Visual Determinants of Reduced Performance on the Stroop Color-Word Test in Normal Aging Individuals

M.P.J. van Boxtel¹, M.P.M. ten Tusscher ², J.F.M. Metsemakers³, B. Willems¹, and J. Jolles¹
¹Department of Psychiatry and Neuropsychology, University of Maastricht, The Netherlands,
²Department of Ophthalmology, University of Maastricht, The Netherlands, and
³Department of General Practice, University of Maastricht, The Netherlands

ABSTRACT

It is unknown to what extent the performance on the Stroop color-word test is affected by reduced visual function in older individuals. We tested the impact of common deficiencies in visual function (reduced distant and close acuity, reduced contrast sensitivity, and color weakness) on Stroop performance among 821 normal individuals aged 53 and older. After adjustment for age, sex, and educational level, low contrast sensitivity was associated with more time needed on card 1 (word naming), red/green color weakness with slower card 2 performance (color naming), and reduced distant acuity with slower performance on card 3 (interference). Half of the age-related variance in speed performance was shared with visual function. The actual impact of reduced visual function may be underestimated in this study when some of this age-related variance in Stroop performance is mediated by visual function decrements. It is suggested that reduced visual function has differential effects on Stroop performance which need to be accounted for when the Stroop test is used both in research and in clinical settings. Stroop performance measured from older individuals with unknown visual status should be interpreted with caution.

INTRODUCTION

The Stroop Color-Word test (SCWT) has established the reputation of being a useful and reliable assessment tool in neuropsychological practice. It seems particularly appropriate as a measure of concentration effectiveness (Lezak, 1995), which taps on the domains of both information processing speed and sustained attention. The test is often incorporated in diagnostic protocols to assess diffuse brain damage such as that found in traumatic brain injury (Bohnen, Twijnstra, & Jolles, 1992; Spreen & Strauss, 1998). In addition, because of its sensitivity to calendar age, the test has become increasingly popular in cognitive aging studies (Houx, Jolles, & Vreeling, 1993; Klein, Ponds, Houx, & Jolles, 1997; Uttl & Graf, 1997; Verhaeghen & De Meersman, 1998). Basically, the outcome of the test reflects a person’s ability to suppress a habitual response in support of an unusual one, i.e., naming the ink color that (incongruously named) color words are printed. For this reason, the Stroop paradigm has been used to study frontal lobe function (Vendrell et al., 1995). Recent evidence suggests that the anterior cingulate cortex is a central structure in the prefrontal areas that subserves multiple attentional circuits during Stroop task performance (Peterson et al., 1999).

Central factors are, however, only one aspect underlying performance in a Stroop paradigm. A prerequisite for reliable performance testing in a Stroop paradigm is an intact visual system, particularly color vision and visual acuity, so that the target words can be identified and processed correctly. Negative associations between visual
function and performance have been demonstrated earlier in neuropsychological tests that rely strongly on visual processing (Kempen, Krichevsky, & Feldman, 1994), but also in tests based on a visual evoked potential paradigm (Kugler, 1999). Although authoritative neuropsychological textbooks warn that visual dysfunctions may interfere with Stroop performance (Lezak, 1995; Spreen & Strauss, 1998), no studies to-date have addressed this issue in a systematic fashion. The prevalence of minor to major visual impairment owing to a reduced static acuity is strongly related to calendar age. In one study among older drivers the prevalence increased rapidly from 23% in individuals aged 64 or younger, to 72 and 96% in people aged 65–75 years and older than 75 years, respectively (Kline & Scialfa, 1996). This loss in visual acuity may in part be associated with elevated contrast discrimination thresholds in older individuals, because of increased blur in the optical media. Although the prevalence is quite stable over age, impairments in color discrimination are common (up to 8% in men; Uvijls, 1998), and even if they often go unnoticed by the individual, a reduced capacity to discriminate between primary colors may affect Stroop performance. In this study, we focused on the subtle defects in visual function that are often observed in normal aging individuals as determinants of performance on the classic version of the Stroop test. More specifically, we evaluated the impact of visual acuity (determined both distant and nearby), color weakness, and contrast sensitivity on Stroop performance. Our objective is to determine the importance of common age-related visual impairments to Stroop test performance (both speed and errors), whether associations between these variables and Stroop performance are mediated or modified by calendar age, and finally, the implications of such relationships in a clinical context.

