International Journal of Gerontology 12 (2018) 336–339

Contents lists available at ScienceDirect

International Journal of Gerontology

journal homepage: www.ijge-online.com

Original Article

Reliability of the Chinese Version of the Trail Making Test and Stroop Color and Word Test among Older Adults

Ray-Yau Wang [a], Jun-Hong Zhou [a], Yuan-Chen Huang [b], Yea-Ru Yang [a, c *]

[a] Department of Physical Therapy and Assistive Technology, National Yang-Ming University, Taipei, 11221, Taiwan, [b] Department of Physical Medicine and Rehabilitation, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi City, 60002, Taiwan, [c] Preventive Medicine Research Center, National Yang-Ming University, Taipei, 11221, Taiwan

ARTICLE INFO

Article history:
Received 15 March 2018
Received in revised form
17 May 2018
Accepted 22 June 2018
Available online 17 July 2018

Keywords:
alternate form reliability,
executive function,
measurement tools,
older adults,
test-retest reliability

SUMMARY

Background: Both Trail Making Test (TMT) and Stroop Color and Word Test (SCWT) are the most popular neuropsychological tests for assessing executive function. This study aimed to examine alternate form reliability of the Chinese version of the TMT Part B (C-TMT-B) and test-retest reliability of the Chinese version of the TMT and SCWT among older adults.

Methods: Twenty participants were recruited in the alternate form reliability study and another 20 participants were recruited in the test-retest reliability study. Original version of the TMT-A and TMT-B and the Chinese version of the TMT-B and SCWT were used as the measurement tools. A retest was conducted 3–7 days later to assess its reliability. The reliability of tests was estimated with intraclass correlation coefficient (ICC) estimates and their 95% confident intervals.

Results: The alternate form reliability of C-TMT-B was moderate to excellent with ICC of 0.89 and 95% confident interval of 0.63–0.96. Test-retest reliability coefficients for TMT-A, C-TMT-B, C-SCWT with congruous condition, and C-SCWT with incongruous condition were estimated as 0.82, 0.93, 0.91, and 0.91, respectively.

Conclusion: Our findings suggest that the Chinese version of the TMT and SCWT are reliable instruments for measuring executive function among older adults.

1. Introduction

Executive function is a part of cognitive function and also known as higher level of cognitive function. Executive function has been defined as a set of cognitive skills necessary for planning, monitoring, and executing a sequence of goal-directed complex actions. Executive function is particularly affected by aging.[1] As age increases, executive function and attention deteriorate, affecting the ability of older people to engage in daily tasks.[2] Therefore, evaluation tools used for assessing executive function among older adults play an important role.

The Trail Making Test (TMT) and Stroop Color and Word Test (SCWT) are most frequently used tools to evaluate executive function among older adults.[3,4] Both tests are extensively used neuropsychological tests. TMT provides information on visual attention, task switching, speed of processing, and mental flexibility.[5] SCWT provides information on selective attention, cognitive flexibility, processing speed, and inhibitory control.[6] The validity and reliability of TMT and SCWT in the healthy older adults has been established.[7–10]

The Chinese versions of the TMT and SCWT have been used in previous studies. Lu and Bigler and Law et al. modified the TMT-B with sequential numbers in the Chinese characters substituting for the English alphabetical sequence.[11–13] Lu et al. and Wei et al. replaced numbers and letters of the TMT-B by circles and squares with numbers.[14,15] Chuang et al. modified the TMT-B with the Chinese zodiacs substituting for the English alphabet.[16] The Chinese version of the SCWT has also been approved as applicable for the Chinese population.[17,18] However, the reliability of these Chinese versions of the TMT-B and SCWT has not been established. Therefore, the purpose of the current study was to establish the alternate form reliability of the Chinese version of the TMT-B and assess the

* Corresponding author. Department of Physical Therapy and Assistive Technology, National Yang-Ming University, Taipei, Taiwan.
E-mail address: yryang@ym.edu.tw (Y.-R. Yang).

https://doi.org/10.1016/j.ijge.2018.06.003

test-retest reliability of the Chinese version of the TMT and SCWT among older adults.

2. Methods

2.1. Participants

Native Chinese-speaking participants were recruited from local community in Taiwan. All participants met the following inclusion criteria: (1) 65 years old or above; (2) a score of the Mini-mental State Examination greater than or equal to 24; and (3) educational level at least 6 years or ability to read the Chinese characters and/or the English alphabets. The exclusion criteria were as follows: (1) inability to grasp a pen; and (2) a diagnosis of hand movement disorders, dysgraphia, agraphia, color blindness or color vision deficiency.

