Saudi normative data for the Wisconsin Card Sorting test, Stroop test, Test of Non-verbal Intelligence-3, Picture Completion and Vocabulary (subtest of the Wechsler Adult Intelligence Scale-Revised)

Ali M. Al-Ghatani, *BA, PhD,* ***Marc C. Obonsawin,*** *MSc, PhD,* ***Basmah A. Binshaig,*** *BA, MSc,* ***Khalaf R. Al-Moutaery,*** *MD, FRCS.*

ABSTRACT

الأهداف: جمع البيانات المعيارية لمجموعة من المقاييس العالمية وتطبيقها على البيئة السعودية وهذه المقاييس كالتالي: اختبار ويسكونسون لتصنيف البطاقات، واختبار ستروب، واختبار الذكاء الغير لفظي، واختبار إكمال الصور، والنسخة المعدلة من اختبار وكسلر لقياس ذكاء البالغين، ومن ثم استخدام هذه البيانات المعيارية من أجل الحصول على معادلات الانحدار المقدرة.

الطريقة: أُجريت هذه الدراسة في قسم الأعصاب بمستشفى القوات المسلحة بالرياض، المملكة العربية السعودية وذلك خلال الفترة من 2000م إلى 2002م، وشملت هذه الدراسة 198 فرداً سليماً تم اختيارهم لتمثيل المتغيرات التالية: العمر، والجنس، وسنوات التعليم، والطبقة الاجتماعية والاقتصادية. ومن ثم تم الحصول على البيانات المعيارية من خلال تطبيق المقاييس التي سبق ذكرها على المشاركين.

النتائج: لقد تم الحصول على البيانات المعيارية من خلال تطبيق الاختبارات القياسية على المشاركين، وعلى ضوء ذلك تم عمل الجداول الخاصة لفهم العلامات التي حصلت عليها الفئات العمرية المختلفة. وتم الحصول على معادلات الانحدار التي تتنبأ بنتائج قدرات المرضى العقلية قبل الإصابة.

خاتمة: قامت هذه الدراسة بجمع البيانات المعيارية الخاصة بثلاثة مقاييس لتحليل وظيفة الفص الجبهي، بالإضافة إلى المقاييس الأخرى التي يتم تطبيقها لتحليل القدرات العقلية العامة وذلك حتى يتم استخدامها في المملكة العربية السعودية. كما تعرض الدراسة طريقة لتحليل القدرات العقلية قبل الإصابة من خلال البيانات المعيارية التي تم الحصول عليها ومن دون اللجوء إلى المقاييس اللفظية.

Objectives: There are 2 aims for this study: first, to collect normative data for the Wisconsin Card Sorting Test (WCST), Stroop test, Test of Non-verbal Intelligence (TONI-3), Picture Completion (PC) and Vocabulary (VOC) sub-test of the Wechsler Adult Intelligence Scale-Revised for use in a Saudi Arabian culture, and second, to use the normative data provided to generate the regression equations.

Methods: To collect the normative data and generate the regression equations, 198 healthy individuals were selected to provide a representative distribution for age, gender, years of education, and socioeconomic class. The WCST, Stroop test, TONI-3, PC, and VOC were administrated to the healthy individuals. This study was carried out at the Department of Clinical Neurosciences, Riyadh Military Hospital, Riyadh, Kingdom of Saudi Arabia from January 2000 to July 2002.

Results: Normative data were obtained for all tests, and tables were constructed to interpret scores for different age groups. Regression equations to predict performance on the 3 tests of frontal function from scores on tests of fluid (TONI-3) and premorbid intelligence were generated from the data from the healthy individuals.

Conclusions: The data collected in this study provide normative tables for 3 tests of frontal lobe function and for tests of general intellectual ability for use in Saudi Arabia. The data also provide a method to estimate pre-injury ability without the use of verbally based tests.

Neurosciences 2011; Vol. 16 (1): 29-41

From the Department of Clinical Neurosciences (Al-Ghatani, Binshaig, Al-Moutaery), Riyadh Military Hospital, Riyadh, Kingdom of Saudi Arabia, and the Department of Psychology (Obonsawin), University of Strathclyde, Glasgow, United Kingdom.

Received 7th December 2010. Accepted 12th December 2010.

Address correspondence and reprint request to: Dr. Ali M. Al-Ghatani, Consultant and Head of Neuropsychology Division, Department of Clinical Neurosciences, Riyadh Military Hospital, Riyadh, Kingdom of Saudi Arabia. Tel. +966 (1) 4777714 Ext. 25404. E-mail: am1221@hotmail.com

From our knowledge and after several investigations, we found limited psychological or neuropsychological tests with normative data modified for Saudi individuals in Saudi Arabia. Therefore, the aim of this study is to describe the work that has been carried out to modify these tests. The availability of Arabic versions of the Wisconsin

Card Sorting Test (WCST), Stroop, Vocabulary (VOC) and Picture Completion (PC) sub-test of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) with normative data for the Saudi healthy population will be very helpful. First, these tests will help the psychologist and neuropsychologist working with Saudi individuals who have Arabic as their first language. Secondly, these tests will give us the chance to improve the care of patients. Thirdly, to generate regression equations from premorbid and fluid intelligence tests to predict scores on the neuropsychological assessments, and to compare these predicted scores from the premorbid and fluid intelligence tests to the scores obtained from the neuropsychological assessments in healthy individuals.

Three categories of tests were chosen to be modified for Saudi individuals. The first category consists of neuropsychological tests known to be sensitive to frontal lobe dysfunction. These tests are the WCST and the Stroop test.[1,2] The second category consists of the Test of Nonverbal Intelligence (TONI-3), a test that is thought to measure the fluid intelligence or 'g'.[3] The third category consists of the PC and VOC sub-tests of the WAIS-R, both of which are thought to measure premorbid intelligence (PI).[4]

Starting with the first category, the WCST was originally developed by Berg[5] to assess the ability to identify abstract categories and the ability to shift cognitive strategies. The first investigators to use the WCST in humans to examine the effects of frontal lobe lesions were Teuber, Battersby, and Bender[6] in 1951. In these early days, the WCST was used to test conceptual ability and abstract thinking, abilities thought to be impaired after damage to the frontal lobes, and the critical measure from the WCST during this time was the number of categories completed by participant as described above. Later, Milner[7] in 1963 observed that individuals with lesions to the frontal lobes made a particular type of error on the WCST, which is the perseverative error. Milner's results with perseveration were replicated by other investigators.[8-11] The perseverative errors score is the most sensitive and significant indicator to differentiate between frontal and nonfrontal groups. According to Milner[7] and Nelson,[12] the most significant measure of both the original WCST and the Modified Card Sorting Test (MCST) is the percentage of errors that are perseverative. The percentage of perseverative error scores on the WCST is the proportion of all the trials that are perseverative errors. It is worth mentioning here that the MCST was developed by Nelson in 1976.[12] The differences between the WCST and the MCST is that in the MCST, there are 24 cards in each set and Nelson changed the method by removing some cards that share more than one aspect with the stimulus card to remove the possibility of having an ambiguous response by the participant. Also, the sorting principles have been changed to be 6 right responses to complete the category, and then the examiner will tell the participant to find another rule to sort the next category. Although there are differences between the 2 versions of the card sorting tests, they are almost measuring the same performance quality.[12,13]