METHODS

Subjects
A group of 838 individuals took part in the study. This group consisted of participants in the 3-year follow-up assessment of the Maastricht Aging Study (MAAS), a study into determinants of usual cognitive aging (Jolles, Houx, van Boxtel & Ponds, 1995; Van Boxtel et al., 1998). Only individuals aged 52 and older were included. Baseline assessment of a wide array of cognitive and medical variables had been performed 3 years earlier in participants aged between 23 and 82 years. They had been randomly recruited from a patient register of collaborating general practices in the Limburg region of The Netherlands (Registration Network Family Practices; Metsemakers, Hoppen, Knottnerus, Kocken, & Limonard, 1992). Individuals with diagnosed morbidity related to brain health (e.g., neurological or psychiatric disease) or who used psychotropic medication were not included in the study. Furthermore, participants were stratified for age, sex, and educational level (Van Boxtel et al., 1998). All participants had received at least primary education and had subjectively no overt difficulties with reading at close range. Of 838 individuals who took part in the assessment, 17 had incomplete data on the Stroop test and were not included in the analyses. Table 1 shows the age distribution of the study sample.

The protocol was approved by the ethics committee of Maastricht University Hospital and written informed consent was obtained from all participants.

Measurements
All measurements used in the analysis took place during a single session in the psychological test laboratory of the Maastricht Brain and Behavior Institute. The extended-vision testing described below was performed at the 3-year follow-up only. All visual tests were executed under standard illumination with appropriate optical correction (i.e., spectacles or contact lenses). Educational level was expressed on an 8-point scale ranging from primary education to university degree (De Bie, 1987).

Vision Tests
– Visual acuity—distant: Binocular visual acuity was measured with a Landolt-C optotype chart (TNO—Netherlands Organization of Applied Science) at a distance of 5 m. Participants are asked to name the orientation of black circles on a white background that contained an opening in one of four possible directions. The size of the circles that were named correctly was the outcome of the test. The score is expressed as the ratio of 5 m and the distance at which a reference
population with normal vision correctly encodes the circle orientation, e.g., 1 is average for young individuals, .5 indicates that individuals with normal vision encode the stimuli of that particular size correctly at 10 m.

- Visual acuity – close: Binocular near vision was tested with the Amsterdam reading chart (Anonymous, 1998). It contains a standard black-on-white text printed in decreasing font sizes. The score is the font size of text that can be read without errors, with a range of .5 (optimal) to 1.25 (poor).

- Color vision: Color blindness or color weakness was identified with the Ishihara color plate set (Ishihara, 1994). Digits consisting of colored dots were presented on a round background with dots of the same size but different hue. Based on errors, total color blindness (achromatopsia) and two different kinds of color weaknesses (red/green or yellow/blue type) can be differentiated from normal color vision. In the population, total color blindness is a rare condition, but 8% of men and 1% of women have some form of color deficiency, mainly of the red/green variety (Uvijls, 1998).

- Contrast sensitivity: The LH–Low Contrast Symbol Test (Hyvärinen, 1992) was used to determine reduced contrast sensitivity, i.e., foreground/background discrimination. Different simple pictograms (e.g., apple or square) printed in decreasing hues of gray had to be identified by the individual at a distance of 1 and 3 m. The score depends on the minimal grey hue that is required to identify the pictograms correctly. The outcome ranges between 0 (high contrast needed = impaired) to 6 (low contrast needed = normal). Blurring of the optical media (e.g., by cataract) is the most prevalent cause of reduced contrast sensitivity in old age.

**Stroop Test**

The Stroop test version described by Hammes (1973) is used widely in Dutch neuropsychological research and practice. It consists of three white cards containing the stimulus material in a 10 x 10 matrix. Card 1 contains color words in random order (red, blue, yellow, green) that are printed in black ink. Card 2 displays solid color patches in one of these four basic colors. Card 3 again contains color words, but were printed in a discongruous ink color. Individuals were instructed to read the words (card 1), name the colors (card 2) and, finally, name the ink color of the printed words (card 3) in three subsequent sessions. Participants were urged to process the stimuli as fast as possible whilst being as accurate as possible. The total time required to process all hundred stimuli was recorded to the nearest tenth of a second, including the number of errors made in each condition. The Stroop interference score is expressed as the time required for card 3 minus card 2 (Hammes, 1973). The test was performed with optical correction for nearby vision (e.g., reading glasses), if necessary.