2.2. Procedure

A cross-sectional study was conducted on older adults between December 2016 and December 2017. The study protocol was approved by the Institutional Review Board of National Yang-Ming University and registered in a clinical trial registry (ACTRN12617000686303). The purpose and nature of the studies were fully explained to the participants. All participants gave written, informed consent before study participation. For the alternate form reliability study, 20 participants were recruited. Ten participants were selected randomly for assessing of the original English version of the TMT-B first, followed by the Chinses version of the TMT-B (C-TMT-B) while the other ten participants followed the reverse order. For the test-retest reliability study, another 20 participants were recruited. TMT-A, C-TMT-B, and C-SCWT were measured twice with a 3–7-day interval between.

2.3. Instruments

2.3.1. Trail Making Test

TMT is one of the neuropsychological tests. It consists of two parts. TMT-A is composed of 25 numbers from 1 to 25. Participants must draw a line sequentially connecting these numbers. The original TMT-B is composed of 12 numbers and 12 letters. Participants must draw a line to connect alternately between numbers and letters (i.e., 1-A-2-B-3-C….). In the C-TMT-B, letters are replaced by 12 Chinese zodiacs (rat, cow, tiger, rabbit, dragon… in Chinese) (Fig. 1). Task requirement of the C-TMT-B is similar to the TMT-B except participants must alternate between numbers and the Chinese zodiacs. The score on each part is the number of seconds required to complete the task.

2.3.2. Stroop Color and Word Test

SCWT is also one of the neuropsychological tests. C-SCWT was used in this study. C-SCWT consists of two subtasks (Fig. 2). The material for each subtask is shown on a white A4 paper with 20 words on one page. The first subtask with congruous condition shows color words (36 mm × 34 mm) in random order (black, blue, red, yellow) printed in the same color ink with the word (i.e., the word blue printed in blue ink). The second subtask with incongruous condition contains color words (36 mm × 34 mm) printed in a different ink color (i.e., the word black printed in blue ink). Participants are required to name the color of the ink as quickly as possible within 45 s. The score is generated using the correct number of items completed on each subtask.

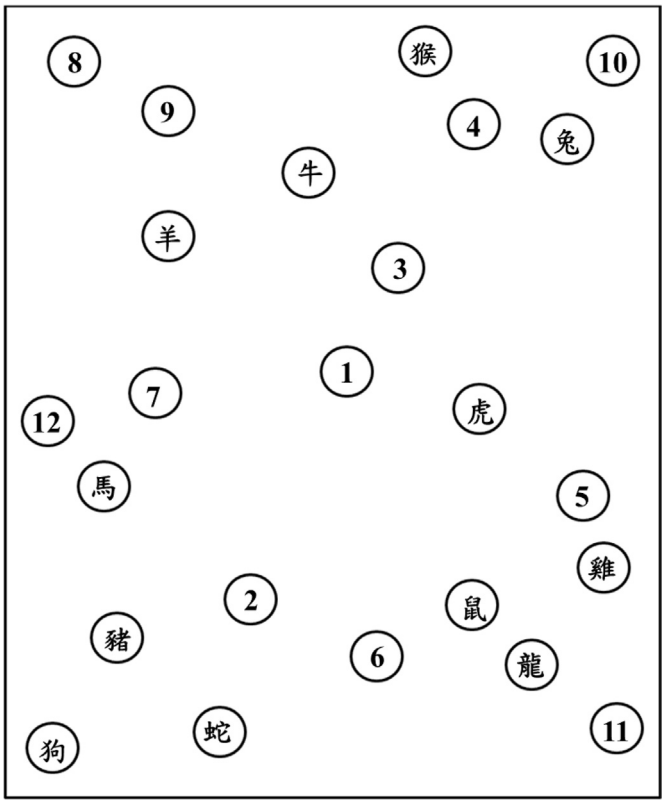

Fig. 1. The Chinese version of the Trail Making Test Part B.

