ORIGINAL

Developmental Changes in Stroop Task Performance among Individuals with High-Functioning Pervasive Developmental Disorders

Ruriko Yamashita[1], Yukie Iwasa[1], and Toshiaki Hashimoto[2]

[1]*Institute of Health Biosciences, The University of Tokushima Graduate School, Tokushima, Japan,* [2]*Japanese Red Cross Tokushima Hinomine Rehabilitation Center for People with Disabilities, Tokushima, Japan*

Abstract : Objective : This study aimed to assess age-related changes in Stroop Color and Word test indices in individuals with high-functioning pervasive developmental disorder (HF-PDD) and compare their performance with typically developing (TD) individuals. Methods : There were a total of 125 participants (57 TD, 68 HF-PDD), aged 6-24. Stimuli were displayed on computer screens, and participants verbally responded with response times recorded via voice key function. Results : Single regression analysis revealed age-associated trends in Stroop test indices for both groups, indicating shorter response times and reduced Stroop interference with age. The age at which the best Stroop test results were obtained ranged from 220 to 260 months for TD and from 195 to 201 months for HF-PDD. While color naming and word reading tasks showed no significant group differences, color word naming task response times were significantly longer for the HF-PDD group. Around 30% of HF-PDD participants scored over two standard deviations above the TD mean. Conclusions : Both TD and HF-PDD groups exhibited age-related changes in Stroop task performance, which were fitted by a quadratic regression curve. Prolongation of the color word naming task in the HF-PDD group suggests that approximately 30% of individuals with HF-PDD have difficulty with stereotype suppression. J. Med. Invest. 71 : 254-259, August, 2024

Keywords **:** *Stroop test, high-functioning pervasive developmental disorder, frontal lobe function, stereotype suppression*

INTRODUCTION

The Stroop effect, first reported by Stroop (1) in 1935, is a phenomenon whereby conflict between words and colors results in increased response latency. In practice, it comprises a cognitive conflict whereby when the word for a color is presented in lettering of a different color, a longer time is required to name the color than when both are congruent. This effect is associated with difficulty in stereotype suppression, and is also recognized in the concept of impaired mental set shifting (2). Clinically, it is used as a test of frontal lobe function, but it does not constitute a diagnostic indicator.

The Stroop test consists of three sets of cards : word cards with the names of colors written in black lettering, color cards with arrays of colored ink, and color word cards with the names of colors printed in inks of different colors. Response time increases in the order, word cards < color cards < color word cards.

Although it is well known that the Stroop test reveals individual differences, response times to the respective cards are also known to be delayed in schizophrenia (3, 4). Comalli *et al*. (5) reported the existence of age differences, and Hirasawa *et al*. (6) attempted to produce standard values for different age groups.

There has as yet been insufficient research, however, on how this effect changes in pervasive developmental disorder (PDD). Although PDD is regarded as a frontal lobe disorder, it is still poorly understood, and Stroop test results may reveal new aspects of the disabilities involved in it. The aim of the present study was to elucidate how Stroop test indices change with age in high-functioning pervasive developmental disorder (HF-PDD), and how these differ from neurotypical development (TD). Participants' response times to stimuli displayed on a computer screen were measured by using a voice key function, which improved the accuracy of measurements at the millisecond level.

Received for publication April 18, 2024 ; accepted May 23, 2024.

Address correspondence and reprint requests to Dr. Yukie Iwasa, Institute of Health Biosciences, The University of Tokushima Graduated School, 3-18-15, Kuramoto-cho, Tokushima 770-8509, Japan and Fax : +81-88-633-9005. E-mail : iwasa@medsci.tokushima-u.ac.jp

METHODS

Participants

Participants included 125 (57 TD and 68 HF-PDD) individuals between the ages of 6 and 24 years. TD participants were defined as individuals who had no problems engaging in social activities and no history of neurological disorder. HF-PDD participants were patients attending outpatient pediatric neurology clinics. Of these, 30 were diagnosed with high-functioning autistic disorder (HFAD), 36 with Asperger's disorder (AS), and 2 with pervasive developmental disorder not otherwise specified (PDD-NOS). Thirteen had also been diagnosed with ADHD. PDD was diagnosed according to the DMS-IV-TR criteria. Characteristics (age, sex, diagnoses, and cognitive performance) of test participants are summarized in Table 1.

