Original Article

The Standardization of Victoria Stroop Color-Word Test among Iranian Bilingual Adolescents

Ayyoub Malek MD•[1], Issa Hekmati MSc[2], Shahrokh Amiri MD[3], Jaber Pirzadeh MSc[4], Hossein Gholizadeh MSc[5]

Abstract

Background: The Stroop Color-Word Test is a classic instrument for the assessment of selective attention and inhibition control and is a highly utilized instrument in research aspects of executive functions of the brain. The purpose of the present study was a preliminary standardization of Stroop test among Iranian bilingual adolescents.

Methods: In this study, 150 subjects, including three groups of adolescents (12 – 13, 14 – 15, and 16 – 17-year-olds) were randomly selected. Also, 30 subjects with attention-deficit hyperactivity disorder (ADHD) were selected for the study of differential validity. The instruments of this study were Victoria Stroop Color- Word Test and Strengths and Difficulties Questionnaire (SDQ).

Results: Correlation coeffecients by test-retest in Stroop test for reaction times of three cards and reaction time interference were 0.86, 0.86, 0.93, and 0.64; and for errors of three cards and error interference were 0.67, 0.37, 0.81, and 0.75 respectively. All of the corellations were significant. Differential validity by comparing ADHD and control group showed that there were significant differences among groups. The results of age effect in Stroop variables showed that there were significant differences between age groups in reaction time of all cards and error of the first and third cards; but sex did not show any significant effect on Stroop variables.

Conclusion: The findings suggest that reaction time and error of almost all cards of Stroop test are reliable and also have a good differential validity to discriminate ADHD from healthy controls in Iranian bilingual adolescents. Based on our findings, age but not sex is influential on performance of Stroop test.

Keywords: Adolescent, bilingual, reliability, Stroop test, validity

Cite this article as: Malek A, Hekmati I, Amiri S, Pirzadeh J, Gholizadeh H. The standardization of Victoria Stroop Color-Word Test among Iranian bilingual adolescents. *Arch Iran Med.* 2013; **16(7):** 380 – 384.

Introduction

The Stroop Color-Word Test was initially developed by John Ridley Stroop in 1935 for the evaluation of interference effect in sequential verbal reactions.[1] The Stroop test is a classic instrument for the assessment of selective attention and it evaluates a construction of executive function that is named "inhibition control", an item which is also considered in the Wisconsin Card Sorting Test and the Verbal Fluency Test.[2,3]

The feasibility and diagnostic importance of Stroop test, especially in the assessment of selective attention and inhibition control, has made this test to be a highly utilized instrument in diagnostic and research aspects of executive functions.[1]

Due to the significant sensitivity of the Stroop test to effects of damage or disturbance of the frontal lobe (executive functions) and attention deficits, it is frequently used for the evaluation of reactivity scales in this region of the brain. On the whole, the evidence related to children, adolescents, and adults shows that the Stroop test is a valid scale for the assessment of executive function and selective attention in population with neurologic deficits. According to the studies, evaluating the function of adults in the Stroop test in different clinical diseases such as epilepsy, multiple sclerosis, Parkinson, brain structural damage, and especially schizophrenia, it is revealed that these disorders result in an abnormal function compared with normal population.[4]

The Stroop Color-Word Test, the Victoria version called Victoria Stroop Test (VST) developed by Spreen and Strauss (1998), is a brief version of the Stroop task. VST has a short administration time (around five minutes), and is a brief, easily administered, and psychometrically sound version of Stroop's original task.[5,6] Also, the VST includes an additional training task (i.e., naming the colors of neutral words) that helps examinees establish the appropriate response set (i.e., color naming) without exposure to the interference condition.[6]

Regarding the verbal status of the test and the importance of the words utterance speed in it, using the norms of other languages, may be problematic because of this matter that even in different accents of a language, the speed of utterance of the words is different. Therefore, using the norms of other languages may result in a decreased psychometric properties and diagnostic value of this neuropsychologic test. Hence, it is necessary that this instrument be standardized among different groups of Iranian accents to increase the accuracy of diagnostic and investigative purposes and use it as a practical instrument to distinguish normal and abnormal subjects.