The Stroop test consists of 3 cards. During this test, the participant is presented with these 3 cards. Two cards have columns of words, and each word is the name of a color. On one card the words are printed in black ink. On the other card, the words are printed in colored ink, but never in the same color as represented by the word. A third card has columns of XXXXs, printed in colored ink. The main point of the Stroop test is that it takes longer to say the color of the ink that a word is printed in than it takes to read the word. This task is more difficult when the word is a color that is different from the color of the ink in which it is printed. The semantic meaning of the word interferes with the naming of the color of the ink. The interference score from the Stroop test will be used in the present study because is thought to be the most sensitive measure of frontal lobe dysfunction. Stroop carried out the original version of the Stroop test,[14] and used 5 colors in the test. Later investigators used 3, 4, and 5 colors.[15,16] There is considerable evidence that performance on the Stroop test is affected by damage to the brain, and the test is very sensitive to brain damage as has been reported by many investigators.[3,17-21]

The second category consists of the TONI-3, which is a test that is thought to measure the fluid intelligence. The TONI-3 is a nonverbal intelligence test and does not require an individual to be able to have speech or to have specific language. It requires from the individual only that he or she uses his/her finger to point to the response. The nonverbal tests have 3 essential features. The first feature is that the individual does not respond to questions on the basis of his/her past experiences. The second feature is that questions require an individual to apply reasoning strategies to the problems on the test. The third feature is that multiple strategies are usually required, and the individuals must be able to change these strategies when the problems start to be more difficult. The nonverbal test is often used as a test of intellectual ability for individuals who have difficulty or a disability in language or speech, for individuals who are from a different culture, and for individuals who have hearing problems. It is also useful for individuals suffering from strokes or any brain insult, because such a test will remove the load of the verbal responses. Although there are many nonverbal intelligence tests like Raven's Progressive Matrices and the Cattell's Culture Fair, the TONI-3 was chosen for this particular

study because it is comparable to other tests of fluid intelligence and has been in use in Saudi Arabia for a long time with no Saudi norms available to interpret the scores obtained. In 1997, the TONI-3 test was developed, and they collected new normative data for TONI-3.[3] They normed the TONI-3 on a sample of 3,451 people from different parts of the United States. This third edition has 2 equivalent forms, and each form contains 45 items arranged in order of difficulty plus 5 items for practicing in the beginning of the picture book. The normative data span the age range between 6 years and 89 years and 11 months. The authors claim that they designed this test not to fit in one specific theory of intelligence, but to be a strong test to measure the general intelligence. They focused on abstract reasoning and problem solving to allow a good measure of g, or fluid ability. They correlated the results of TONI-3 with the results of 3 intelligence tests: the Wechsler Intelligence Scale for Children (third edition), the WAIS-R, and the Comprehensive Test of Nonverbal Intelligence (CTONI).[21-23] Performance on the TONI-3 correlates in the moderate to high range with performance on these tests. Scores on the performance scale of the WAIS-R correlate at 0.75 and 0.76 with Forms A and B of the TONI-3, and the Full Scale IQ scores of the WAIS-R correlate at 0.73 and 0.71 of Forms A and B of the TONI-3. Scores on the TONI-3 appear to be strongly associated with scores on other tests of intellectual ability.

The third category consists of the PC and VOC sub-tests of the WAIS-R, both of which are thought to measure PI. The PC and VOC are sub-tests of the WAIS-R. The WAIS-R 1981 is the updated version of the WAIS that was developed in 1955.[24,25] The WAIS-R measures general intelligence and is commonly used by psychologists. It is composed of 11 sub-tests, and these sub-tests can be grouped into 6 verbal and 5 performance subtests. The verbal sub-tests measure our store of knowledge (crystallized intelligence), and all the performance sub-tests except PC measure our ability in solving new problems (fluid intelligence).[4,24] Each sub-test has items ranging from easy to increasingly more difficult. The VOC sub-test was chosen because it is claimed to be the best single measure of PI.[26-29] Because the VOC sub-test had to be translated, it was not sure if the resulting Arabic version of the VOC subtest would be an adequate measure of PI, so a second test of PI, PC, was also administered. Another good reason to include the PC in this study as a premorbid sub-test is because of the need for such nonverbal tests for clinicians to use as predictor with their patients with speech or language problems. The VOC and PC sub-tests were chosen in this study because the VOC and PC are highly correlated, and many investigators reported that performance on both these sub-tests is highly resistant to brain damage.[4,26,30-33] Also, the VOC and PC were suggested be the best subtests of the WAIS-R to predict PI for normal participants between 16-74 years of age, and also to predict pre-injury intelligence for participants with brain damage.[33,34] More investigators have confirmed these findings, and added new equations for participants older than 75 years of age by using the VOC and PC sub-tests with high reliability.[35,36] The VOC sub-test of the WAIS-R consists of 35 English words as items, and each item is scored 2, 1, or 0. The maximum score is 70 points. The instructions to administer this test are very simple. For participants with poor verbal ability, the examiner starts with the first item. For other participants, the examiner starts with the fourth item. The test is discontinued after 5 consecutive failures (responses scored as zero). The original version of the PC sub-test of the WAIS-R contains 20 cards printed with a picture and bound into a booklet. Each picture has one important missing part. The participant is allowed 20 seconds to give the name or to point to the missing part. If the participant does not indicate the missing part, the item is scored (zero) as a failure, and the next card will be presented. After 5 consecutive failures, the test is discontinued. Scoring will be one point for each correct response, and the maximum score is 20 points.

As we mentioned earlier, there were problems with usage of some of these tests in Saudi Arabia, and none of these tests have been modified for the Saudi Arabian population. The non-verbal tests (WCST, TONI-3, and PC) are not a problem, because the only verbal components are the instructions. However, the other tests (Stroop and VOC) involve language. There are 2 problems with them, first is the dissimilarity in the language and second is the validity of the test in the Arabic culture (solving these problems will be seen in the methods section).

Methods. This study included 198 healthy volunteers, and the number of the male participants (N=99) was equal to the females (N=99). All participants were recruited through the Riyadh Al-Kharj Hospital Program (RKH) and from the Medical Services Department (MSD), Riyadh, Kingdom of Saudi Arabia between January 2000 and July 2002. All participants were right handed according to the Annett Handedness Questionnaire that has been translated to Arabic and used in an earlier study.[37,38] Participants were divided into 3 different age groups: i) from 16-30 years of age, ii) from 31-50 years of age, and iii) from 51-65 years of age, and each group has 66 participants. Healthy participants were recruited on the basis of the following inclusion criteria. These participants had to belong to

the Saudi community, and to be Arabic native speakers, to be between 16-65 years of age, to have received education at Primary class and above, to have intact vision, and not be color blind, to be from different socio-economic backgrounds, and to have no history of hospitalization due to head injury, mental illness, stroke, alcoholism, or any other neurological problem, and not currently be on any psychotropic medication. All participants were from different levels of education and from different social-economic backgrounds (class 1, 2, 3, 4, and 5) as shown in **Table 1**. The definition of social class was based on occupation, as defined by the Great Britain Office of Population Censuses and Surveys (1991).[39] A 5-level social class system adopted from this census was used: (1) professional, (2) managerial, (3) skilled, non-manual and manual, (4) partly skilled, and (5) unskilled. The following tests were used:

The PC sub-test of WAIS-R. The original version of the PC sub-test of the WAIS-R contains 20 cards printed with a picture and bound into a booklet. Each picture has one important missing part. The participant is allowed 20 seconds to give the name or to point to the missing part. If the participant does not indicate the missing part, the item is scored (zero) as a failure, and the next card will be presented. After 5 consecutive failures, the test is discontinued. Scoring will be one point for each correct response, and the maximum score is 20 points. Nothing was changed in this subtest, except for the instructions, which had been translated to the Arabic language from the manual of the test and all participants received the Arabic instructions.