Participants were provided a written explanation of the study, which included the study objective and methods, informed individuals that their participation was voluntary, and affirmed they would not be placed at any disadvantage by declining. Consent was obtained after it had been confirmed that the participants fully understood the content. For minors, the content was explained in accordance with their level of comprehension, and consent was obtained from their parents or guardians. This study was approved by the Ethics Committee of Tokushima University Hospital.

Experimental procedure

A Stroop Color and Word Test, consisting of three tasks and

Table 1. Characteristics of test subjects

Age Group		Pre Puberty (6–10)	Early Puberty (11–15)	Late Puberty (16–20)	(21–24)
TD group					
	Male	9	7	12	5
	Female	4	7	11	2
	Total	13	14	23	7
Age (mean ± SD)		8.2 ± 1.4	12.9 ± 1.5	18.0 ± 1.4	22.0 ± 1.2
Cognitive ability			NA		
HF-PDD group					
	Male	21	29	10	1
	Female	3	3	1	0
	Total	24	32	11	1
Diagnosis	HFAD	7	16	7	0
	AS	17	15	3	1
	PDD-NOS	0	1	1	0
Age (mean ± SD)		8.2 ± 1.1	12.6 ± 1.4	17.3 ± 1.4	23
Cognitive ability index (mean ± SD)	All tests	91.9 ± 14.2	97.5 ± 19.6	88.5 ± 15.8	110
	Verbal	91.3 ± 14.2	99.2 ± 19.2	83.8 ± 17.2	103
	Perform.	93.3 ± 15.9	93.1 ± 21.0	92.0 ± 17.2	118

Note. All values represent numbers of individuals, unless otherwise stated.

four colors (red, blue, green, and yellow), was administered to test participants. The first task was a color-naming (C) task, in which participants were asked to name the colors of circles displayed in the four colors. The second task was a word reading (W) task, in which participants were asked to read the names of the four colors displayed in black *hiragana* Japanese syllables. The third was a color word naming (CW) task, in which the names of the four colors were presented in font of incongruent colors, and participants were asked to name the font colors in which the words were displayed. The expected order of response times for the three tasks was W < C < CW.

Tests were administered on a personal computer and participants were seated during the tests. Stimuli were displayed and response times measured using a sound/image stimulus system (MTS0400 MultiTrigger System, MedicalTrySystem Co. Ltd.). Stimuli appeared in the center of the computer screen, and participants responded by speaking the name of the words or colors. Participants were presented 30 stimuli for each task, and performance was assessed as the average response time for each task. Stimuli were displayed for 6000 ± 985 ms intervals, with each task taking 3 min to complete. Response times were measured using a voice key function and the number of incorrect responses was also recorded.

Analytical methods

The response time for each task (C, W, and CW), difference in response times between color word naming and color naming tasks (CW – C), ratio of response times for color word naming and color naming tasks (CW/C), and ratio of the difference in response times between color word naming and color naming tasks to word reading tasks ((CW – C)/W) were calculated. Regression analyses were performed for the TD and HF-PDD groups with Stroop test score as the dependent variable and age as the explanatory variable. Participants between the ages of 6 and 20, in each group (TD and HF-PDD), were divided by age into pre-puberty (6–10 years), early puberty (11–15 years), and late puberty (16–20 years) age groups. The mean responses of each group were calculated and differences between groups were tested using nonparametric, Mann-Whitney *U* tests. The proportions of participants in the TD and HF-PDD groups whose indices exceeded the mean TD score by more than two standard deviations were calculated, and Fisher's exact test was used to test for differences between these rates. SPSS Advanced Models 15.0 software was used for statistical analysis.