The original version of the Stroop test is translated to different languages and it has been a matter of debate about this test.[7] Rosselli, et al. evaluated English and Spanish monolingual subjects

Authors' affiliations: [1]Clinical Psychiatry Research Center (CPRC), Tabriz University of Medical Sciences, Tabriz, Iran, [2]Department of Psychology, Tabriz University, Tabriz, Iran, [3]Clinical Psychiatry Research Center (CPRC), Tabriz University of Medical Sciences- Tabriz- Iran, [4]Department of Psychology, Tarbiat Moallem University, Tehran, Iran, [5] Department of Counselling Psychology, Allame Tabatabaei University, Tehran, Iran.
•Corresponding author and reprints: Ayyoub Malek MD, Department of Psychiatry, Razi Mental Hospital, El Goli Boulevard, Tabriz, Iran, P. O. Box: 5456. Tel: 0098-914-314-4720, Fax: 0098-411-3803353,
E-mail: maleka@tbzmed.ac.ir.
Accepted for publication: 11 May 2013

and compared with Spanish-English bilingual subjects in order to study the effect of language in the Stroop test.[8] They found no significant differences in Stroop scores between the two groups except for color naming performance so that the bilinguals were slower than monolinguals. Also, Moering, et al. found the significant effect of age, education, sex, and the interaction of education with sex in Stroop performance in their study on African-American adults aged 60 to 84 years.[9]

Despite recent studies on the Stroop interference effect, the role of intelligence, age, sex, and language is not so clear because of small sample size in most studies and also different versions of the test used in them. The various versions of the test would emphasize the need for the repetition of the results of the studies about the effect of different variables on the test.

So, the present research was aimed to study the Persian (Farsi) version of the Stroop test among the bilingual Turkish-Persian adolescents in Tabriz, north-west of Iran to determine its validity and reliability and investigate the normalized scores for this age group.

Materials and Methods

The target population included the bilingual citizens in Tabriz, north-west of Iran, aged 12 to 17 years.The total sample size was 200 students, 50 of whom were excluded due to higher scores in Strengths and Difficulties Questionnaire (SDQ) subscales and failure to complete the tasks; then the sample size consisted of 150 subjects in three age groups (12 – 13, 14 – 15, and 16 – 17 years). This age categorization was due to assessing the impact of age variations on the Stroop task and based on different developmental characteristics of adolescents in early, middle, and late adolescence. The participants were selected by random cluster sampling in the first stage and categorical random sampling in the final stage. According to this sampling method, three of five districts were randomly selected and two guidance schools (for males and females) and two high schools (for males and females) were chosen in each district. In fact, 25 students from guidance schools (males and females) and 25 students from high schools (males and females) (a total of 50 students in each district) were selected finally and the Stroop test was administered. Since the presnece of any psychiatric disorder might affect the ability of subjects in the Stroop test, the adolescents were screened for psychiatric disorders by SDQ and those with scores higher than cut-off points in each subscale were excluded. Also, the history of neurologic disorders such as brain injury, epilepsy, head trauma, as well as the drug abuse were considered as exclusion criteria.

Also, 38 adolescents from among the excluded adolescents who obtained higher scores in hyperactivity subscale of SDQ were selected and evaluated to determine the discriminative validity. They were interviewd by a psychologist to establish the clinical criteria based on DSM-IV and as a result 30 of them met the criteria (11 females and 19 males, mean age: 14.6 years). In order to determine the test-retest reliability, after two weeks a second round of evaluation with the same measure was undertaken among 30 subjects who were randomly selected from among previous subjects.

Measures

Stroop test

The Stroop test, developed by Spreen and Strauss[10] (named as VST), consists of three cards, 21.5 cm x 14 cm, that were presented to the subjects with this order: Card D (Dot), Card W (Word), and Card C (Color). Each card has six rows and four columns. Card D includes color dots (red, green, blue, and yellow) and the participants were asked to name the colors of dots with a maximum speed possible. Card W includes words: “this”, “water”, “to”, and “up” in Farsi that was printed with red, green, blue, and yellow colors and the participants should rapidly name the color of the words. Card C includes color-words and the color of the words is red, yellow, blue, and green so that the color- words are printed with incongruent color of the name (for example, the yellow word is printed with green color) and the participants should rapidly name the color of the words. The naming order in all cards is in length of the rows and from left to right. The reaction time and the number of errors in each card were recorded for each participant and the time difference and error between Cards C and D were calculated.

Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a screening tool for the diagnosis of emotional and behavioral disorders in children and adolescents and includes 25 items and evaluates five main subgroups of psychiatric symptoms i.e., conduct problems, hyperactivity, emotional symptoms, peer problems, and prosocial behavior. This questionnaire has an impact score showing the severity of problems to indicate whether the severity of problems has resulted in daily and family life dysfunction.[11] This questionnaire has three versions of parent, teacher, and self-report; in this study the parent version was used. Tehranidoost, et al. showed good psychometric properties for parent and teacher versions in the evaluation of Iranian chidren’s psychiatric disorders.[12]

Procedure

One hundred and fifty subjects in age groups of 12 – 13, 14 – 15, and 16 – 17 years and among districts 1, 3, and 5 were selected and SDQ and the Stroop Color-Word Test were administered among them. Each age group consisted of 50 subjects (25 males and 25 females). To analyze the data, the Pearson correlation coefficient (for test-retest relaibility), independent-sample T (for discriminative validity), and one-way MANOVA (for the comparison of normalized variables in different age groups and gender) were used.

Results

The test-retest correlation coefficient of the Stroop test are shown in Table 1.

According to Table 1, the corrrelation coefficient of all variables of the Stroop test were significant and this showed reliability of this test among the participants. The highest and the lowest corrrelation coefficients were related to Card 3 reaction time and Card 2 error with 0.93 and 0.37, respectively. The results of discriminative validity of Stroop test are shown in Table 2.

According to this table, there was a significant difference in reaction time, error rate of any card, reaction time and error interference between healthy and ADHD adolescents. According to the mean of the groups, the differences showed that ADHD children have poor performance in reaction time and error.

For the comparison of normalized indices of the Stroop test as dependent variables among different age groups, we used MANOVA. Regarding the significance level of the Wilks Lambda

Table 1. The test-retest correlation coefficient of the Stroop test among sample

Card 1 reaction time	Card 2 reaction time	Card 3 reaction time	Reaction time interference	Card 1 error	Card 2 error	Card 3 error	Error interference
0.86**	0.86**	0.93**	0.64**	0.67**	0.37*	0.81**	0.75**

*P < 0.05; **P < 0.001.

Table 2. The results of independent-sample T- test of the Stroop test in different groups

	Descriptive results		T- test for equality of means		
	Mean	SD	Significance level	df	t
Card 1 reaction time			0.001	58	8.13
Normal	16.73	5.01			
ADHD	30.38	7.70			
Card 2 reaction time			0.001	58	7.61
Normal	23.65	5.38			
ADHD	37.36	8.26			
Card 3 reaction time			0.001	58	8.65
Normal	36.35	6.22			
ADHD	54.84	9.91			
Reaction time interference			0.002	58	3.33
Normal	19.63	3.53			
ADHD	24.60	7.38			
Card 1 error			0.008	58	2.76
Normal	0.06	0.25			
ADHD	0.50	0.82			
Card 2 error			0.003	58	3.05
Normal	0.7	0.83			
ADHD	1.5	1.16			
Card 3 error			0.001	58	5.88
Normal	1.93	1.41			
ADHD	4.40	1.81			
Error interference			0.001	58	4.57
Normal	1.86	1.40			
ADHD	3.90	1.98			

Table 3. The results of between-subjects effects of the Stroop test in study groups

Source of changes	Dependent variable	Sum of squares	df	Mean square	F
Group	Card 1 reaction time	813.79	2	406.89	92.53**
	Card 2 reaction time	938.39	2	469.19	95.57**
	Card 3 reaction time	1318.54	2	659.27	106.81**
	Reaction time interference	160.01	2	80	15.88**
	Card 1 error	0.49	2	0.24	3.18*
	Card 2 error	0.52	2	0.26	0.76
	Card 3 error	13.08	2	6.54	4.18*
	Error interference	10.61	2	5.30	3.44*
Error	Card 1 reaction time	646.29	147	4.39	
	Card 2 reaction time	721.62	147	4.90	
	Card 3 reaction time	907.29	147	6.17	
	Reaction time interference	704.40	147	5.03	
	Card 1 error	11.38	147	0.07	
	Card 2 error	49.72	147	0.33	
	Card 3 error	229.78	147	1.56	
	Error interfencree	226.16	147	1.53	

* P < 0.05; ** P < 0.001.

in the analysis of variance (F = 13.75, P < 0.001), the general effect of groups on indices is determined. So, the results of between-subjects effects of the Stroop test among study groups are shown in Table 3. According to this table, there is a significant difference among groups for all variables except Card 2 error.

Post-hoc comparison (Tukey) showed that there were significant differences between third group (16-17- year- olds) and two other groups in the first and second cards reaction time, while in the third card reaction time, there was a signicicant difference between three age groups. And finally, in reaction time subtraction, the first group (12 – 13-year-olds) had a significant difference with two other groups (14 – 15 and 16 – 17-year-olds). It means that the third group has spent less time for responding in Cards 1 and 2; but in Card 3, the third, second, and the first groups had least to the most response time. About reaction time interference, the second and the third groups had spent less time compared with the first group.