2.4. Statistical analysis

All data were analyzed using SPSS 20.0. The distributions of the variables were presented as medians and percentiles of 25 and 75. The Wilcoxon signed-rank test was used to compare the medians between the TMT-B and C-TMT-B and to test the change in performance at retest. Intraclass correlation coefficient (ICC) estimates and their 95% confident intervals were calculated based on an absolute-agreement, 2-way mixed-effects model. ICCs ≥0.90 were considered as excellent reliability, values in the range of ≥0.75 to <0.90 were considered as good reliability, while values in the range of ≥0.50 to <0.75 were considered as moderate reliability.[19]

3. Results

The demographic characteristics of all participants are presented in Table 1. Some participants had a history of cardiac disease, diabetes, hypertension or hypothyroidism. No one had a history of neurological or psychiatric disorder. The scores of the TMT-B and C-TMT-B were 82.00 (48.25, 126.75) seconds and 63.50 (42.25, 89.25) seconds, respectively. Results of the Wilcoxon signed-rank test revealed that the scores of C-TMT-B were significant lower than the scores of the TMT-B ($p = 0.005$). In the alternate form reliability study, ICC value was 0.89 (95% confident interval = 0.63–0.96) that was considered as moderate to excellent reliability.

The Wilcoxon signed-rank test revealed a significant difference between the first test and the second test for the C-TMT-B ($p = 0.02$), with participants being faster at the second test compared to the first test. There was no significant difference between the test 1 and 2 for remaining measures. The results of the test-retest reliability are shown in Table 2. Good to excellent reliability (ICC ≥ 0.75) was found for the C-TMT-B, C-SCWT with

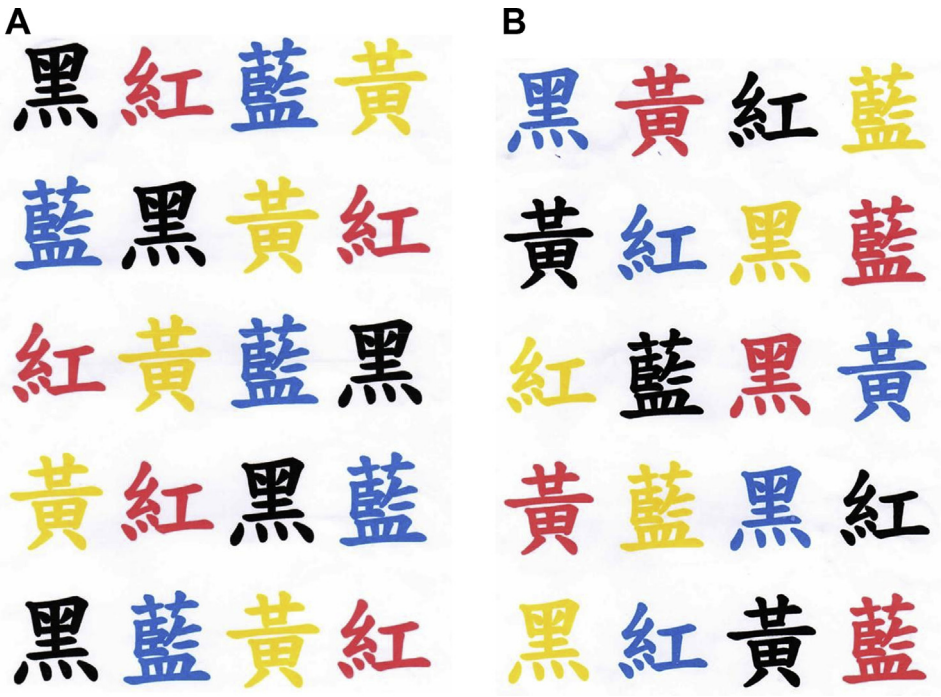

Fig. 2. The Chinese version of the Stroop Color and Word Test. (A) Congruous condition and (B) Incongruous condition.

Table 1
Demographic characteristics of the participants.

Characteristics	Alternate form reliability (n = 20)	Test-retest reliability (n = 20)
Age (years)	68.00 (66.00, 71.75)	78.00 (73.25, 84.00)
Gender (male/female)	9/11	8/12
Education (years)	13.00 (12.00, 16.00)	12.00 (9.00, 15.00)

Data are presented as the median (percentiles of 25, percentiles of 75) or number.

Table 2
Test-retest reliability of the Chinses version of the Trail Making Test and Stroop Color and Word Test among older adults.

Measures	The first test (n = 20)	The second test (n = 20)	Intraclass correlation coefficient	95% Confident interval
Chinses version of the Trail Making Test				
Part A (s)	43.50 (37.25, 53.00)	41.50 (31.50, 46.75)	0.82	0.56–0.93
Part B (s)	87.00 (75.25, 113.50)	76.50 (59.00, 96.50)	0.93	0.77–0.98
Chinses version of the Stroop Color and Word Test				
Congruous (number)	76.00 (70.00, 93.25)	80.00 (75.75, 89.50)	0.91	0.77–0.96
Incongruous (number)	30.00 (24.00, 36.00)	30.00 (24.75, 41.50)	0.91	0.76–0.97

Data are presented as the median (percentiles of 25, percentiles of 75).

congruous condition, and C-SCWT with incongruous condition. Moderate to excellent (ICC $\geq$ 0.50) was found for the TMT-A.