RESULTS

Age-dependent changes in Stroop test indices

The relationship between age and response time, for each task, of individuals in the TD group is shown in Figure 1, and age-related changes in interference effect indices in Figure 2. Regression analysis showed a significant correlation between Stroop test indices and age (C : $R^2 = 0.233$, $p = 0.001$; W : $R^2 = 0.177$, $p = 0.005$; CW : $R^2 = 0.347$, $p = 0.000$; CW – C : $R^2 = 0.204$, $p = 0.002$; CW/C : $R^2 = 0.118$, $p = 0.034$; (CW – C)/W : $R^2 = 0.130$, $p = 0.023$). Response time and Stroop interference decreased until teenage. After reaching the minimum values, response times and Stroop interference increased.

Changes by age in the response time for each task in

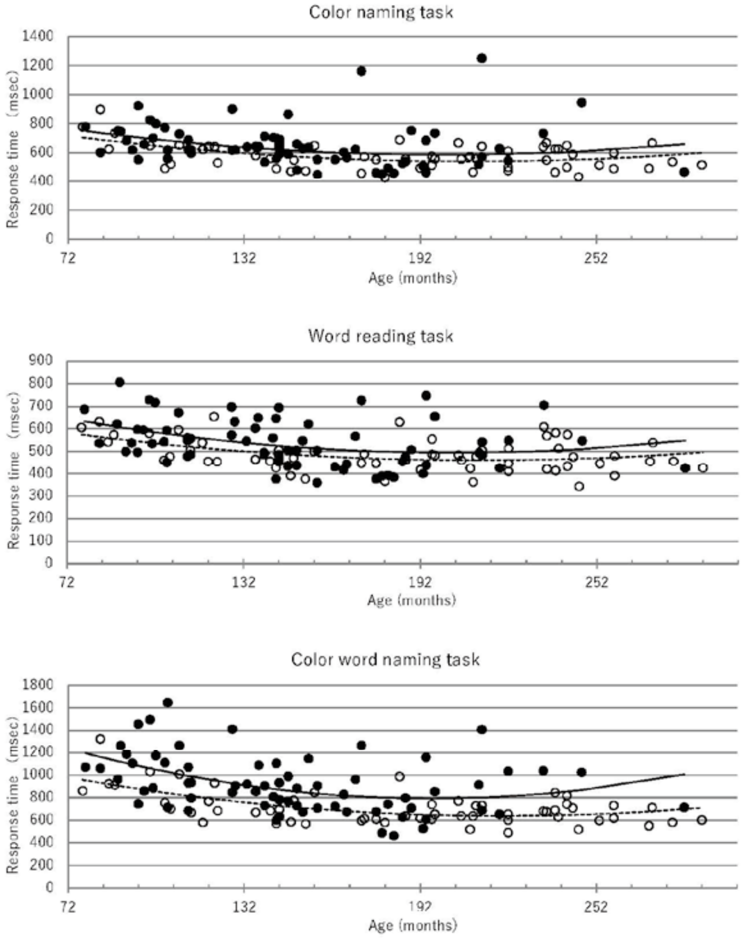

Figure 1. Age-dependent changes in Stroop test response times
Circles (○) show response times for TD children and adolescents. Circles (●) show response times for HF-PDD children and adolescents. Dotted lines (TD) and Solid lines (HF-PDD) show regression curves for response time with respect to age. All indices followed quadratic regression curves, and were significantly correlated with age.
TD : Color naming, $p = .001$; Word reading, $p = .005$; Color word naming, $p < .000$
HF-PDD : Color naming, $p = .004$; Word reading, $p = .008$; Color word naming, $p = .001$.

the HF-PDD group are shown in Figure 1, and age-related changes in interference effect indices in Figure 2. Regression analysis showed a significant correlation between Stroop test indices and age (C : $R^2 = 0.159$, $p = 0.004$; W : $R^2 = 0.139$, $p = 0.008$; CW : $R^2 = 0.196$, $p = 0.001$; CW – C : $R^2 = 0.189$, $p = 0001$; CW/C : $R^2 = 0.111$, $p = 0.001$; (CW – C)/W : $R^2 = 0.137$, $p = 0.022$), with the same developmental changes evident as in the TD group.