The VOC sub-test of WAIS-R. There was a problem as we mentioned earlier with using this test in the English language. The original version of the VOC sub-test of the WAIS-R consists of 35 English words. Each item is scored 2, 1, or 0 (see the manual of the test for the scoring principles and sample responses). The maximum score is 70 points. The test is discontinued after 5 consecutive failures (responses scored zero). All the English words in the original English version were translated to Arabic. In translating, we found that there were sometimes 4 or 5 words in the Arabic language for some English words, and all these words had similar meanings, but most of them were not commonly used. To help in the selection of the appropriate words, we consulted 5 different Arabic-English linguists. We explained the nature of the test, and we asked them to choose the most useful Arabic words, while considering 2 points: first, to chose only original Arabic words that are in the Arabic dictionary, and second, to arrange all the Arabic words in order of difficulty as provided in the English version. An Arabic translation of the same instructions and the original scoring methods was used. In addition, a potential measure of PI was obtained by adding the raw scores of the PC and VOC subtests. In fact, Lucas used VOC + PC as an estimate of PI and called it the Sydney Premorbid Intelligence Estimate (SPIE). They found that the SPIE was the best predictor of premorbid ability.[40,41]

Table 1 - Number of participants according to socio-economic level and educational level.

SOCIOECONOMIC CLASS						
Class One		Class Two		Class Three	Class Four	Class Five
30		58		62	39	9
EDUCATIONAL LEVEL						
Primary	Secondary	High School	Diploma	3-4 years University	5-6 years University	7+ years University
17	35	58	10	47	7	24

The TONI-3 forms A and B. The TONI-3 items are in a picture book, and every page contains one item. This picture book is placed between the subject and the examiner. The instructions provided in the manual are given to the test subject. There is no time limit for this test. We used the same original test booklet and the Arabic instructions.[4]

The Stroop test. We used the Arabic Version of the Stroop test that has been modified for the Saudi population.[42]

The WCST. We used the original version with original instructions translated into the Arabic language.[5]

Recruitment of participants was from the RKH Program and from the MSD of RMH. We had an introductory letter from the Program Director, directed to the RKH and MSD Personnel Department, explaining the nature of the project. We obtained a list of their Saudi employees with reference to their jobs, ages, and qualifications, and we obtained permission for these individuals to attend the screening sessions at specific times. Individuals (n=198) from RKH and the MSD were selected according to the recruitment criteria. These individuals were approached on a personal basis, and were given a full explanation of the project, and a description of the tests that were to be administered with signed informed consent. Ethical permission to proceed with the study was provided by the Hospital Ethics Committee (code 256). The participant's biographical data were then obtained. Administration of the test was randomised by using a list of patterns generated by a BASIC program. These patterns allowed us to vary the order of the tests for each individual. These tests were: VOC sub-test of WAIS-R, PC sub-test of WAIS-R, TONI-3 forms A and B, Arabic Stroop test, and WCST. This session lasted from 60-90 minutes with a 15 minutes coffee break, provided that

Table 2 - Description of the names of tests.

Short name	Full name description
VOC	Vocabulary sub-test of WAIS-R
PC	Picture Completion sub-test of WAIS-R
PI	VOC + PC = Premorbid intelligence
WCST-tc	Wisconsin card sorting test total correct responses
WCST-te	Wisconsin card sorting test total error responses
WCST-pr	Wisconsin card sorting test perseverative responses
WCST-pe	Wisconsin card sorting test perseverative errors
WCST-npe	Wisconsin card sorting test non-perseverative errors
WCST-cat	Wisconsin card sorting test categories
WCST-ppe	Wisconsin card sorting test percent perseverative errors WCST-ppe = (WCST-pe / (WCST-tc + WCST-te))
VF-1	Verbal Fluency test 1 (total of letters: waaw, raa, and taa)
VF-2	Verbal Fluency test 2 (total of letters: siin, daal, and qaaf)
Stroop-w	Stroop test word naming
Stroop-c	Stroop test color naming
Stroop-cw	Stroop test word-word naming
Stroop-prdcw	Stroop test predicted word-word naming
Stroop-int	(prdcw = ((Stroop-c*Stroop-w) / (Stroop-c + Stroop-w)) Stroop test interference (into = (Stroop-cw – stroop-prdcw))
TONI-3A	Test of non verbal intelligence third edition form A
TONI-3B	Test of non verbal intelligence third edition form B

the individual was not exhausted or tired. The statistical analysis comprised t-test, ANOVA, and ANCOVA and was performed with the Statistical Package for Social Sciences (SPSS Inc., Chicago, IL, USA) version 10 for windows.

Results. The abbreviated names for all the test measures are provided in **Table 2**. It will be more useful to go through this table as a first step for easier understanding of the following results. It is also useful to introduce the Arabic verbal fluency (VF) test scores that were previously carried out to generate regression equations for future studies.[37,38]

Distribution of scores. The distribution parameters for each measure using criteria of (-1 or +1) for Skewness showed that most variables are normally distributed except, the number of categories obtained on the WCST are not normally distributed, therefore, the categories of the WCST were analyzed separately using a nonparametric test (Kruskal-Wallis test), the Spearman correlation test was also used. Association of test performance with demographic variables shows that years of education and socio-economic class correlate significantly with all variables. The correlations range from -0.279 to 0.784. The results demonstrate that good performance on all tests is positively associated with the number of years of education and high socio-economic class. An apparent exception is that a high number of correct responses on the WCST are associated with lower socio-economic class. Age is also correlated significantly with years of education and most psychological variables. The correlations range from -0.113 to 0.505, except with the VOC sub-test of WAIS-R and the number of correct responses on the WCST. In addition, **Table 3** shows the mean and standard deviation for the test raw scores. Two-way ANOVA showed no effects of gender on performance on all tests after all test scores were transformed to Z scores. There were no effects of test, gender, nor any interaction.

Association between age groups and test performance. Due to the significant correlations between the different demographic variables and between the demographic variables and test performance, any correction for the effects of age alone could be compounded by the effects of years of education and socioeconomic status. To avoid these confounding effects, the test scores that we used to correct for age were not the original test scores, but the estimated marginal means scores covariated with years of education and socioeconomic status. These estimated means and standard errors are provided in

Table 3 - Mean and standard deviation (SD) for the test's raw scores for males and females.

Tests	Male		Female	
	Mean	SD	Mean	SD
VOC	63.27	3.62	62.87	3.96
PC	15.21	1.40	15.18	1.51
PI	78.48	4.81	78.05	5.29
VF-1	36.23	3.98	35.29	4.25
WCST-tc	77.16	10.07	81.12	10.29
WCST-te	26.01	9.16	26.37	7.20
WCST-pr	19.47	8.07	20.09	6.50
WCST-pe	15.17	6.25	15.52	4.99
WCST-npe	10.87	4.54	11.02	3.59
WCST-cat	5.84	0.40	5.81	0.42
WCST-ppe	14.33	4.40	14.20	3.28
Stroop-w	85.61	14.17	84.24	14.55
Stroop-c	66.65	13.71	64.50	14.44
Stroop-cw	41.63	8.57	40.97	9.60
Stroop-int	4.19	2.18	4.48	2.89
TONI-3A	28.37	3.09	28.70	3.77

Table 4 - Performance (mean estimate and standard error) on each measure according to age group. The mean estimate and the standard error for the following tests before the age correction with class and education were used as covariates.