4. Discussion

This study established the alternate form reliability of the Chinese version of the TMT-B and the test-retest reliability of the Chinese version of the TMT and SCWT. Our results showed that the C-TMT-B had an acceptable alternate form reliability and both the C-TMT and C-SCWT had stable test-retest reliabilities.

Although our participants were familiar with English alphabet, it took them significantly longer to complete the standard TMT-B comparing with the C-TMT-B. Consistently, previous studies found that those non-native English-speakers performed the original version of the TMT-B leading to a poor performance.[20–22] Therefore, linguistic and cultural factors may not be overlooked for neuropsychological measurements. Considering linguistic and cultural factors, a sequence of the Chinese zodiacs was learned in childhood for native Chinese-speakers and may be a valid substitute to the letters in the standard TMT-B. Our results showed a high correlation between the C-TMT-B and the original version of the TMT-B. This finding supports the use of the C-TMT-B for native Chinese speakers.

Aging results in decline in executive function.[23,24] In this study, test-retest reliability of two measures of executive function was evaluated among older adults. Our results showed that both the C-TMT and C-SCWT achieved stable test-retest reliability in an interval of 3–7 days among older adults (ICC range: 0.82–0.93). Cangoz et al. examined the test-retest reliability of the Turkish version of the TMT over a 1-month interval in older adults (age range: 51–68 years) and found that the test-retest reliabilities of the TMT-A and TMT-B was 0.78 and 0.73 for score A and score B

with Pearson Correlation Coefficient, respectively.[8] Lemay et al. examined the test-retest reliability of the French version of the SCWT with an inter-assessment interval of 14 days in older adults (age range: 52–80 years) and found that the test-retest reliabilities (ICC) of the SCWT ranged from 0.48 to 0.80.[10] These findings indicate that different language versions of the TMT and SCWT demonstrates moderate to excellent reliability for clinical use among older adults.

The retest interval is one of factors that can influence test–retest reliability estimates. Shorter retest intervals may produce practice effect especially for tests with higher cognitive demands.[25,26] Separating practice effect from true change is critical in the interpretation of repeated assessment data. In the current study, a significant improvement was observed for the C-TMT-B upon retest. No significant differences were found between the first test and the second test for the TMT-A and C-SCWT. The practice effect of the TMT had also been found in a previous study.[27] Therefore, the unexpected finding of improvement across time should be taken into consideration when use the C-TMT-B as a measurement tool. To determine the practice effects at very brief test-retest intervals, Collie et al. examined the test–retest reliability on 4 occasions in 1 day.[28] They found that practice effects were evident between the first and second assessment, as performance remained more stable between the second, third and fourth assessments.[28] A single familiarization session may be useful in attenuating practice effects.[10]

There are several limitations in this study. First, the sample size was small. Studies with sample size more than 50 participants would be preferable and provide more conclusive evidence.[29,30] Second, the participants in this study were recruited form local community. The results might not be representative of the Chinese population in other regions. Third, our finding may not generalize to different test-retest intervals because retest interval may influence magnitude of performance changes as well as test-retest reliability estimated.

In summary, the current study suggests the Chinese version of the TMT-B is a reliable alternate instrument from the original version and both the C-TMT and C-SCWT have good to excellent test-retest reliabilities among older adults. Our findings support these two instruments for assessing executive function of older adults who are native Chinese speakers.

Conflicts of interest

There are no conflicts of interest.

Acknowledgments

The authors acknowledge the National Science Council (NSC100-2314-B-010-021-MY2) for supporting this work.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ijge.2018.06.003.