Table 2 shows regression formulae, coefficients of determination, levels of significance, and ages at which minimum Stroop interference was observed. These values were calculated from regression formulae, obtained from regression analysis. In regression formulae, Y represents interference effect score and X represents age. The age at which response time was shortest ranged from 209.2 to 260.0 months in the TD group and from 195.3 to 203.5 months in the HF-PDD group, somewhat younger in the HF-PDD group.

Comparison of TD and HF-PDD groups

Table 3 shows average response times and interference effect indices for each task analyzed by age group. Mann-Whitney *U* tests were used to test for differences in Stroop test indices between the TD and HF-PDD groups. Although there was no significant difference in the color naming or word reading tasks, response times for the color word naming task were significantly longer for the HF-PDD individuals in all age groups (pre-puberty, $p = .016$; early puberty, $p = .006$; late puberty, $p = .016$). CW – C and (CW – C)/W, indices of the Stroop interference effect, were significantly higher in the HF-PDD group than in the TD group during pre-puberty (CW – C, $p = .014$; (CW – C)/W, $p = .033$). CW – C and (CW – C)/W were also higher in the HF-PDD participants in other age groups, but these differences were not statistically significant. The number of incorrect responses was, likewise, higher in the HF-PDD group during pre-puberty and early puberty (pre-puberty, $p = .009$; early puberty, $p = .032$; late puberty, $p = .713$).

Table 4 shows the proportions of participants in the TD and HF-PDD groups whose indices exceeded the mean scores of the TD group by more than two standard deviations and the results of the Fisher's exact test. Around 30% of participants in the HF-PDD group exceeded the mean plus two standard deviations, and a comparison between the TD and HF-PDD groups showed that the HF-PDD group included a significantly larger number of participants who deviated from the norm (CW, $p = .001$; CW – C, $p < .001$; CW/C, $p < .001$; (CW – C)/W, $p < .001$). There was no particular bias with respect to type of diagnosis.