Tests	Age group 16-30		Age group 31-50		Age group 51-65	
	Mean	Std. Error	Mean	Std. Error	Mean	Std. Error
VOC	62.9	0.329	63.4	0.327	62.9	0.328
PC	15.4	0.111	15.5	0.111	14.7	0.111
PI	78.3	0.409	78.9	0.405	77.6	0.407
VF-1	36.2	0.352	35.9	0.349	35.1	0.351
Stroop-w	88.0	1.110	86.9	1.100	79.9	1.100
Stroop-c	68.5	1.090	67.7	1.080	60.4	1.090
Stroop-cw	42.9	0.686	42.8	0.681	38.1	0.683
WCST-tc	80.4	1.055	75.9	1.047	81.1	1.050
WCST-pr	17.5	0.722	18.9	0.717	23.0	0.719
WCST-te	23.8	0.793	24.5	0.787	30.3	0.790
WCST-pe	13.0	0.521	14.4	0.516	18.5	0.518
WCST-npe	10.7	0.478	10.2	0.474	11.9	0.476
TONI-3A	28.9	0.292	28.5	0.290	28.2	0.291

Table 5 - Significance values of multiple comparisons Sidak test comparing performance on tests between different age groups.

Tests	16-30 vs. 31-50	16-30 vs. 51-65	31-50 vs. 51-65
VOC	0.704	> 0.999	0.686
PC	0.843	< 0.001	< 0.001
PI	0.695	0.519	0.076
VF-1	0.936	0.103	0.290
Stroop-w	0.860	< 0.001	< 0.001
Stroop-c	0.937	< 0.001	< 0.001
Stroop-cw	0.997	< 0.001	< 0.001
WCST-tc	0.008	0.952	< 0.001
WCST-pr	0.479	< 0.001	< 0.001
WCST-te	0.909	< 0.001	< 0.001
WCST-pe	0.173	< 0.001	< 0.001
WCST-npe	0.778	0.211	0.025
TONI-3A	0.693	0.329	0.921

Table 4. The scores for WCST-cat are absent from this table, and different analyses and tables will be provided later for WCST-cat. It is the estimated means presented in **Table 4**, corrected for the effects of years of education and socioeconomic status that were used to make the correction for age alone. Between the 3 age groups, using the years of education and socioeconomic class as covariates, most of the tests scores were significantly different, except on the VOC, PI, TONI-3A, and Verbal Fluency (VF-1) tests.

Multiple post hoc comparisons were carried out using the Sidak correction (as recommended by Field).[43] **Table 5** shows the significant values for performance on the different tests between different age groups. Between the youngest 16-30 and the middle 31-50 age groups, the only difference was on WCST total correct. Between the 16-30 and 51-65 age groups, the differences were seen on WCST-pr, WCST-pe, and WCST-te. Also there are changes between the 31-50 and 51-65 age groups on most of the WCST (WCST-tc, WCST-pr, WCST-pe, and WCST-te). The interference measures on the Stroop test show changes between the 31-50 and 51-65 age groups (Stroop-w, Stroop-c, and Stroop-cw). On the PC, there are changes between the youngest and the oldest age groups (16-30, 51-65) and the between the middle and the oldest age group (31-50, 50-65).

For the WCST-cat, a nonparametric Kruskal-Wallis test of variance was carried out between the WCST-cat and the 3 age groups. The results of the analysis in **Table 6** show the mean rank for each age group, and show that the Kruskal-Wallis test indicates significant differences between the median of the 3 age groups and their performance on the WCST-cat. Because of this significance, pair-wise comparisons were carried out using the Mann-Whitney test. As seen in the same table, all of the 3 age groups differ from each other. With aging, people are more likely to achieve fewer than 6 categories, but most people still achieve 6 categories. Although the median number of categories completed in each age group is 6, older people are more likely to complete fewer categories than younger people. Because the median number of categories is the same in each age group, a correction for age was not applied for the WCST-cat.

Table 6 - Performance (mean rank) on WCST-cat according to age group.

	Age group 16-30	Age group 31-50	Age group 51-65
WCST-cat	Mean rank	Mean rank	Mean rank
	111.55	100.90	86.05
Kruskal-Wallis test for the test below			
WCST-cat	Degrees of freedom	Chi-square	Significance
	2	15.787	<0.001
Significance values of multiple comparisons Mann-Whitney test comparing performance on test between different age groups.			
WCST-cat	16-30 vs. 31-50	16-30 vs. 51-65	31-50 vs. 51-65
	0.040	<0.001	0.042

Frequencies, number of categories completed and the median number of categories completed on the WCST for each age group.

	Age groups							
	16-30			31-50			51-65	
Number of participants	3	63	1	9	56	1	19	46
Number of categories completed	5	6	4	5	6	4	5	6
Median	6		6			6		

Due to the difference in performance with age on most other tests, the age of the individual completing the test should be taken into consideration when interpreting any normative score. To correct the age group differences, we subtracted the mean estimate score (education and class used as covariates) of the age group affected from the score of the age group with the highest mean (which was mostly the youngest age group) and added that difference to the affected age group. For example, for the Stroop-w, the mean performance for each of the 3 age groups is 88, 86.9, and 79.9. To correct the scores for the 51-65 age group, we subtracted the mean score for this group from the mean score of the group with the highest mean score, which is the youngest age group (88-79.9=8.1) and added the rounded difference (to remove the age effect) to the score of the 51-65 age group, or added 8 points to the raw score. As the 51-65 age group is almost 8 points lower than the scores of the other 2 age groups, the mean of the oldest age group was subtracted from the mean of the youngest age group.

Table 7 - Corrections for age for the scores on all tests must be carried out before using the percentile norms tables.

Tests	Age group 16-30	Age group 31-50	Age group 51-65
TONI-3A	Nothing to add	Nothing to add	Nothing to add
WCST-tc	Subtract 5 points from the total score	Nothing to subtract	Subtract 5 points from the total score
WCST-te	Nothing to subtract	Nothing to subtract	Subtract 7 points from the total score
WCST-pr	Nothing to subtract	Nothing to subtract	Subtract 4 points from the total score
WCST-pe	Nothing to subtract	Nothing to subtract	Subtract 4 points from the total score
WCST-npe	Nothing to subtract	Nothing to subtract	Subtract 2 points from the total score
Stroop-c	Nothing to add	Nothing to add	Add 8 points to the total score
Stroop-w	Nothing to add	Nothing to add	Add 8 points to the total score
Stroop-cw	Nothing to add	Nothing to add	Add 5 points to the total score
VF-1	Nothing to add	Nothing to add	Nothing to add
PC	Nothing to add	Nothing to add	Add one point to the total raw score
VOC	Nothing to add	Nothing to add	Nothing to add
PI	Nothing to add if you already added the points to PC	Nothing to add if you already added the points to PC	Nothing to add if you already added the points to PC
WCST-cat	Nothing to add	Nothing to add	Nothing to add

Note: some of the tests above do not require any age correction

Table 7 shows how to carry out the age corrections for every single test (to enable the examiner to interpret the scores obtained on any test by using the percentile norms tables provided later), but not for WCST-cat. The means and standard deviations from the descriptive statistics obtained for the 3 age groups after correcting the age for all by using the ANCOVA test shows no differences in the mean scores between the 3 age groups after the age correction.

Percentile norms for each neuropsychological test. Percentile norms tables (age corrected) are available now for individuals living in a Saudi Arabian culture. **Table 8** shows the percentile norms for the Arabic version of VOC, PC, PI (VOC+PC) sub-tests of WAIS-R that can be used as measurements of PI, TONI-3 test (form A and B), VF-1, and Stroop test (interference). Percentile norms for the Arabic versions of the WCST are shown in **Table 9**.