About error rate in Card 1, the second and third groups were significantly different and in the Card 3 error and also error subtraction, the difference was significant between the first and third

groups. It means that in Card 1, the third group had fewer errors compared with second group, but in Card 3 error and error subtraction, the third group had done fewer errors only compared with the first group.

For the comparison of normalized indices of the Stroop test among males and females, MANOVA was used. Since the Wilks Lambda in the analysis of variance was not significant ($F = 1.51$, $P > 0.05$), it may be concluded that the gender had no effect on Stroop indices.

Discussion

The aim of the present study was a preliminary standardization of Victoria Stroop Color-Word Test among Iranian bilingual Turkish-Persian adolescents. The results showed that test-retest correlation coefficient of this test was high for almost all of cards, and it showed the good reliability of the Stroop test among Turkish-Persian bilingual adolescents. However, to date there has been no study to evaluate the psychometric properties of this test among bilingual subjects, but our findings are in congruence with those reports about its good reliability among children and adolescents. For example, Graf, et al., Sack, et al., and Feinstein, et al. reported a reliability higher than 0/80 for scores obtained in Stroop Color-Word Test.[13–15] However, Ludvig, et al. who used the modified version of the Stroop test among adults, showed that only the time indices had a good reliability but there was no good reliability for the error rate.[16]

For the assessment of validity of the Stroop test, different methods have been used among which the most common one is differential validity focusing more on ADHD and learning disorders.[5] The differential validity in this study showed that the Stroop test may differentiate ADHD children from healthy ones and then this test has a good differential validity. This finding is in congruence with studies that show ADHD subjects have poor executive function and are subsequently poor in the Stroop test performance. Lufi, et al. showed that ADHD compared with emotional disorders had a negative impact on color-word and interference performance.[17] In other words, the children with emotional disorders had a better performance compared with ADHD children. Lavoie and Charlebois similarly reported that the children with ADHD had poorer performance in the Stroop test indices compared with healthy children.[18] Also, MacLeod and Proir demonstrated that interference index scores may differentiate intermediatly between ADHD and conduct disorder among adolescents.[19] Totally, these studies show that the Stroop test is intermediately sensitive to neuropsychologic deficits observed in subjects with ADHD. The previous studies have reported that subjects with ADHD have deficits in executive functions.[20,21] These subjects often show problems in concentration, attention, and response inhibition and the magnetic resonance imaging (MRI) studies have revealed that these subjects' brain, in the frontal lobe, is not normaly assymetried and the right hemispher is greater than the left hemispher.[7]

The results of our study also showed that the age is effective in the performance of bilingual adolescents in the Stroop test so that the age group of 16 – 17-year-olds have less reaction time and error in most of indices. Of course, it was more significant about reaction times compared with errors, and the errors in Card 2 showed no difference. As mentioned by Ludvig, et al.[16] these findings may show that in the Stroop test the reaction time is more appropriate index compared with error rate. Although the previous studies have reported the role of age in the Stroop test,[7] most of these studies have been performed among adults. The life span pattern effect on the Stroop test related to inhibition control, is often seen as an increasing one during childhood[22] and decreasing at the end of adulthood.[23] According to the findings of the present study, it may be concluded that the inhibition control has also an increasing trend in adolescence.

The results of this study showed that the gender of the subjects had no effect on participants' performance in the Stroop test, which is in contrast to those reported by previous studies.[9] Also, Mitrushina, et al. reported a controversy in effect of gender in the Stroop test performance.[1] Moering, et al. mentioned that the small sample size for gender differences analysis in Stroop studies may induce invalid results.[9] Accordingly, further studies should be carried out to determine the exact impact of gender on the Stroop test.

Acknowledgment

This study was undertaken by the financial support of the Research Office of Tabriz University of Medical Sciences.