**Statistical Analysis**

Close visual acuity and contrast sensitivity scores were reduced to dummy variables for further analysis, based on the frequency distribution of these variables. Thus, 13.8% of the participants were classified as having reduced close visual acuity (score > .5), and 25.9 and 15.8% showed reduced or weak contrast sensitivity (scores 4 and < 4), respec-
Associations between demographic, visual, and cognitive variables were analyzed in several hierarchical regression procedures. Demographical variables (age, sex and educational level) were entered together in the first step of all analyses, and visual function parameters in one block thereafter. Associations between the same variables and errors made on the test were analyzed by logistic regression (outcome: 0 or 1 or more errors). Between group comparisons of individuals with color weakness and normal color vision (matched for age, sex, and educational level) were performed with paired t-tests (speed measures) and Pearson’s $\chi^2$-analysis (errors). Analyses were performed with the SPSS for Macintosh program series with significance levels of $p \leq .05$.

RESULTS

The prevalence of reduced visual function as a function of age is illustrated in Figure 1. Visual acuity and contrast sensitivity decreased steadily with age, in contrast with the low but stable prevalence of color weakness over all age groups, ranging from 0 to 5.5%. In total 25 out of 424 men (6%) had a color weakness defect (red/green type), a proportion which was slightly lower than expected from population estimates (8%; Uvijls, 1998).

Table 1 summarizes the mean Stroop performance and interference scores by age level: all linear age trends in performance were significant at $p < .001$ level (cf. the regression models in Table 2).

The results of the regression of Stroop performance on demographic and visual function parameters are displayed in Table 2. As expected, higher educational level and lower age were associated with better performance for all measures. Women were faster on card 3. The additional variance explained by visual parameters was mainly accounted for by distant visual acuity, in that lower acuity was associated with better performance on card 3 and with less interference. In addition, low contrast sensitivity predicted reduced card 1 performance and color weakness was associated with lower scores on card 2. Introduction of visual function variables significantly increased the prediction (adjusted $R^2$) of the four regression models by 1.5, 1.3, 1.6 and 1.3% for performance on cards 1, 2, 3 and the interference score, respectively. The total variance explained for each Stroop parameter is

Fig. 1. Prevalence of eye function deficits by age level.
shown at the bottom of Table 2. Additional analyses were performed to investigate the decrease in age-related variance in Stroop performance after parameters of visual function were taken into account. The contribution of calendar age to the prediction of performance was greatly reduced when visual function parameters were entered into the models first, with values ranging from 55 (card 1) to 47% (interference). This indicates that approximately half of the age-related variance in Stroop performance could also be accounted for by the variance associated with

Table 2. Regression Results After Prediction of Stroop Performance from Demographic Variables (Age, Sex, Educational Level) in Step 1 and the Additional Set of Visual Function Parameters in Step 2. The Regression Coefficient $B$ and its Standard Error, the Associated $p$-value (Indicated only when $< .1$) and the Total Variance in Test Performance Explained After Each Step.

<table>
<thead>
<tr>
<th>Stroop parameter</th>
<th>Card 1</th>
<th>Card 2</th>
<th>Card 3</th>
<th>Interference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$B$</td>
<td>$(SE)$</td>
<td>$p$</td>
<td>$B$</td>
</tr>
<tr>
<td><strong>Step 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (range 52–84)</td>
<td>.26</td>
<td>(.03)</td>
<td>.00</td>
<td>.39</td>
</tr>
<tr>
<td>Sex (0: male/1: female)</td>
<td>$- .15$</td>
<td>(.52)</td>
<td>–</td>
<td>$- 1.10$</td>
</tr>
<tr>
<td>Education (range 1–8)</td>
<td>–1.02</td>
<td>(.15)</td>
<td>.00</td>
<td>$- 1.20$</td>
</tr>
<tr>
<td>$R^2$ intermediate model</td>
<td>.156</td>
<td></td>
<td></td>
<td>.162</td>
</tr>
<tr>
<td><strong>Step 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (range 52–84)</td>
<td>.21</td>
<td>(.03)</td>
<td>.00</td>
<td>.33</td>
</tr>
<tr>
<td>Sex (0: male/1: female)</td>
<td>–.07</td>
<td>(.52)</td>
<td>–</td>
<td>–.98</td>
</tr>
<tr>
<td>Education (range 1–8)</td>
<td>–1.01</td>
<td>(.15)</td>
<td>.00</td>
<td>$- 1.18$</td>
</tr>
<tr>
<td>Acuity: distant</td>
<td>–1.07</td>
<td>(.92)</td>
<td>–</td>
<td>$- 2.30$</td>
</tr>
<tr>
<td>Acuity: close (0/1)</td>
<td>1.35</td>
<td>(.82)</td>
<td>–</td>
<td>.32</td>
</tr>
<tr>
<td>Contrast: medium (0/1)</td>
<td>–.53</td>
<td>(.69)</td>
<td>–</td>
<td>$- .94$</td>
</tr>
<tr>
<td>Contrast: low (0/1)</td>
<td>1.93</td>
<td>(.87)</td>
<td>.03</td>
<td>2.05</td>
</tr>
<tr>
<td>Color weakness (0/1)</td>
<td>1.97</td>
<td>(1.46)</td>
<td>–</td>
<td>4.34</td>
</tr>
<tr>
<td>$R^2$ overall model</td>
<td>.171</td>
<td></td>
<td></td>
<td>.175</td>
</tr>
</tbody>
</table>