Table 8 - Percentile norms for the Arabic versions of the Vocabulary, Picture Completion, Premorbid Intelligence, TONI-3, Verbal Fluency and Stroop (Interference) (*age corrected), for Saudi individuals between the ages of 16 and 65 years.

Percentiles	Vocabulary raw scores	Picture Completion raw scores	PI (VOC + PC)	TONI-3A	TONI-3B	Total VF-1 counted words	Interference scores
>99	70	19	88	39	40	46	12
95	69	18	86	35	36	43	9
90	68	17	85	33	34	42	8
85	68	17	85	31	33	41	7
80	67	17	84	31	32	39	7
75	66	16	83	30	31	38	6
70	66	16	82	30	31	37	6
65	65	16	81	29	31	37	5
60	64	16	80	29	30	36	5
55	64	16	79	30	30	36	4
50	63	15	78	28	29	35	4
45	62	15	78	28	29	35	4
40	62	15	77	28	28	35	4
35	61	15	76	27	28	34	4
30	60	15	75	27	28	33	3
25	60	15	75	26	27	32	3
20	60	14	74	26	27	32	2
15	59	14	73	25	26	32	2
10	58	14	72	24	25	31	2
5	56	13	70	23	23	29	1
<1	54	12	67	21	21	27	0

*To use this table to interpret the scores obtained on the test, age correction has to be carried out first by using the correction method for the test on Table 7.

Table 9 - Percentile norms for the Arabic version of the WCST (*age corrected) for Saudi individuals between the ages of 16 and 65 years.

Percentiles	Correct	Total errors	Perseverative responses	Perseverative errors	Non-perseverative errors	Categories	Percent perseverative errors
>99	57	8	3	2	2	6	3
95	62	11	7	7	4	6	8
90	64	15	10	8	6	6	9
85	65	16	12	9	6	6	10
80	66	18	13	10	7	6	11
75	68	19	14	10	8	6	11
70	69	20	15	11	8	6	12
65	70	21	15	12	9	6	12
60	72	22	16	12	9	6	13
55	73	23	17	13	9	6	13
50	74	23	18	14	10	6	14
45	76	24	19	14	10	6	14
40	77	25	19	15	11	6	14
35	79	26	20	15	11	6	15
30	81	27	21	16	12	6	15
25	83	28	22	17	12	6	16
20	85	29	24	18	13	6	16
15	87	31	26	19	14	5	17
10	91	33	27	20	16	5	18
5	94	39	31	24	18	5	21
<1	101	43	39	27	20	4	24

*To use this table to interpret the scores obtained on the test, age correction has to be carried out first by using the correction method for the test on Table 7.

The main aim of this study is to collect normative data for use in Saudi Arabia for tests of frontal lobe function, fluid intelligence and PI, and to develop methods to predict the estimate premorbid ability. To satisfy the aim, one method of regression was used to predict scores from the premorbid and from the fluid intelligence scores. This method was the linear regression for the prediction of scores on tests of frontal lobe function (VF, Stroop, WCST) by using scores from one other variable, scores of either premorbid or fluid intelligence. Linear regression is to be used for future study tests by assessing the ability of tests of fluid and PI alone to predict scores on tests of frontal lobe function. The aim of this exercise is not to generate the model that predicts the most variance, but to investigate how much variance is predicted by scores of fluid and PI. To allow us to carry out the best comparison between performance of frontal lobe tests as predicted by a test of fluid intelligence and as predicted by a test of crystallized intelligence, it was first necessary to identify the test of crystallized intelligence that correlated best with the test of fluid intelligence (TONI-3). There were 2 reasons for doing this. First, it was necessary to have 2 predictors for the healthy people, one predictor on the basis of performance on a test of fluid intelligence, and one predictor on the basis of performance on a test that estimates PI. Second, if performance on the tests of fluid and crystallized intelligence are highly correlated, then they can be used in future study to predict scores on neuropsychological tests from the performance on the premorbid and fluid intelligence tests. There are various possible criteria to choose between PC, VOC, or PI as a test of PI. One possible criterion is that the test is not sensitive to brain injury. Another criterion is the extent of similarity of the measure of PI with the measure of fluid intelligence, TONI-3. The PC shows that is the test with the highest correlation with TONI-3 (r=0.821), the second was the PI (r=0.790), and the third was the VOC (r=0.742). The higher correlation between TONI-3 and PC rather than with VOC or PI indicates that TONI-3 and PC are more similar than TONI-3 and VOC or TONI-3 and PI. To have more than one equation from different predictors, PC, VOC and PI were selected to be the measures of PI and to be used as predictors. Scores from TONI-3, PC, VOC, and PI were used to predict scores on the VF-1, Stroop interference, and WCST percent perseverative errors for the researchers and for future study.

Linear regression equations were generated for future study for each test, using the age-corrected scores for each variable, and the methods are shown in **Table 10**. **Table 11** shows significant values for the predicted equations for future study. **Table 12** shows the predicted equations for VF-1, Stroop, and WCST, and the percentage variance accounted for by the equations based on PC, VOC, PI, and TONI-3 (for future study). **Table 13** shows the means, and standard deviations for the predicted scores from PC, predicted scores from

Table 10 - Methods used in the linear regression equations to predict performance on the VF, Stroop, and WCST from crystallized or fluid intelligence.

Dependent variables	Predictors	Methods	Variables		R	R. Squ.	Std. error of the estimate
			Entered	Excluded			
VF-1	PC	Enter	PC	-	0.724	0.524	2.862
VF-1	TONI-A	Enter	TONI-A	-	0.717	0.515	2.889
VF-1	VOC	Enter	VOC	-	0.738	0.545	2.797
VF-1	PI	Enter	PI	-	0.761	0.579	2.691
Stroop-Int	PC	Enter	PC	-	0.685	0.470	1.825
Stroop-Int	TONI-A	Enter	TONI-A	-	0.678	0.460	1.840
Stroop-Int	VOC	Enter	VOC	-	0.636	0.404	1.934
Stroop-Int	PI	Enter	PI	-	0.672	0.452	1.855
WCST-ppe	PC	Enter	PC	-	0.275	0.075	3.655
WCST-ppe	TONI-A	Enter	TONI-A	-	0.357	0.127	3.551
WCST-ppe	VOC	Enter	VOC	-	0.305	0.093	3.621
WCST-ppe	PI	Enter	PI	-	0.307	0.094	3.617

R = Correlation, R Squ = R Square

Table 11 - Significance values for the linear regression equations.

