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

Sophie Bayard1,*, Jérôme Erkes2, Christine Moroni3, the Collège des Psychologues Cliniciens spécialisés en Neuropsychologie du Languedoc Roussillon (CPCN Languedoc Roussillon)

1Service de Neurologie, Unité des Troubles du Sommeil, Hôpital Gui-de-Chauliac, Centre hospitalier universitaire, Montpellier Cedex 5, France
2Pôle de Gériatrie, Centre Antonin Balmes, Centre hospitalier universitaire, Montpellier, France
3Université Nord de France, Jeune Equipe 2497 : Neuropsychologie et Cognition Auditive, Villeneuve d’Ascq, France

*Corresponding author at: Service de Neurologie, Unité des Troubles du Sommeil, Hôpital Gui-de-Chauliac, Centre hospitalier universitaire, Montpellier Cedex 5, France. Tel.: (33) 4 67 33 94 68; fax: (33) 4 67 33 72 85.
E-mail address: bayardsophie@yahoo.fr (S. Bayard).

Accepted 12 June 2011

Abstract

The Stroop Color-Word Test-Victoria version (VST) is a measure of executive function commonly used in neuropsychological evaluation. Because of its short administration time, the VST seems particularly appropriate for use in geriatric populations and with those suffering from dementia and who are prone to fatigue during neuropsychological examination. In this study, we examine the influence of demographic characteristics on VST score and present descriptive data for a sample of 244 elderly French speakers (50–94 years of age). Normative data corrected for age and education are provided for clinical use. Furthermore, by comparing the VST performance of patients with the Alzheimer-type dementia to that of 40 matched healthy controls, we provide clinical evidence suggesting that the VST has a clinical utility in the assessment of inhibition in AD.

Keywords: Stroop Color-Word Test-Victoria version; normative data; inhibition; Alzheimer’s disease

Introduction

Executive functions are among the first cognitive functions to bear the negative effects of normal aging (Belleville, Chertkow, & Gauthier, 2007; Belleville, Rouleau, & Van der, 2006; Collette, Schmidt, Scherrer, Adam, & Salmon, 2009; Daigeneault, Braun, & Whitaker, 1992) and are severely impaired in early Alzheimer’s disease (AD; for review, see Collette & Van der Linden, 2004). Inhibitory control is classically considered as representing an important executive function (e.g., Miyake et al., 2000). Inhibitory impairments have been reported in several neuropsychological conditions, as well as in normal aging (e.g., Hasher & Zacks, 1988) and have frequently been reported in the first stage of AD (for review, see Amieva, Phillips, Della, & Henry, 2004). Typically, the Stroop interference effect is considerably greater in AD patients than in healthy older people, even after having been corrected for processing speed (Amieva, Lafont, et al., 2004; Bondi et al., 2002). Consequently, the assessment of executive functioning is a critical component of the neuropsychological evaluation of the elderly.

In North America and in French-speaking communities, the Stroop Test is among the top 10 most frequently used means of evaluating in executive functioning (Groupe de Réflexion sur l’Évaluation des Fonctions Exécutives, 2001; Rabin, Barr, & Burton, 2005). The Stroop effect originally described by Stroop (1935) has traditionally been viewed as an expression of executive functioning, and more specifically, of the ability to inhibit an over-learned response in favor of an unusual one (Zacks & Hasher, 1994). The classical effect is that response latency to name the color of ink of printed words is exceptionally long when
the written word and the name of the ink are different colors (called hereinafter Interference Stroop Effect), thereby highlighting the difficulty of inhibiting an automatic response.