*Note.* $^1$Code ‘0’ indicates that the function not diminished, and ‘1’ that function is diminished.
this set of visual parameters. Addition of specific interaction terms for visual function parameters with age or color weakness with other visual function parameters did not improve the predictive power of any of the four models (results not shown in the table).

The same set of predictors was used to predict errors on Stroop performance with logistic regression, using the dichotomous variable no errors/one or more errors (0/1) as outcome measure. Making errors on card 2 was associated with older age ($B_{\hat{}} = .025$, Wald (1) $\hat{=} 4.07$, $p_{\hat{}} = .044$); errors on card 3 with older age ($B_{\hat{}} = .037$, Wald (1) $\hat{=} 11.39$, $p_{\hat{}} = .001$), lower education ($B_{\hat{}} = -.254$, Wald (1) $\hat{=} 25.37$, $p < .001$), and male sex ($B_{\hat{}} = -.399$, Wald (1) $\hat{=} 5.76$, $p = .016$). Visual function parameters, however, did not significantly improve the prediction of errors.

To evaluate the performance of the 25 individuals with color weakness in more detail, we matched this group by age and educational level to men with normal color vision (Table 3). Both groups had the same visual acuity. In line with results from the regression analysis the men with color weakness performed significantly worse on card 2 only, although a trend towards lower performance on other Stroop parameters was apparent. $\chi^2$-analysis of errors in these groups (0 = no errors, 1 = one error, 2 = two or more errors) showed that men with defective color vision made more errors on card 3 (Pearson’s $\chi^2$ (df $\hat{=}$ 2) $\hat{=} 10.9$, $p < .01$), with no differences in the number of errors made on the other two cards.

DISCUSSION

We studied to what extent performance of the classic Stroop color-word test of normal aging individuals with no overt reading difficulties is determined by age-related visual factors. It was found that calendar age and educational level were the most robust predictors of both speed and errors. Individuals with low distant visual acuity were slower on Stroop card 3 – an effect of 1 unit on the acuity scale was comparable to that of being 7.5 years older (cf. $B$ coefficients of these variables in Table 2). Optimal contrast sensitivity was more important for performance on card 1 – individuals with low contrast sensitivity were slower on this task. A small but significant association was found between red/green color weakness (present in 25 male participants) and slower performance on color patch identification (card 2). In addition, these individuals made more errors on card 3 than did matched controls with normal color vision. Compared to the contribution of demographic variables the unique variance explained by visual parameters was relatively small, ranging from 1.3 to 1.6%. However, half of the variance in Stroop performance explained by age alone could also be explained by visual function variables. Combined with earlier observations which have demonstrated reduced performance on tests that rely on visual processing in individuals with impaired visual function, this finding suggests that some of the effect of reduced visual function probably is incorporated in the effect of calendar