Dependent variables	Predictors	Entered	Coefficient	B	Std. Error	T	Significance of the model
VF-1	PC	PC	(Constant)	1.986	2.310	0.860	0.391
			PC	2.175	0.148	14.680	<0.001
VF-1	TONI-A	TONI-A	(Constant)	11.154	1.719	6.487	<0.001
			TONI-A	0.862	0.060	14.416	<0.001
VF-1	VOC	VOC	(Constant)	-15.048	3.320	-4.532	<0.001
			VOC	0.806	0.053	15.332	<0.001
VF-1	PI	PI	(Constant)	-13.817	3.027	-4.565	<0.001
			PI	0.631	0.038	16.413	<0.001
Stroop-Int	PC	PC	(Constant)	-14.680	1.473	-9.968	<0.001
			PC	1.244	0.094	13.171	<0.001
Stroop-Int	TONI-A	TONI-A	(Constant)	-9.419	1.095	-8.599	<0.001
			TONI-A	0.493	0.038	12.928	<0.001
Stroop-Int	VOC	VOC	(Constant)	-21.783	2.296	-9.486	<0.001
			VOC	0.419	0.036	11.528	<0.001
Stroop-Int	PI	PI	(Constant)	-21.819	2.086	-10.458	<0.001
			PI	0.337	0.026	12.708	<0.001
WCST-ppe	PC	PC	(Constant)	25.524	2.950	8.653	<0.001
			PC	-0.756	0.189	-3.997	<0.001
WCST-ppe	TONI-A	TONI-A	(Constant)	25.005	2.113	11.835	<0.001
			TONI-A	-0.393	0.074	-5.351	<0.001
WCST-ppe	VOC	VOC	(Constant)	33.001	4.298	7.678	<0.001
			VOC	-0.305	0.068	-4.480	<0.001
WCST-ppe	PI	PI	(Constant)	32.137	4.068	7.900	<0.001
			PI	-0.234	0.052	-4.522	<0.001

B = probability of type II error, T = Wilcoxon's matched pairs signed rank statistics

VOC, predicted scores from PI, and predicted scores from TONI-3A, and the obtained scores from the VF-1, Stroop-int, and WCST-ppe. The results of ANOVA tests show no differences between the scores predicted by the equations based on TONI-3, PC, VOC, or PI, and the obtained scores from VF-1, Stroop interference, and WCST percent perseverative errors (for future study).

Discussion. The purpose of this study was to provide normative data for WCST, Stroop test, TONI-3, PC and VOC sub-tests of the WAIS-R for the Saudi Arabian individuals. The availability of the Arabic versions of these tests will then be used to generate regression equations in future study. It is the time to welcome the first Arabic normative data for the WCST, Stroop, VF, TONI-3, PC, VOC, and PI tests that are available now for Saudi Arabia. These norms are appropriate for use in Saudi Arabia for individuals of Saudi origin. Their use in other Arabic cultures is possible, but remains to be tested. In addition, regression equations have been generated to predict the performance on the WCST, Stroop, and VF tests on the basis of performance on tests of PI and on tests of fluid intelligence.

The norms are described for 3 age groups. The 3 age groups were chosen as they reflect 3 periods of change in life circumstances in Saudi Arabia. The early age range

Table 12 - The equations to predict scores from PC (age corrected), VOC (age corrected), PI (age corrected), and TONI-3A (age corrected) to VF-1 (age corrected), Stroop-int (age corrected) and WCST-ppe (age corrected) for future study.

The equations to predict scores from PC	The equations to predict scores from VOC	The equations to predict scores from TONI-3A	The equations to predict scores from PI
VF-1 = 2.175 x (PC) + 1.986 % Variance explained = 52.4	VF-1 = 0.806 x (VOC) – 15.048 % Variance explained = 54.5	VF-1 = 0.862 x (TONI) + 11.154 % Variance explained = 51.5	VF-1 = 0.631 x (PI) - 13.817 % Variance explained = 57.9
Stroop-int = 1.244 x (PC) - 14.680 % Variance explained = 47.0	Stroop-int = 0.419 x (VOC) - 21.783 % Variance explained = 40.4	Stroop-int = 0.493 x (TONI) - 9.419 % Variance explained = 46.0	Stroop-int = 0.337 x (PI) - 21.819 % Variance explained = 45.2
WCST-ppe = -0.756 x (PC) + 25.524 % Variance explained = 07.5	WCST-ppe = -0.305 x (VOC) + 33.001 % Variance explained = 09.3	WCST-ppe = -0.393 x (TONI) + 25.005 % Variance explained = 12.7	WCST-ppe = -0.234 x (PI) + 32.137 % Variance explained = 09.4

Table 13 - The means and standard deviation for the predicted scores from PC, the predicted scores from VOC, the predicted scores from PI, the predicted scores from TONI-3A, and the obtained scores from the neuropsychological tests.

Test	Obtained Scores		Predicted Scores from PC		Predicted Scores from VOC		Predicted Scores from PI		Predicted Scores from TONI-3A	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
VF-1	35.76	4.14	35.76	2.99	35.79	3.06	35.75	2.96	35.78	3.15
Stroop-int	4.64	2.50	4.64	1.71	4.64	1.59	4.65	1.70	4.67	1.68
WCST-ppe	13.78	3.79	13.78	1.04	13.76	1.16	13.79	1.35	13.74	1.17

(16-30 years) is the period to complete education (high school and university), to find a job, and to save money for marriage. The middle age range (31-50) is the period to create your own family, to develop your career, to plan for the future, and secure the life of your family. The oldest age range (51-65) is the period for early retirement from service, to create personal business, and to enjoy the rest of your life.

There is no reliable census information in Saudi Arabia at the present time, so it is difficult to construct a picture of socioeconomic levels. However, the British census does provide a clear description of socioeconomic levels, and this system was adopted for this study. It seemed that it would be useful to have stages of education that reflected Saudi Arabian society. Many individuals, particularly older individuals, have only a primary school education. It is in the past 50 years or so that education past the primary level has become more widespread. The levels of years of education in the present study were divided into 7 levels for 2 reasons. The first is to have a sample from most of the Saudi population especially from the old ages and secondly to differentiate between a person with primary level and between a person with a doctorate level of education.

The normative results of the present study can be compared to other normative results in 2 ways: first, the values for each variable can be compared to those obtained from other studies in other cultures, and second, the relationship between these scores and the demographic variables (age, gender, education, and socioeconomic level) can also be compared. First, the mean scores of all tests were compared with the mean scores on the original normative data. All the variables, with the exception of the number of categories obtained on the WCST were distributed normally, so there was no floor or ceiling effects other than for the number of categories.

The results with the WCST indicate that the pattern of the number of categories, correct responses, total errors, perseverative errors, and non perseverative errors are similar to the results provided by Heaton and Boone.[44,45] The results with the Stroop test indicate that the healthy participants took longer to say the color of the word than to read the word, and took even longer to say the color of an incongruous color word. The results definitely demonstrate the Stroop effect, and are very similar to the norms provided by other investigators.[2,14,46]

The results with the TONI-3 test indicate that the scores of the healthy participants on forms A and B are very similar to the scores described by previous investigators.[47-49] The results with PC are similar to the norms results provided by Kaufman[4] and Lezak.[27] However, the VOC results are slightly different. The scores in the Arabic Vocabulary version are higher than the scores on the original version. This difference is due

to the difference in language. Although the words in the Arabic version were chosen by Arabic-English linguists, the Arabic words looked easier than the English words. For example, the word Tirade is a very difficult word in the English language, and is still difficult in the Arabic language, but may be easier than the English word. The increasing difficulty of each item in the Arabic version of the VOC subtest was not tested, so the version described in this study may be easier. However, the holy Qur'an contains many difficult Arabic words, and the Muslim people have to read the Qur'an and have to understand every single word. It is possible that the scores in the Arabic version were higher than the scores on the English version because of this familiarity with the words of the Qur'an. Despite this, the higher mean score VOC versus the distribution of scores is still normal, as indicated by the low Skewness score. Also, the potential measure of PI obtained in this study by adding the raw scores of the PC and VOC subtests showed that the combination of these 2 sub-tests is a good measure of PI, which is similar to the result obtained by previous investigators.[40]