There are a number of versions of the Stroop test (Mitrushina, Boone, & D'Elia, 2005). These versions vary in the number of stimuli cards, items, in the size and presentation of stimuli cards, the administration procedures, the method of scoring, and the normative data calculation. For example, few standardized Stroop tests provide normative data for errors scores and not all versions correct for generalized slowing on the interference condition. Despite the presence of numerous versions, a classical Stroop test consists of an interference task and control tasks. The Stroop Color-Word Test-Victoria version (VST) developed by Spreen and Strauss (1998) is a brief version of the Stroop task. VST uses three conditions that consist in naming the color of dots, of neutral words, and of color words printed in incongruent colors. Each condition contains 24 items. Due to the brief administration time (~5 min), this version seems particularly appropriate for use with geriatric and brain-damaged populations who are prone to fatigue during neuropsychological examination. The VST demonstrates good psychometric properties, including excellent test–retest reliability (Troyer, Leach, & Strauss, 2006), sensitivity to frontal lobe versus non-frontal damage (for review, see Strauss, Sherman, & Spreen, 2006), amnesic mild cognitive impairment and AD (Joubert et al., 2010), and aging (Moroni & Bayard, 2009; Troyer et al., 2006). In spite of the popularity of the VST, few comprehensive sets of norms exist. To the best of our knowledge, there is only one published study proposing VST data norms (Troyer et al., 2006). The authors of this study have reported performances on the VST among a large group of adults ranging in age from 18 to 94 (N = 272) and living in the same community: namely Ontario (Canada).

In this study, we provide descriptive and normative data for the VST in an elderly francophone population (50–94 years old, N = 244). For this purpose, the French adaptation of the VST (f-VST) was used. Because previous studies carried out with English speakers have shown that age and education are potential correlates of performance on the Stroop tasks (Mitrushina et al., 2005; Strauss et al., 2006; Troyer et al., 2006), we examined the influence of these demographic characteristics on f-VST performances (Study 1). Additionally, we assessed the clinical implications of the f-VST for neurodegenerative conditions in which inhibition deficits are frequently reported, that is, AD (Study 2).

**Study 1: Normative data for elderly francophone adults**

**Methods**

**Participants.** The participants were 244 healthy community-dwelling adults living in Montpellier and Lille (France). Their mean age was 65.83 (SD = 10.71) and 49% were men. All were native French speakers with normal or corrected hearing and vision at the time of testing. They were recruited from an adult participants’ pool and senior-citizen associations. Expert clinicians interviewed each participant to screen for neurological and psychiatric disorders that could affect cognitive abilities. Interviews were conducted by registered psychologists (S.B., J.E., and other psychologists from “Les Membres du Collège des Psychologues Cliniciens Spécialisés en Neuropsychologie du Languedoc Roussillon”). All participants had a Mini-Mental State Examination (MMSE) score above the 10th percentile taking into account their level of education (participants aged from 50 to 79: Kalafat, Hugonot-Diener, & Poitrenaud, 2003; participants aged >79: Lechevalier-Michel, Fabrigoule, Lafont, Letenneur, & Dartigues, 2004). The healthy participants carried out only the f-VST and received no financial compensation for their participation.