References

1. Mitrushina M, Boone KB, Razani J, D'Elia LF. *Handbook of Normative Data for Neuropsychological Assessment.* 2nd Ed, New York: Oxford University Press, Inc, 2005; 389 – 394.
2. Barch DM, Carter CS, Hachten PC, Usher M, Cohen JD. The "benefits" of distractibility: mechanisms underlying increased Stroop effects in schizophernia. *Journal of Schizophrenia Bulletin.* 1999; **25:** 749 – 762.
3. Kosmidis MH, Bozicas VP, Zafiri M, Karavatos A. Shared cognitive processes underlying performance on the Wisconsin Card Sorting Test and the Stroop Test in patients with schizophrenia: A measurement artifact? *Neuroscience Letters.* 2006; **409:** 234 – 238.
4. Marnat GG. *Neuropsychological Assessment in Clinical Practice: a guide to test interpretation and integration*. New York: John Willy & Sons, Inc, 2000; 426 – 452.
5. Bayard S, Erkes J, Moroni C. Victoria Stroop Test: normative data in a sample group of older people and the study of their clinical applications in the assessment of inhibition in Alzheimer's Disease. *Archives of Clinical Neuropsychology*. 2011; **26:** 653 – 661.
6. Troyer AK, Leach L, Strauss S. Aging and response inhibition: Normative data for the Victoria Stroop Test. *Aging, Neuropsychology, and Cognition*. 2006; **13:** 20 – 35.
7. Homack S, Riccio CA. A meta-analysis of the sensivity and specificity of the Stroop Color-Word Test with children. *Archive of Clinical Neuropsychology.* 2004; **19:** 725 – 743.
8. Rosselli M, Ardila A, Santisi MN, Arecco MR, Salvatierra J, Conde A, et al. Stroop effect in Spanish- English bilinguals. *Journal of the International Neuropsychology Society.* 2002; **8:** 819 – 827.
9. Moering RG, Schinka JA, Mortimer JA, Graves AB. Normative data for elderly African- Americans for the Stroop Color and Word Test. *Archives of Clinical Neuropsychology.* 2004; **19:** 61 – 71.
10. Spreen O, Strauss E. A Compendium of Neuropsychological Tests. 2nd Ed. New York: Oxford University Press, Inc; 1998; 477 – 490.
11. Goodman R. The Strengthss and Dificulties Questionnare. *Journal of Child Psychology and Psychiatry.* 1997; **38:** 581 – 586.
12. Tehranidoost M, Shahrivar Z, Pakbaz B, Rezaei A, Ahmadi F. Validity of Persian version of Strengthss and Dificulties Questionnare (SDQ). *Iranian Journal* of *Advances in Cognitive Science.* 2006; **8:** 33 – 39.
13. Graf P, Utte B, Tuokko H. Color and picture-word Stroop tests: Performance changes in old age. *Journal of Clinical and Experimental Neuropsychology.* 1995; **17:** 390 – 415.
14. Sacks TL, Clar CR, Pols RG, Geffen LB. Comparibility and stability of performance of six alternate forms of the Dodrill-Stroop color-word test. *The Clinical Neuropsychologist.* 1991; **5:** 220 – 225.
15. Feinstein A, Brown R, Ron M. Effects of practice of serial test of attention in healthy subjects. *Journal of Clinical and Experimental Neuropsycholoy.* 1994; **16:** 436 – 447.
16. Ludvig C, Borella E, Tettamanti M, De Ribaupierre A. Adult age differences in the Stroop test: A comparison between item-by-item

and blocked version. *Archive of Gerentology and Geriatrics.* 2009; doi:10.1016/j.archger.2009.09.040.
17. Lufi D, Cohen A, Parsh-Plass J. Identifying attention deficit hyperactivity with WISC-R and the Stroop color and word test. *Psychology in Schools.* 1990; **27:** 28 – 34.
18. Lavoie ME, Charlebois P. The discriminant validity of Stroop color and word test: Toward a cost-effective strategy to distinguish subgroups of disruptive preadolescents. *Psychology in Schools.* 1994; **31:** 98 – 107.
19. MacLeod CM, Prior M. Attention deficit in adolescents with ADHD and other clinical groups. *Child Neuropsychology.* 1996; **2:** 1 – 10.
20. Kerns KA, Mc Inerney RJ, Wild NJ. Time reproduction, working memory, and behavioral inhibition in children with ADHD. *Child Neuropsychology.* 2001; **7:** 21 – 31.
21. Helliman KM, Voeller KK, Nadeau SE. A possible patho-physiological subtract of attention deficit hyperactivity disorder. *Journal of Child Neurology.* 1991; **6:** 76 – 81.
22. Harnishfeger KK. *The development of cognitive inhibition: theories, definitions, and research evidence*. In: Dempster FN, Brainerd CJ. (Eds). New Perspectives on Interference and Inhibition in Cognition. San Diego: Academic Press, 1995; 176 – 199.
23. Hasher L, Lustig C, Zacks RT. *Inhibitory mechanisms and the control of attention.* In: Conway AR, Jarrold C, Kane MJ, Miyake A, Towse JN. (Eds.). Variation in Working Memory. New York: Oxford University Press, 2007; 227 –240.