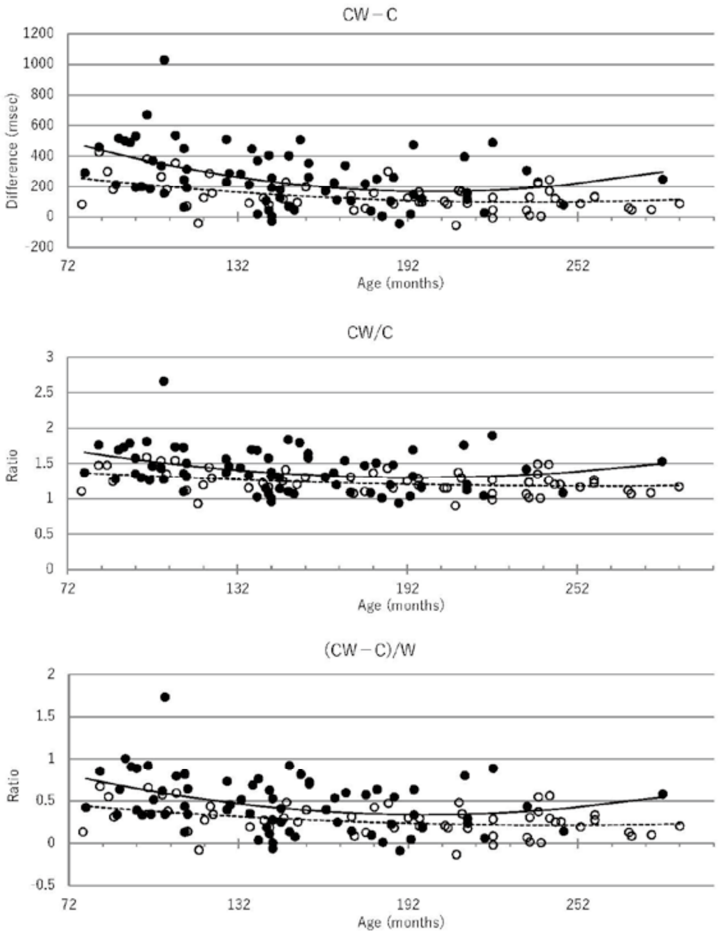

Figure 2. Age-dependent changes in Stroop interference indices
Circles (○) indicate differences between response times for color word naming tasks and color naming tasks (CW – C) for TD children and adolescents, and ratios of response times for color word naming and color naming tasks (CW/C), and the ratio of response times for word reading task in the differences in the response time between color word naming and color naming tasks ((CW-C)/W). Circles (●) show response for HF-PDD children and adolescents. Dotted lines (TD) and Solid lines (HF-PDD) show regression curves for each index with respect to age. All indices followed quadratic regression curves, and were significantly correlated with age.
TD : CW – C, $p = .002$; CW/C, $p = .034$; (CW – C)/W, $p = .023$.
HF-PDD : CW – C, $p = .001$; CW/C, $p = .022$; (CW – C)/W, $p = .008$.

Table 2. Age-related changes in Stroop test indices and best scores

		Regression formula	R^2	p value	Optimum Age (months)
TD group	C	$Y = 0.009X^2 - 3.938X + 951.610$	.233	.001	209.2
	W	$Y = 0.006X^2 - 2.649X + 741.295$	.177	.005	212.4
	CW	$Y = 0.016X^2 - 6.926X + 1397.838$	.347	< .001	219.5
	CW-C	$Y = 0.006X^2 - 2.988X + 446.213$	.204	.002	234.7
	CW/C	$Y = 0.000X^2 - 0.003X + 1.558$	.118	.034	260.0
	(CW-C)/W	$Y = 0.004X^2 - 0.000X + 0.695$	.130	.023	245.8
HF-PDD group	C	$Y = 0.011X^2 - 4.398X + 1025.790$	.159	.004	200.4
	W	$Y = 0.009X^2 - 3.615X + 861.333$	.139	.008	203.5
	CW	$Y = 0.029X^2 - 11.336X + 1903.789$	.196	.001	196.0
	CW-C	$Y = 0.020X^2 - 7.976X + 969.340$	.189	.001	201.2
	CW/C	$Y = 0.000X^2 - 0.010X + 2.2973$	.111	.001	195.3
	(CW-C)/W	$Y = 0.000X^2 - 0.012X + 1.505$	.137	.022	198.2

Note. In the regression equation, Y represents each index for interference effects, while X represents the age in months.

Table 3. Response times for Stroop tasks and Stroop effect indices

		TD group ($n = 50$)		HF-PDD group ($n = 67$)			
		Mean	*SD*	*Mean*	*SD*	*p* value	
Pre-puberty		(13)		(24)			
(6–10 years)	Age	8.23	1.42	8.17	1.13	.838	
	C (msec)	644.81	110.06	692.25	102.32	.324	
	W (msec)	544.35	69.55	589.66	90.26	.227	
	CW (msec)	857.71	197.49	1065.72	258.91	.016	*
	CW-C (msec)	212.89	137.10	373.47	209.59	.014	*
	CW/C	1.33	0.20	1.54	0.31	.052	
	(CW-C)/W	0.38	0.23	0.63	0.34	.033	*
	Incorrect responses	1.62	1.98	4.46	3.76	.009	**
Early puberty		(14)		(32)			
(11–15 years)	Age	12.93	1.54	12.63	1.41	.556	
	C (msec)	550.62	83.66	616.89	137.78	.115	
	W (msec)	459.06	65.49	497.75	96.37	.242	
	CW (msec)	673.05	121.57	805.05	184.80	.006	**
	CW-C (msec)	122.43	78.09	188.16	148.53	.189	
	CW/C	1.23	0.13	1.32	0.26	.340	
	(CW-C)/W	0.26	0.15	0.38	0.29	.252	
	Incorrect responses	1.00	1.41	2.66	2.60	.032	*
Late puberty		(23)		(11)			
(16–20 years)	Age	18.00	1.41	17.27	1.35	.158	
	C (msec)	562.09	71.88	689.21	231.92	.118	
	W (msec)	476.36	70.20	544.34	114.75	.127	
	CW (msec)	669.00	89.10	900.82	266.46	.016	*
	CW-C (msec)	106.91	77.22	211.60	174.10	.191	
	CW/C	1.20	0.15	1.34	0.31	.408	
	(CW-C)/W	0.23	0.18	0.37	0.29	.329	
	Incorrect responses	0.70	0.93	1.00	1.41	.713	