References

1. Colcombe SJ, Kramer AF, Erickson KI, et al. The implications of cortical recruitment and brain morphology for individual differences in inhibitory function in aging humans. *Psychol Aging*. 2005;20:363–375.
2. Smith-Ray RL, Makowski-Woidan B, Hughes SL. A randomized trial to measure the impact of a community-based cognitive training intervention on balance and gait in cognitively intact black older adults. *Health Educ Behav*. 2014;41:62s–69s.
3. Chan RCK, Shum D, Toulopoulou T, et al. Assessment of executive functions: review of instruments and identification of critical issues. *Arch Clin Neuropsychol*. 2008;23:201–216.
4. Faria CA, Alves HVD, Charchat-Fichman H. The most frequently used tests for assessing executive functions in aging. *Dement Neuropsychol*. 2015;9:149–155.
5. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. *Arch Clin Neuropsychol*. 2004;19:203–214.
6. Scarpina F, Tagini S. The Stroop color and word test. *Front Psychol*. 2017;8:557.
7. Sanchez-Cubillo I, Perianez JA, Adrover-Roig D, et al. Construct validity of the Trail Making Test: role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J Int Neuropsychol Soc*. 2009;15:438–450.
8. Cangoz B, Karakoc E, Selekler K. Trail Making Test: normative data for Turkish elderly population by age, sex and education. *J Neurol Sci*. 2009;283:73–78.
9. Kim TY, Kim S, Sohn JE, et al. Development of the Korean Stroop Test and study of the validity and the reliability. *J Korean Geriatr Soc*. 2004;8:233–240.
10. Lemay S, Bedard MA, Rouleau I, et al. Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *Clin Neuropsychol*. 2004;18:284–302.
11. Lu L, Bigler ED. Performance on original and a Chinese version of Trail Making Test part B: a normative bilingual sample. *Appl Neuropsychol*. 2000;7:243–246.
12. Lu L, Bigler ED. Normative data on Trail Making Test for neurologically normal, Chinese-speaking adults. *Appl Neuropsychol*. 2002;9:219–225.
13. Law LLF, Barnett F, Yau MK, et al. Effects of functional tasks exercise on older adults with cognitive impairment at risk of Alzheimer's disease: a randomised controlled trial. *Age Ageing*. 2014;43:813–820.
14. Lu JC, Guo QH, Hong Z, et al. Trail Making Test used by Chinese elderly patients with mild cognitive impairment and mild Alzheimer' dementia. *Chin J Clin Psychol*. 2006;14:118–120.
15. Wei M, Shi J, Li T, et al. Diagnostic accuracy of the Chinese version of the Trail-Making Test for screening cognitive impairment. *J Am Geriatr Soc*. 2018;66:92–99.
16. Chuang HH, Wang YY, Ku YC, et al. Executive dysfunction in old men with cardiovascular diseases and major depression. *J Evid Based Nurs*. 2008;4:118–126.
17. Guo Q, Hong Z, Lu C, et al. Application of Stroop Color-Word Test on Chinese elderly patients with mild cognitive impairment and mild Alzheimer's dementia. *Chin J Neuromed*. 2005;4:701–704.
18. Feng H, Li G, Xu C, et al. Training rehabilitation as an effective treatment for patients with vascular cognitive impairment with no dementia. *Rehabil Nurs*. 2017;42:290–297.
19. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155–163.
20. Arnold BR, Montgomery GT, Castañeda I, et al. Acculturation and performance of hispanics on selected halstead-reitan neuropsychological tests. *Assessment*. 1994;1:239–248.
21. Kim HJ, Baek MJ, Kim S. Alternative type of the Trail making test in nonnative English-speakers: the Trail making test-black & white. *PLoS One*. 2014;9,e89078.
22. Avila JF, Verney SP, Kauzor K, et al. Normative data for Farsi-speaking Iranians in the United States on measures of executive functioning. *Appl Neuropsychol Adult*. 2018:1–7.
23. Salthouse TA, Atkinson TM, Berish DE. Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J Exp Psychol Gen*. 2003;132:566–594.
24. Kirova AM, Bays RB, Lagalwar S. Working memory and executive function decline across normal aging, mild cognitive impairment, and Alzheimer's disease. *BioMed Res Int*. 2015;2015:748212.
25. Salthouse TA, Tucker-Drob EM. Implications of short-term retest effects for the interpretation of longitudinal change. *Neuropsychology*. 2008;22:800–811.
26. Vincent AS, Roebuck-Spencer TM, Fuenzalida E, et al. Test-retest reliability and practice effects for the ANAM General Neuropsychological Screening battery. *Clin Neuropsychol*. 2018;32:479–494.
27. Palmer CE, Langbehn D, Tabrizi SJ, et al. Test–retest reliability of measures commonly used to measure striatal dysfunction across multiple testing sessions: a longitudinal study. *Front Psychol*. 2018;8:2363.
28. Collie A, Maruff P, Darby DG, et al. The effects of practice on the cognitive test performance of neurologically normal individuals assessed at brief test-retest intervals. *J Int Neuropsychol Soc*. 2003;9:419–428.
29. Atkinson G, Nevill A. Typical error versus limits of agreement. *Sports Med*. 2000;30:375–381.
30. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med*. 2000;30:1–15.