**French adaptation of the VST.** Regard’s (1985) original task was initially translated into French (for a detailed description, see Strauss et al., 2006). We added a practice row for the three conditions as no practice stimuli were present in the original version of the VST. Stimuli were presented on three different cards, one for each condition. Four colors were used: blue, green, yellow, and red. In the “Dot” condition, color dots were presented. In the “Word” condition, the words *mais* (but), *pour* (for), *donc* (thus), and *quand* (when) were written in a random order in one of the four color inks listed. In the “Interference” condition, the words *bleu* (blue), *vert* (green), *jaune* (yellow), and *rouge* (red) were written in one of the three other colors (e.g., the word green would be written in yellow ink). Twenty-four stimuli were presented on each card and were arranged in a 4 × 6 matrix. Card conditions were presented in the following order: Dot, Word, and Interference. In each of the three conditions respectively, participants were asked to name as quickly as possible the color of the dots (Dot condition) and the color of the written words (Word condition and Interference condition). In the Word and Interference conditions, participants had to inhibit the written word in order to correctly name the color of the ink. The Word condition may be considered as an intermediate inhibition condition as the classical interference effect between the written word and the color name is not present. For each condition, we measured the completion time and the number of errors (corrected, non-corrected, and total errors). Two interference scores were computed: (a) Word/Dot for time and (b) Interference/Dot for time.
Statistical analyses. The statistical analyses were carried out with SPSS version 16.0 for Windows (SPSS, Inc., Chicago, USA). Because many distributions of the f-VST measures were skewed, logarithmic transformations were performed using either log10(X) when the measure is expressed in seconds (such as the time to complete each condition or the interference scores) or log10(X + 1) when the measure is a number of errors (as recommended by Tabachnik and Fidell). For most of these transformed scores, skewness values fell within the acceptable range (i.e., between -1 and 1). However, exceptions were observed for errors on the Dot, Word, and Interference conditions. Therefore, the normative data for errors were proposed in percentiles. To determine the association between transformed f-VST variables and age, the Pearson correlations were performed with the entire sample and effect sizes were examined (with rs of .10, .30, and .50 defined as small, medium, and large effect size, respectively, Cohen, 1988). Education level was divided dichotomously into participants with <12 years of schooling (Level 1) and participants with a number of years of schooling ≥12 years (Level 2). Group differences in f-VST scores were analyzed with one-way between-groups analysis of variance (ANOVA) with age as a covariate. Gender effect was analyzed with a parametric ANOVA. Finally, in order to correct for demographic variables that contribute to the variability of the f-VST, normative data were calculated using a regression-based approach. Standard multiple regressions were performed between f-VST measures as dependent variables, and age and education level as independent variables. The level of significance was α < 0.05.

Results

Error rates. Errors on the Dot condition (inter-quartile range = 0, range 0–2) and the Word condition (inter-quartile range = 0, range 0–2) were rare, with 94% of participants making no errors on the Dot condition, and 95% of participants making no errors on the Word condition. Consequently, no further analyses were performed on the error scores of these two conditions.

Because of the restricted range and resulting skewed distribution of error scores on the Interference condition (mean = 1.43, SD = 1.96, range 0–12 with more than 90% of participants making fewer than five errors), we did not explore the relationship between this f-VST variable and demographic factors. Table 1 presents the percentile score for the f-VST error rates in the Interference condition.

Completion time. Determination of the influence of demographic variables on the f-VST: Correlations between age and time required to complete Dot, Word, and Interference conditions were positive and statistically significant (r = .53, .61, and .57, respectively, all ps < .001). However, no correlation between age and Word/Dot or Interference/Dot index was observed. The ANCOVAs (with age as a covariate) showed that participants with <12 years of education (Level 1) performed worse than participants with a higher level of education (≥12 years; Level 2). This was true for all conditions (transformed time) and interference scores: Dot, F(1, 242) = 7.82, p = .006; Word, F(1, 242) = 26.83, p < .001; Inference, F(1, 242) = 24.68, p < .001; Word/Dot, F(1, 242) = 4.58, p = .033; and Interference/Dot, F(1, 242) = 7.37, p = .007. Men and women did not significantly differ on f-VST indices. Furthermore, for each measure of the f-VST, a mean comparison was conducted to analyze the potential influence of the gender as a demographic factor. Men and women did not significantly differ on the f-VST indices (p = .78 for the Dot condition, p = .32 for the Word condition, p = .29 for the Inference condition, p = .34 for the Word/Dot, and p = .29 for the Interference/Dot).

Regression-based normative data. Since both age and education (based on the number of years of education) show sizeable associations with the f-VST indices, these variables were considered separately in regression analyses. Models are presented in Table 2. Age appears to exert a significant influence on f-VST indices (from 1% to 37% of the variance) in all indices apart from the interference indices. Education accounts for between 2% and 10% of the f-VST indices variance.