Note. Although there was no significant difference in the color naming or word reading tasks, response times in the color word naming task were significantly longer for the HF-PDD group in all age groups. CW – C and (CW – C)/W, which are indices of the Stroop interference effect, were significantly higher in the HF-PDD group compared with the TD group during pre-puberty, and were also higher in the HF-PDD group in other age groups, although these differences were not statistically significant. The number of incorrect responses was also higher in the HF-PDD group in pre-puberty and early puberty. Mann-Whitney test, *$p < 0.05$, **$p < 0.01$. Figures in parentheses are the numbers of subjects.

Table 4. Proportions of subjects whose indices exceeded the mean TD score by more than two standard deviations : Comparison of TD and HF-PDD groups

	TD group ($n = 50$)	HF-PDD group ($n = 67$)	*p* value
CW	3 (6.0%)	20 (29.9%)	.001
CW-C	1 (2.0%)	20 (29.9%)	< .001
CW/C	3 (6.0%)	23 (34.3%)	< .001
(CW-C)/W	2 (4.0%)	25 (37.3%)	< .001

Note. Approximately 30% of subjects in the HF-PDD group exceeded the mean of the TD group by more than two standard deviations. Furthermore, a comparison of the population rates between the TD and HF-PDD groups revealed a significantly higher proportion of subjects deviating from the norm in the HF-PDD group. Fisher's exact test.

DISCUSSION

Regression analysis of age-related changes in Stroop test indices showed developmental changes in all Stroop test indices, with the fastest response times and lowest interference observed between ages 17 and 21 in TD group. These results are consistent with previous studies. Commalli *et al.* (5) found that the best results for all tasks were achieved between the ages of 17 and 19 years, Hama *et al.* (7) between 19 and 22 years, and Hirasawa *et al.* (6) between 16 and 17 years. In both the HF-PDD and TD groups, the relationship between age and Stroop indices was modeled best by a quadratic regression curve. In HF-PDD group, the fastest responses and minimum interference of all Stroop test indices were observed at the slightly younger age, 16–17 years.

HF-PDD individuals required longer times for color word naming tasks but significant difference between the TD and HF-PDD groups were not observed for color naming and word reading tasks. This differs from the effects reported in patients with schizophrenia, in which response times are longer for all tasks (3, 4). The performance of the interference task in the Stroop test is thought to involve two functions : selective attention, which directs attention to the color of the printed letter of the color name word, and stereotype inhibition, which suppresses the habitual response of reading the letters while calling out the color of the printed letter of the color name word (8). Our result may reflect frontal lobe dysfunction in ASD (9, 10).

We also found that around 30% of children and adolescents in the HF-PDD group deviated from the norm. There was no particular bias with respect to diagnosis. Stroop interference is associated with stereotype suppression disability, and approximately 30% of individuals with HF-PDD may have difficulty in stereotype suppression.

Previous studies investigating the performance of ADHD and ASD individuals have reported equivocal results. In a study of developmental disability and Stroop interference, Ozonoff and Jensen (11) compared the results of Stroop tests taken by children with ADHD and HF-PDD. The authors reported that children with ADHD showed poorer performance than TD children, but they found no differences between TD individuals and those with autism. However, in follow-up experiments, Goldberg *et al.* (12) found no differences in Stroop task performance between ADHD, high-functioning autism, and TD children. In contrast, recent studies have shown impairment of Stroop interference among individuals with autistic spectrum disorders (ASDs) (13-15).

There are a couple of likely explanations for the discrepancy in Stroop test results among studies of ADHD and ASD individuals. One possible reason for these contradictory findings may be differences in the nature of test. Performance on the conventional, Victoria version of the test (16) (in which 24 stimuli are printed on a single card and response times measured) and Golden's version (17) (in which 100 stimuli are printed on a single card, and the number of responses in 45 s is counted) is likely affected by the fact that they include not only the response time for a single stimulus, but also the processing time involved when participants' eyes track (pick up) the next stimulus. The function of the frontal lobe, known as executive function, maintains appropriate sets for achieving future goals. This is a multidimensional act that includes functions such as planning, set-shifting, working memory, inhibition, and verbal fluency (18, 19). In conventional methods such as the Victoria and Goldman versions, problems with other elements of executive function may negatively affect the results even in the absence of difficulty with stereotype suppression. This problem is eliminated by computer-administration of tests.