Secondly, the effects of demographic variables on performance on these tests can also be compared. Four demographic variables were considered: age, gender, level of education, and socioeconomic level. Two of these variables (age, gender) are biological characteristic of the individuals, and 2 variables (education and socioeconomic level) reflect the individual in a social context. Because of the relatively small sample in the present study, correction to the norms for any effects of these variables was only carried out for the relevant biological variable, in this case age. The years of education and socioeconomic class were used as covariates to avoid any correction for education and socioeconomic class when the correction of age was made to the age groups. Performance on most tests, with the exception of VOC, VF, and TONI-3, was affected by age, and in all tests, performance decreased with age. These results are identical to those reported in other studies for each of the tests, the WCST, the Stroop, and PC.[2,4,14,43,46,49-53] It is worth mentioning that there was no significant effect of age on the performance of the VF test and TONI-3, but other studies reported the affect of age on the performance of VF.[54-56] However, evidence from other and more recent studies reported that performance on the oral VF test (COWAT) does not decline with age for subjects up to 90 years old.[57-59] There are likely to be a number of reasons why the performance on VF in the present study showed no association with age. First, the version in the present study required the participant to go through 3 levels of difficulty. Second, the ANOVA comparisons between age groups were covaried for the years of education and socioeconomic class. These demographic variables affect performance on VF much more than age does. Regarding the lack of association between age and performance on the TONI-3, previous studies have shown an effect of age on TONI-3, but in these studies, the effects of age were investigated without first correcting for the effects of years of education and socioeconomic state, as was carried out in the current study.[45,60] It is possible then that the effects of age on TONI-3 reported in previous studies reflect more the effects of years of education, or even socioeconomic status, rather than age itself.

Scores for all these variables were corrected to account for the age effects only. There are different ways to correct the age, but for the present study, age was corrected using the method of Golden[2] because it is very clear for researchers and clinicians to use.

Performance on all tests was not significantly different between males and females. These findings are identical to those reported in other studies for each test.[2,43,46,51]

Education and socioeconomic levels affected performance on all tests in the present study. Many studies have reported similar findings for the WCST, the Stroop, VF, TONI-3, PC, and VOC.[2,4,43,46,49,51,61,62] It is because of these effects that the years of education and socioeconomic class are used as covariates in this study.

In contrast, the Saudi Arabian normative data for the WCST, Stroop, VF, TONI-3, PC, VOC, and PI tests are very similar to the original versions of these tests, and these similarities make the normative data presented in this study appropriate for use in Saudi Arabia for individuals of Saudi origin.

Regarding the use of regression, 2 different methods of regression have been used in this study for the reasons described earlier in the results section. The regression equations generated using the PC, VOC PI (as PI tests) and TONI-3 (as fluid intelligence tests) all appear to predict scores for WCST, the STROOP and the VF tests. For each of these tests, there is no difference between the 4 predicted scores and the obtained scores. The regression equations based on PC, VOC, and PI predict a much lower percentage of the variance for perseverative errors on the WCST than for words produced on the VF task or interference on the Stroop test. This result reflects the lower correlations reported between tests of PI and perseveration than with other measures of frontal lobe function.[63,64] Also, the regression equations based on TONI-3 predict a much lower percentage of the variance for perseverative errors on the WCST, but it is a reasonable predictor of scores for the Stroop test and the VF test. These equations are not for clinical purposes, but for future study. For clinical use, regression equations were generated with demographic variables in addition to a measure of PI.

The work carried out in this study has given us the chance to achieve 4 aims, to modify the WCST, Stroop, VF, TONI-3, PC, and VOC for use in the Saudi Arabian culture, to construct normative tables describing the performance on these tests of Saudi people, to generate regression equations to predict performance on WCST, Stroop, and VF for future study and for clinical usage, and to compare these predicted performances to the obtained scores. Also, these normative data will give us the chance to carry out more studies to investigate more neuropsychological studies.

References

1. Parker DM, Crawford JR. Assessment of frontal lobe dysfunction. In: Crawford JR, Parker DM, McKinlat WW, editors. A Handbook of Neuropsychological Assessment. Hove: Lawrence Erlbaum Association; 1992. p. 267-291.
2. Golden CJ. Diagnosis and rehabilitation in clinical neuropsychology. Springfield (IL): Charles C. Thomas; 1978.
3. Brown L, Sherbenou RJ, Johnsen SK. Test of Nonverbal Intelligence. 3rd ed. Austin (TX): PRO-ED; 1997.
4. Kaufman AS. Assessing adolescent and adult intelligence. Boston (MA): Allyn and Bacon; 1990.
5. Berg EA. A simple objective test for measuring flexibility in thinking. ***J Gen Psychol*** 1948; 39: 15-22.
6. Teuber HL, Battersby WS, Bender MB. Performance of complex visual tasks after cerebral lesions. ***J Nerv Ment Dis*** 1951; 114: 413-429.
7. Milner B. Effects of different brain lesions on card sorting. ***Arch Neurol*** 1963; 9: 90-100.
8. Drewe EA. The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. ***Cortex*** 1974; 10: 159-170.
9. Robinson AL, Heaton RK, Lehman RA, Stilson DW. The utility of the Wisconsin Card Sorting test in detecting and localizing frontal brain lesions. ***J Consult Clin Psychol*** 1980; 48: 605-614.
10. Weinberger DR, Berman KF, Zec RF. Physiological dysfunction of dorsolateral prefrontal cortex in schizophrenia: I. Regional cerebral blood flow evidence. ***Arch Gen Psychiatry*** 1986; 43: 114-124.
11. Wildgruber D, Kischka U, Fassbender K, Ettlin T. The frontal lobe score: part II: evaluation of its clinical validity. ***Clin Rehabil*** 2000; 14: 272-278.
12. Nelson H. A modified card sorting test sensitive to frontal lobe defects. ***Cortex*** 1976; 12: 313-324.
13. Greve KW, Smith MC. A comparison of the Wisconsin Card Sorting Test with the Modified Card Sorting Test for use with older adults. ***Gerontol Geriatr Educ*** 1991; 11: 57-65.
14. Stroop JR. The basis of Ligon's theory. ***Am J Psychol*** 1935; 47: 499-504.
15. Thurstone LL, Mellinger JJ. The Stroop test. Raleigh (NC): The Psychometric Laboratory, University of North Carolina; 1953.
16. Broverman DM. Cognitive style and intra-individual variations in ability. ***J Personality*** 1960; 28: 240-256.
17. Golden CJ. The diagnosis of brain damage by Stroop test. ***J Clin Psychol*** 1976; 32: 654-658.
18. Perret E. The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. ***Neuropsychologia*** 1974; 12: 323-330.
19. Holst P, Vilkki J. Effect of frontomedial lesions on performance on the Stroop Test and word fluency tasks. ***J Clin Exp Neuropsychol*** 1988; 10: 79.
20. Vendrell P, Junque C, Pujol J, Jurado MA, Molet J, Grafman J. The role of prefrontal regions in the Stroop task. ***Neuropsychologia*** 1995; 33: 341-352.
21. Stuss DT, Floden D, Alexander MP, Levine B, Katz D. Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. ***Neuropsychologia*** 2001; 39: 771-786.
22. Wechsler D. Wechsler Intelligence Scale for children. 3rd ed. San Antonio (TX): Psychological Corporation; 1991.
23. Hammill DD, Person NA, Wiederholt JL. The Comprehensive Test of Nonverbal Intelligence. Austin (TX): PRO-ED; 1996.
24. Wechsler D. WAIS-R manual. New York (NY): The Psychological Corporation; 1981.
25. Wechsler D. WAIS manual. New York (NY): The Psychological Corporation; 1955.
26. Belsky JK. The psychology of aging theory, research, and interventions. 2nd ed. Pacific Grove (CA): BrooksCole Pub. Co. 1990.
27. Lezak MD. Neuropsychological assessment. 3rd ed. New York (NY): Oxford University Press; 1995.
28. Toivonen, J, Kuikka P, Kaukinen S. Effects of deliberate hypotension induced by labetalol with isoflurane on neuropsychological function. ***Acta Anaesthesiol Scand*** 1993; 37: 7-11.
29. Muldoon MF, Ryan CM, Matthews KA, Manuck SB. Serum cholesterol and intellectual performance. ***Psychosom Med*** 1997; 59: 382-387.
30. Crawford JR, Besson JAO, Bremner M, Ebmeier KP, Cochrane RHB, Kirkwood K. Estimation of premorbid intelligence in schizophrenia. ***Br J Psychiatry*** 1992; 161: 69-74.
31. Mangone CA, Sica REP, Pereyra S, Genovese O, Segura E, Riarte A, et al. Cognitive impairment in human chronic Chagas' disease. ***Arq Neuropsiquiatr*** 1994; 52: 200-203.
32. Alekoumbides A, Charter RA, Adkins T, Seacat G. The diagnosis of brain damage by the WAIS, WMS, and Reitan battery utilizing standardized scores corrected for age and education. ***International Journal of Clinical Neuropsychology*** 1987; 9: 11-28.
33. Goldstein G, Watson JR. Test-retest reliability of Halstead-Reitan Battery and the WAIS in a neuropsychiatric population. ***Clin Neuropsychol*** 1989; 3: 265-272.
34. Krull KR, Sherer M, Adams RL. A comparison of indices of premorbid intelligence in clinical population. ***Appl Neuropsychol*** 1995; 2: 35-38.
35. Vanderploeg RD, Schinka JA. Predicting WAIS-R IQ premorbid ability: combining subtest performance and demographic variable predictors. ***Arch Clin Neuropsychol*** 1995; 10: 225-239.
36. Sattler J. Assessment of children. San Diego (CA): Jerome M. Sattler Publisher; 1988.
37. Al-Ghatani AM, Obonsawin MC, Al-Moutaery KR. Normative data for the two equivalent forms of the Arabic Verbal Fluency Test. ***Pan Arab Journal of Neurosurgery*** 2009; 13: 57-65.
38. Obonsawin MC, Al-Ghatani AM, Akhdar FM, Al-Moutaery KR. Two equivalent forms of the Verbal Fluency test in the Arabic language. ***Neurosciences (Riyadh)*** 1999; 4: 115-119.
39. Simpson S. Editorial: Coverage of the Great Britain Census of Population and Housing. ***J R Statist Soc A*** 1994; 157: 313-316.
40. Krull KR, Scott JG, Sherer M. Estimation of premorbid intelligence from combined performance and demographic variables. ***Clin Neuropsychol*** 1995; 9: 83-88.