When a regression-based approach is used in a clinical application, the raw scores of a person are converted into standardized residuals in three steps. First, the predicted scores of the person are calculated; second, the residuals (e_\text{id}) are calculated; and third,
the residuals are standardized ($Z_i = e_{id} / SD_{[residual]}$). We can illustrate this method with the following example: a 73-year-old man with 12 years education took 15 s on the Dot condition, 26 s on Word condition, and 69 s on Interference condition. The interference scores were 1.73 (Word/Dot) and 4.60 (Interference/Dot). To determine whether or not this participant obtained a normal Interference/Dot score, a predicted score was calculated using the equation: $0.35 + (0.002 \times 73) - (0.069 \times 2) = 0.358$. The residual was $2 - 0.305$ ($= 0.358 - \log_{10}[4.60]$) and the standardized residual was $2 - 2.77$ ($= 0.305/0.11$), which can be considered as a pathological score (calculations for all the f-VST scores are presented in the Appendix).

Because the regression-based approach requires some calculations, we suggest downloading automatic formulas that are available on the following website: http://nca.recherche.univ-lille3.fr/index.php?page=materiel-de-test-2.

**Study 2: f-VST for the Alzheimer-type dementia**

**Method**

**Participants.** Two groups of participants were recruited for this study: patients with AD and normal elderly participants. The AD patients consisted of 40 patients (19 men and 21 women, aged between 62 and 89 years old—mean age: 77 [± 5.6]). They were recruited among outpatients referred for neuropsychological assessment in two Memory Clinics (Lille and Montpellier) and met the DSM-IV (American Psychiatric Association, 2000) clinical criteria and the NINCDS-ADRDA research criteria for probable AD (McKhann et al., 1984). All patients had suffered from progressively worsening memory problems for the last 6 months minimum. The diagnosis was based on general medical, neurological, and neuropsychological examinations. Patients’ mean MMSE score was 23.6 ± 3.5 (16–29). Forty normal older participants matched for age, sex, and education level were selected from the Study 1 set. The control participants were non-institutionalized alert and had no history of neurological problems, alcohol abuse, or psychiatric disorders. They had normal or corrected vision and normal or corrected hearing. The average age for the control group was 77.5 ± 3.8 years. All control participants had a total MMSE score above the 10th percentile when taking their level of education into consideration (Kalafat et al., 2003). They were all native French speakers.

**French adaptation of the VST.** The f-VST was administered following the procedure described in Study 1 (see Methods section). It was systematically introduced into a battery of standardized neuropsychological tests used in routine clinical procedure.

**Statistical analyses.** The statistical analyses were carried out with SPSS version 16.0 for Windows (SPSS, Inc.). The data were examined for normal distribution and homogeneity of variance. Group differences were analyzed with Student’s $t$-tests for independent samples for continuous variables, the Mann–Whitney tests for ordinal data, and the $\chi^2$ test for categorical variables. Analyses were carried out using the f-VST untransformed raw scores. Correlations were evaluated with the Bravais–Spearman coefficients. The level of significance was $\alpha < 0.05$.

**Results**

The participant’s results for the f-VST are described in Table 3. The AD patients were slower than control participants and produced more errors in the Word and Interference conditions (all $ps < .05$). The AD patients’ interference ratio indices (Word/
Dot and Interference/Dot) were significantly higher than the control group’s (all \( p < .01 \)). Note that group differences on the interference ratio were greater than those on the Word and Interference conditions.

When controlling for age and education level, a significant negative correlation was observed between Word/Dot and MMSE \( (r^2 = - .33, p = .003) \), with higher Word/Dot indices associated with lower MMSE scores. The same pattern of association was observed for Interference/Dot and MMSE score \( (r^2 = - .22, p = .052) \) but was not statistically significant.

**General discussion**

The results of this study provide the only source of normative data for the VST for an elderly francophone population. The f-VST was used on a normal elderly population to measure the inhibition of pre-potent responses. Furthermore, we provide clinical evidence suggesting that the f-VST is of interest in the assessment of inhibition in AD patients.