Table 3. Mean (+SD) Background Characteristics and Stroop Performance in 25 men With red/green Color Weakness, Compared With Age- and Education-Matched Controls.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Color Weakness</th>
<th>$t^1$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yr)</td>
<td>64.2 (8.1)</td>
<td>64.3 (8.1)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Education (yr)</td>
<td>4 (2)</td>
<td>4 (2)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Acuity: distant</td>
<td>1.11 (.29)</td>
<td>.96 (.42)</td>
<td>–1.65</td>
<td>–</td>
</tr>
<tr>
<td>Card 1 (sec)</td>
<td>44.5 (7.8)</td>
<td>48.1 (8.3)</td>
<td>1.77</td>
<td>–</td>
</tr>
<tr>
<td>Card 2 (sec)</td>
<td>59.1 (9.3)</td>
<td>65.4 (11.2)</td>
<td>2.23</td>
<td>.04</td>
</tr>
<tr>
<td>Card 3 (sec)</td>
<td>108.0 (36.5)</td>
<td>117.8 (32.0)</td>
<td>1.03</td>
<td>–</td>
</tr>
<tr>
<td>Interference (sec)</td>
<td>48.8 (30.7)</td>
<td>52.5 (24.4)</td>
<td>.46</td>
<td>–</td>
</tr>
</tbody>
</table>

Note. $^1$Paired $t$-test, 2-tailed, df $\hat{=} 24$; only $p \leq .05$. 
As a result, the actual impact of parameters related to visual function may be larger than what is implied by the increase in explained variance in step 2 of the analysis (Table 2). This may be particularly relevant for the interpretation of Stroop results of individual subjects with visual dysfunctions, e.g., many older people, in whom the measurement error in Stroop performance is directly related to the extent of visual function decrements.

Although, the effect of defective color vision on Stroop performance is a subject of much speculation, this is to our knowledge the first study that addresses this issue directly. It appears that red/green color weakness increases the time needed to complete card 3, and the number of errors made. These individuals performed at the same level as comparable individuals with no color weakness who were 13 year older (Table 2). Color weakness did not change the performance of individuals on card 3 and also did not affect the interference score. Thus, the most common defect of color vision seems to have only a moderate effect on Stroop performance, but in our view this finding still calls for adequate assessment of color vision in situations when a deficiency is suspected.

The reduced performance on card 1 in individuals with low contrast sensitivity did not come as a great surprise because this aspect of visual function is vital for stimulus identification. This effect was comparable to 9 calendar years (Table 2). Visual acuity and contrast sensitivity are to some extent interrelated, as is reflected by the overall increase in association between contrast sensitivity and performance when visual acuity was omitted from the regression models (results not shown). The importance of adequate visual acuity for an unbiased estimate of Stroop performance is obvious, but also to some extent puzzling with regard to card 3 performance. It could be argued that reduced acuity may put a respondent at an advantage when degraded color words due to the visual impairment interfere less with the task to name ink colors. This is, however, not the case for individuals with normal reading ability or is at least overshadowed by the importance of adequate acuity for ink color detection.

Participants in this study were not screened by an ophthalmologist for specific ocular diseases. All individuals reported that they did not have impairments of visual function that interfered with reading ability. However, many prevalent ocular diseases are typically degenerative and thus age-related in nature (e.g., clouding of the lens in cataract, or stiffening of the lens, resulting in presbyopia). They have an insidious onset and may go unnoticed for a long period. In fact, intrinsic aging of the ocular system was considered one of the primary variables in this study. We felt that specialized screening for ocular disease was superfluous as it is the actual (remaining) function that matters in daily life.

It could be argued that the effect of different aspects of visual function on cognitive test performance should be tested by manipulating the physical properties of stimulus material, in order to control extra-experimental variability more rigorously. However, the effect of the continuous adaptation to age-related changes in sensory function must be taken into account as this process may partly compensate for a reduced quality of stimulus material. We consider this study as the best approximation of the situation in daily neuropsychological practice in which individuals with an unknown visual status need to be evaluated.

This study illustrates that basic information about the visual status is of theoretical and practical importance to the neuropsychologist. On the basis of our results, we expect that Stroop interference scores are likely to be underestimated in individuals with low visual acuity. The impact of color weakness is moderate but emphasizes that it is necessary to identify individuals with true color blindness before a score can be interpreted. We therefore suggest that a routine screening of both color vision (subjective) and visual acuity (objective) should be part of an assessment with the Stroop test. This information can be used to adjust the clinical interpretation of performance scores, e.g., by increasing the index age for use in normative reference tables proportionally to the amount of reduced visual function, as outlined above. In any case, the outcome on the Stroop test should be interpreted cautiously in older people of whom the visual status has not been ascertained.
REFERENCES


