 529–535.

Geisinger KF. 1994. Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychol Assess* **6**: 304–312.

Giovagnoli AR, Del Pesce M, Masheroni S, Simoncelli M, Laiacona M, Capitani E. 1996. Trail Making Test: normative values from 287 normal adult controls. *Ital J of Neurolo Sci* **17**: 305–309.

Greenfield PM. 1997. You can't take it with you: why ability assessment don't cross cultures. *Am Psychol* **52**: 1115–1124.

Ivnik RJ, Malec JF, Smith GE, Tangalos EG, Petersen RC. 1996. Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, STROOP, TMT, and JLO. *Clin Neuropsychol* **10**: 262–278.

Kennedy KJ. 1981. Age effects on Trail Making Test performance. *Percept Mot Skills* **52**: 671–675.

Lee DY, Yoon JC, Lee KU, *et al*. 1999. Reliability and validity of the Korean version of Short Blessed Test (SBT-K) as a dementia screening instrument. *J Korean Neuropsychiatric Association* **38**: 1297–1307.

Lee JH, Lee KU, Lee DY, *et al*. 2002. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. *Gerontol B Psychol Sci Soc Sci* **57**: 47–53.

Lee DY, Lee KU, Lee JH, *et al*. 2004. A normative study of the CERAD neuropsychological assessment battery in the Korean elderly. *J Int Neuropsychol Soc* **10**: 72–81.

Lucas JA, Ivnik RJ, Smith GE, *et al*. 2005. Mayo's older African Americans normative studies: norms for Boston Naming Test, Controlled Oral Word Association, Category Fluency, Animal Naming, Token Test, WRAT-3 Reading, Trail Making Test, Stroop test, and Judgement of Line Orientation. *Clin Neuropsychol* **19**: 243–269.

McCurry SM, Gibbons LE, Umoto JM, *et al*. 2001. Neuropsychological test performance in a cognitively intact sample of older Japanese American adults. *Arch Clin Neuropsychol* **16**: 447–459.

Mitrushina MN, Boone KB, D'Ella L. 1999. *Handbook of Normative Data for Neuropsychological Assessment* Oxford University Press: New York; 33–64.

Mitrushina MN, Satz D. 1991. Effect of repeated administration of a neuropsychological battery in the elderly. *J Clin Psychol* **47**: 790–801.

Morris JC, Heyman A, Mohs RC, *et al*. 1989. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* **39**: 1159–1165.

O'Donnell JP, MacGregor LA, Dabrowski JJ, Oestreicher JM, Romero JJ. 1994. Construct validity of neuropsychological tests of conceptual and attentional abilities. *J Clin Psychol* **50**: 596–600.

Pauker JD. 1988. Constructing overlapping cell tables to maximize the clinical usefulness of normative data: rational and an example from neuropsychology. *J Clin Psychol* **44**: 930–933.

Rossini ED, Karl MA. 1994. The Trail Making Test A and B: a technical note on structural nonequivalence. *Percept Mot Skills* **78**: 625–626.

Snow WG, Tierney MC, Zorzitto ML, Fisher RH, Reid DW. 1988. One-year test-retest reliability of selected neuropsychological tests in older adults. *J Clin Exp Neuropsychol* **10**: 60 (abstract).

Soukup VM, Ingram R, Grady J, Schiess MC. 1998. Trail Making Test: issues in normative data selection. *Appl Neuropsychol* **5**: 65–73.

Spreen O, Strauss E. 1998. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*, 2nd edn. Oxford University Press: New York.

Steinberg BA, Bieliauskas LA, Smith GE, Ivnik RJ. 2005. Mayo's older African Americans normative studies: age- and IQ-adjusted norms for Trail Making Test, Stroop Test, and MAE Controlled Oral Word Association Test. *Clin Neuropsychol* **19**: 329–377.

Stuss DT, Bisschop SM, Alexander MP, Levine B, Katz D, Izukawa D. 2001. The Trail Making Test: a study in focal lesion patients. *Psychol Assess* **13**: 230–239.

The Consortium to Establish a Registry for Alzheimer's Disease. 1994. *CERAD Assessment Packet: Protocol 4b revised.* Duke University Medical Center: Durham, NC.

Tombaugh TN. 2004. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol* **19**: 203–214.

Waldmann BW, Dickson AL, Monahan MC, Kazelskis R. 1992. The relationship between intellectual ability and adult performance on the Trail Making Test and the Digit Symbol Modalities Test. *J Clin Psychol* **48**: 360–363.

Wiederholt WC, Cahn D, Butter N, Salmon DP, Kritz-Silverstein D, Barret-Conner E. 1993. Effects of age, gender, and education on selected neuropsychological tests in an elderly community cohort. *J Am Geriatr Soc* **41**: 639–647.