In this study, we decided to use a regression-based approach to calculate our normative data for time measures. We have several reasons for preferring our approach to the “traditional” method in which raw scores are converted into \( Z \)-scores to evaluate a person’s performance. The first reason is that the regression-based norms provide more accurate estimates of population statistics because they are based on equations that are derived from all demographic groups (Van Breukelen & Vaeyen, 2005; Zachary & Gorsuch, 1985). The second reason is that with the regression-based approach, norms can even be provided for people with certain demographic characteristics that were not in the sample groups. For example, in this study, despite our best efforts, we had difficulties recruiting very old people (i.e., 80 years and more) with a high level of education. This problem stems from the fact that among earlier generations, further education was not the norm. Finally, with this approach, the total sample size does not need to be subdivided to provide norms that are corrected for demographic variables. This dramatically decreases the sample size on which the normative statistics are calculated. The heavily skewed error score for the Interference condition, however, did not allow us to use the regression-based norms. As the error rate was extremely low for this f-VST score, cumulative percentages were proposed for the interference error score without splitting the sample by age group and level of education.

In this sample, gender was not found to be significantly associated with performance, a result that is inconsistently reported in the past (see Mitrushina et al., 2005). Although women outperform men (Stroop, 1935; Van der Elst, Van Boxtel, Van Breukelen, & Jolles, 2006), gender differences are not always present in the literature (e.g., Ivnik, Malec, Smith, & Tangalos, 1996). Consistent with past studies, we found that education was related to Stroop performance (time scores). Participants with a higher level of education performed better than those with fewer years of schooling. In their study presenting normative data for anglophone Canadians, Troyer and colleagues (2006) found that education had a negligible association with VST indices (correlation ranging from \(- .14 \) to \(- .24 \)). In our population, education accounted for 2%–10% of the variance of f-VST, which is also a small effect.

Concerning the influence of age on the f-VST scores, we found a very large effect of this demographic variable on Dot, Word, and Interference time scores (from 33% to 37% of the variance). This effect falls dramatically for the Interference/Dot ratio (only 1% of the variance). Moreover, age did not influence the Word/Dot ratio score at all. These results may suggest that the classical effect of age on time measures in the Stroop test is in fact contaminated by a general slowness (Saltheuse, 1996). Inhibitory processes might then not be directly influenced by age. This point of view has already been proposed by Verhaeghen and De Meersman (1998) in a meta-analysis of the effects of age on Stroop performance. However, this
absence of effect of age on a “purer” measure of inhibition is inconsistent with several reports using interference time scores that are corrected for general slowing (Delis, Kaplan, & Kramer, 2001; Ivnik et al., 1996; Klein, Ponds, Houx, & Jolles, 1997; West, 1999). Several reasons may explain this difference. First, as we proposed and in accordance with Verhaeghen and De Meersman’s meta-analysis (1998), it may be that inhibitory processes as measured by the Stroop test are not particularly age sensitive (Moroni & Bayard, 2009). Second, it is possible that the f-VST presents a lower difficulty level for older people compared with other versions of the Stroop test (such as Golden, Perret, or Delis–Kaplan versions). Indeed, in contrast to other versions, which have a large number of items (i.e., 60–112) in each condition, the f-VST has only 24 items on each of the three tasks. Therefore, in the f-VST, sustained attention is probably less involved and this may impact on the age effect. Moreover, the f-VST presents another important methodological difference compared with classical Stroop tests. In fact, order and content of the three conditions are different in the f-VST. In classical Stroop tests, the interference condition is often preceded by a simple word reading condition. Subsequently, in the interference condition, the participant has not only to cope with the interference effect but also to change a previously initiated behavior (i.e., reading words). This probably increases the executive aspect of the task by adding a flexibility component. It has recently been demonstrated that the order in which the conditions are administered in the Stroop test influences performance in this way (Amieva, Lafont, et al., 2004). In the f-VST, the interference condition is preceded by a “low” interference condition (Word condition) in which the subject has to inhibit the reading behavior, while this behavior has not been previously initiated and while the interference produced by the written words is low. This Word condition could act as a preparation for the inhibition behavior, with a lower complexity level that may facilitate subsequent performance. In the psychology of learning, it is a well-known effect that some first easier trials facilitate the subsequent ones and the learning of a skill (Dutta, Schweickert, Choi, & Proctor, 1995). In consequence, we can suggest that the f-VST is easier to perform than the other versions. This could have caused the Stroop effect, as assessed by the f-VST corrected time scores, to show no effect of age. Nevertheless, this lower level of the difficulty of the f-VST remains hypothetical.