41. Lucas SK, Carstairs JR, Shores EA. A comparison of methods to estimate premorbid intelligence in an Australian sample: Data from the Macquarrie University Neuropsychological Normative Study (MUNNS). ***Aust Psychol*** 2003; 38: 3.
42. Al-Ghatani AM, Obonsawin MC, Al Moutaery KR. The Arabic version of the Stroop test and its equivalency to the English version. ***Pan Arab Journal of Neurosurgery*** 2010; 14: 112-115.
43. Field A. Discovering statistics using SPSS for Windows. London (UK): Sage; 2000.
44. Heaton RK. A manual for the Wisconsin Card Sorting Test. Odessa (FL): Psychological Assessment Resources; 1981.
45. Boone KB, Ghaffarian S, Lesser IM, Hill-Gutierrez E, Berman NG. Wisconsin card sorting performance in healthy, older adults: Relationship to age, sex, education, and IQ. ***J Clin Psychol*** 1993; 49: 54-60.
46. Golden CJ. A group form of the Stroop colour and word test. ***J Pers Assess*** 1975; 39: 386-388
47. Anastasi A. Psychological testing. 6th ed. New York (NY): Macmillan; 1988.
48. Brown L, Bryant BR. The why and how of special norms. ***Remedial and Special Education*** 1984; 5: 52-61.
49. Brown L, Sherbenou RJ, Johnsen SK. Test of Nonverbal Intelligence. 2nd ed. Austin (TX): PRO-ED; 1990.
50. Heaton RK, Grant I, Matthews CG. Comprehensive norms for an expanded Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications. Odessa (FL): Psychological Assessment Resources; 1991.
51. Chelune GJ, Baer RL. Developmental of norms for the Wisconsin Card Sorting Test. ***J Clin Exp Neuropsychol*** 1986; 8: 219-228.
52. Brown W. Practice in associating colour-names with colours. ***Psychol Rev*** 1915; 22: 45-55.
53. Daigneault S, Braun CMJ, Whitaker HA. Early effects of normal aging on perseverative and non-perseverative prefrontal measures. ***Dev Neuropsychol*** 1992; 8: 99-114.
54. Welsh MC, Pennington BF, Groisser DB. A normative-developmental study of executive function: A window on prefrontal function in children. ***Dev Psychol*** 1991; 7: 131-149.
55. Comalli PE, Wapner S, Werner H. Interference effects of Stroop colour-word test in childhood, adulthood, and aging. ***J Genet Psychol*** 1962; 100: 47-53.
56. Schaie KW. Rigidity-flexibility and intelligence: A cross-sectional study of the life span from 20 to 70 years. ***Psychol Monogr*** 1958; 72: Whole No. 462.
57. Schaie KW, Strother GR. Cognitive and personality variables in college graduates of advanced age. In: Talland GA, editors. Human Aging and Behaviour. New York (NY): Academic Press; 1968
58. Furrey CA, Baltes PB. The effect of age differences in ability-extraneous performance variables on the assessment of intelligence in children, adult and the elderly. ***J Gerontol*** 1973; 28: 73-80.
59. Boone KB, Miller BL, Lesser IM, Hill E, D'Elial. Performance on frontal lobe tests in healthy older individuals. ***Dev Neuropsychol*** 1990; 6: 215-223.
60. Parkin AJ, Walter BM. Recollective experience, normal aging and frontal dysfunction. ***Psychol Aging*** 1992; 7: 290-298.
61. Aiken LR. Psychological testing and assessment. 6th ed. Boston (MA): Allyn and Bacon; 1994.
62. Salvia J, Ysseldyke JE. Assessment. 6th ed. Boston (MA): Houghton Mifflin; 1995.
63. Malec JF, Ivnik RJ, Smith GE. Mayo's older American normative studies: utility of correction for age and education for the WAIS-R. ***Clinical Neuropsychologist*** 1992; 6: 31-47.
64. Obonsawin MC, Crawford JR, Page J, Chalmers P, Low G, Marsh P. Performance on the Modified Card Sorting Test by normal, healthy individuals: relationship to general intellectual ability and demographic variables. ***Br J Clin Psychol*** 1999; 38: 27-41.

Authorship entitlement

Excerpts from the Uniform Requirements for Manuscripts Submitted to Biomedical Journals updated November 2003.
Available from www.icmje.org

The international Committee of Medical Journal Editors has recommended the following criteria for authorship; these criteria are still appropriate for those journals that distinguish authors from other contributors.

Authorship credit should be based on 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) intellectual content; and 3) final approval of the version to be published. Authors should meet conditions 1, 2, and 3.

Acquisition of funding, collection of data, or general supervision of the research group, alone, does not justify authorship.

An author should be prepared to explain the order in which authors are listed.