CONFLICT OF INTEREST STATEMENT

All authors declare no conflict of interest.

REFERENCES

1. Stroop JR : Studies of interference in serial verbal reactions. J Exp Psychol 18 : 643-662, 1935
2. Kashima H : Frontal Lobe Syndrome : Frontal Lobe Symptoms as the Format of This Disability from the Perspective of Neuropsychological Examination. Neuropsychology and Psychiatry (in Japanese). Japan Scientific Societies Press, Tokyo, 1996, pp. 140-153
3. Wapner S, Kurus DM : Effects of lysergic acid diethylamide, and difference between normals and schizophrenics on the Stroop color word. J Neuropsychiatr 2 : 76-81, 1960
4. Hama H : Experimental Abnormal Psychology (in Japanese). Seishin Shobo, Tokyo, 1969, pp. 86-99
5. Comalli REJ, Wapner S, Werner H : Interference effects of Stroop color word test in childhood, adulthood, and aging. J Genet Psychol 100 : 45-53, 1962
6. Hirasawa T, Sanada T, Yanagihara M, Tsushima Y, Kado Y, Ogino T, Nakano K, Watanabe K, Ohtsuka Y : Standard value and developmental changes in the indices of interference effect in the modified Stroop test (in Japanese). No To Hattatsu 41 : 426-430, 2009
7. Hama H, Hashimoto E : Change in conflict strength measured by Color-Word Test in childhood, adulthood, and the aged (in Japanese). Jpn J Psychol 56 : 175-179, 1985
8. Hlavat P, Kašpárek T, Linhartová P, Ošlejšková H, Bares M : Autism, impulsivity and inhibition a review of the literature. Basal Ganglia 14 : 44-53, 2018
9. Sanders J, Johnson KA, Garavan H, Gill M, Gallagher L : A review of neuropsycho-logical and neuroimaging research in autistic spectrum disorders : Attention, inhibition and cognitive flexibility. Research in Autism spectrum disorders 2 : 1-16, 2008
10. Park HR, Lee JM, Moon HE, Lee DS, Kim BN, Kim J, Kim DG, Paek SH : A short review on the current understanding of autism spectrum disorders. Experimental Neurobiology 25(1) : 1-13, 2016
11. Ozonoff S, Jensen J : Specific executive function profiles in three neurodevelopmental disorders. J Autism Dev Disord 29 : 171-177, 1999
12. Goldberg MC, Mostofsky SH, Cutting LE, Mahone EM, Astor BC, Denckla MB, Landa RJ : Subtle executive impairment in children with autism and children with ADHD. J Autism Dev Disord 35 : 279-293, 2005
13. Saino H, Kawai T, Kurokawa S, Denda K : Characteristics of executive functioning in pervasive developmental disorder (in Japanese). Jpn J Child Adolesc Psychiatry 48 : 493-502, 2007
14. Corbett BA, Constantine LJ, Hendren R, Rocke D, Ozonoff S : Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. Psychiatry Res 166 : 210-222, 2009
15. Robinson S, Goddard L, Dritschel B, Wisley M, Howlin P : Executive functions in children with autism spectrum disorders. Brain Cogn 71 : 362-368, 2009
16. Regard M : Cognitive rigidity and flexibility ; A neuropsychological study. Un-published PhD. dissertation, University of Victoria : 1981
17. Golden JC : Identification of brain disorders by the Stroop color and word test. J Clin Psychol 32 : 654-658, 1976
18. Welsh MC, Pennington BF : Assessing frontal lobe functioning in children : Views from developmental psychology. Developmental Neuropsychology 4 : 199-230, 1988
19. Welsh MC, Pennington BF, Groisser DB : A normative-developmental study of executive function : A window on prefrontal function in children. Developmental Neuropsychology 7 : 131-149, 1991