In this study, we were also interested in exploring the clinical utility and the sensitivity of the f-VST in AD patients. As age and prevalence of neurodegenerative diseases increase in the general population, the neuropsychological evaluation of the elderly has grown considerably in the last few years. The assessment of executive functioning has major clinical interest in pathological geriatric conditions. In fact, increasing numbers of studies suggest that executive functions appear to be affected early in neurodegenerative diseases (Amieva, Phillips, et al., 2004; Bherer, Belleville, & Hudon, 2004; Dujardin & Defebvre, 2007). However, executive assessment may induce several methodological problems. Some are directly related to the definition of the concept of executive functions (e.g., construct validity and the test–retest problems; for review, see Strauss et al., 2006). Others are specifically related to the neuropsychological evaluation of elderly people with neurodegenerative disorders. Cognitive tasks that depend on executive processes are generally time-consuming. It is therefore very difficult to carry out such tasks with patients with low attentional resources such as AD patients (Berardi, Parasuraman, & Haxby, 2005; McGuinness, Barrett, Craig, Lawson, & Passmore, 2010). In our study, patients with AD performed normally on the Dot condition but their performances were impaired on the Word and Interference cards of the f-VST (times and error rates). Patients also showed impairment in ratio scores that took into account their general cognitive processing speed. Our results on the f-VST replicate the findings of Amieva and colleagues (2002) and Amieva, Phillips, and colleagues (2004) and those of Spieler, Balota, and Faust (1996), who also reported Stroop deficits in persons with AD.

Our study has several limitations. The Study 1 participants were not recruited from a population sample but from an adult participants’ pool and senior-citizen associations. Ideally, a random sampling method would have been preferable, maximizing the representativeness of the sample. The clinicians must be aware of this limitation because it reduces the generalization of the f-VST normative data use. The f-VST is not sensitive to age differences in inhibitory control in healthy adults from 50 to 94 years old. As mentioned previously, it is possible that the VST might be easier than an other Stroop task. It would have been interesting to administer jointly to the f-VST another Stroop procedure such as the DKEFS Color-Word Interference Test version (Delis et al., 2001) in order to test this hypothesis and to explore the construct validity of the f-VST in the assessment of inhibition. A final limitation of the present study is the relatively small number of participants in each age–education cluster.

In conclusion, the main objective of this study was to propose a valid and norm-based cognitive test that exploits response inhibition to clinicians working with elderly patients. The f-VST is of great clinical interest. It requires only a short administration time (∼5 min) so it may be ideal for detecting response inhibition in geriatric populations with cognitive fatigability. Both measures, recording error rates and calculating interference ratios, are important in determining that deteriorating performances on the f-VST are specifically related to response inhibition impairment rather than a general cognitive slowing frequently reported in normal aging (Salthouse & Meinz, 1995). Furthermore, the intermediate interference condition (i.e., Word condition) is of interest for individuals who are not able to complete the interference condition. The results of our study also suggest that the f-VST has strong potential for characterizing response inhibition impairment in AD.
Conflict of interest

None declared.

Acknowledgements

We wish to thank E. Ó Duíll and A. Arthley for their suggestions on improving the manuscript.

Appendix

To determine whether or not a participant obtains normal scores on the f-VST, several steps are needed.

Step 1: Collect performance in each condition of the f-VST

A 73-year-old man with a level of education of 12 years (level 2) scores 15 s on Dot condition, 26 s on Word condition and 69 s on Interference condition.

From these scores, both interference scores are calculated: Word/Dot index, \( \frac{26}{15} = 1.73 \); and Interference/Dot index, \( \frac{69}{15} = 4.60 \).

Step 2: Logarithmic transformations

In this step, all f-VST scores undergo a \( \log_{10}(X) \) transformation.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Raw scores</th>
<th>( \log_{10}(X) ) transformation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dot condition</td>
<td>15 s</td>
<td>1.176</td>
</tr>
<tr>
<td>Word condition</td>
<td>26 s</td>
<td>1.415</td>
</tr>
<tr>
<td>Interference condition</td>
<td>69 s</td>
<td>1.839</td>
</tr>
<tr>
<td>Word/Dot</td>
<td>1.73</td>
<td>0.239</td>
</tr>
<tr>
<td>Interference/Dot</td>
<td>4.60</td>
<td>0.663</td>
</tr>
</tbody>
</table>

Step 3: Calculate for each index a predicted score using a specific regression equation

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Equation</th>
<th>Predicted score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dot condition</td>
<td>( \hat{y} = 0.80 + (0.006 \times \text{age}) - (0.038 \times \text{level of education})^{a} )</td>
<td>1.162</td>
</tr>
<tr>
<td>Word condition</td>
<td>( \hat{y} = 0.94 + (0.006 \times \text{age}) - (0.073 \times \text{level of education})^{a} )</td>
<td>1.232</td>
</tr>
<tr>
<td>Interference condition</td>
<td>( \hat{y} = 1.15 + (0.008 \times \text{age}) - (0.107 \times \text{level of education})^{a} )</td>
<td>1.52</td>
</tr>
<tr>
<td>Word/Dot</td>
<td>( \hat{y} = 0.19 - (0.038 \times \text{level of education})^{a} )</td>
<td>0.114</td>
</tr>
<tr>
<td>Interference/Dot</td>
<td>( \hat{y} = 0.35 + (0.002 \times \text{age}) - (0.069 \times \text{level of education})^{a} )</td>
<td>0.358</td>
</tr>
</tbody>
</table>

\(^{a}\text{Level of education: 1, <12 years of education; 2, \geq 12 years of education.}\)

Step 4: Calculate residuals and Z-score for each f-VST score

<table>
<thead>
<tr>
<th>Conditions</th>
<th>( e_{id} = (\text{predicted score} - \text{observed score})^{b} )</th>
<th>Residual standard deviation</th>
<th>Z-score</th>
<th>Clinical decision(^{b})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dot condition</td>
<td>1.162 – 1.176 = −0.014</td>
<td>0.105</td>
<td>−0.13</td>
<td>Average</td>
</tr>
<tr>
<td>Word condition</td>
<td>1.232 – 1.415 = −0.182</td>
<td>0.09</td>
<td>−2.03</td>
<td>Deficit</td>
</tr>
<tr>
<td>Interference condition</td>
<td>1.52 – 1.839 = −0.319</td>
<td>0.12</td>
<td>−2.65</td>
<td>Deficit</td>
</tr>
<tr>
<td>Word/Dot</td>
<td>0.114 – 0.239 = −0.125</td>
<td>0.09</td>
<td>−1.38</td>
<td>Borderline</td>
</tr>
<tr>
<td>Interference/Dot</td>
<td>0.358 – 0.663 = −0.305</td>
<td>0.11</td>
<td>−2.77</td>
<td>Deficit</td>
</tr>
</tbody>
</table>

\(^{b}\text{Clinical decisions could be either “deficit” if the Z-score is inferior to −1.65, or “borderline” if the Z-score is between −1.64 and −0.9, or “average” if the Z-score is between −0.9 and 0.9, or “superior” if the Z-score is between 0.9 and 1.64, or “very superior” if the Z-score is superior to 1.65.}\)
References


