California Older Adult
Stroop Test (COAST):
Development of a Stroop Test
Adapted for Geriatric Populations

Nancy A. Pachana, PhD
Larry W. Thompson, PhD
Bernice A. Marcopulos, PhD
Ruth Yoash-Gantz, PsyD

ABSTRACT. An adaptation of the traditional Stroop test, the California Older Adult Stroop Test (COAST) (Pachana, Marcopulos, Yoash-Gantz & Thompson, 1995), has been developed specifically for use with a geriatric population, utilizing larger typeface, fewer items (50) per task, and more easily distinguished colors (red, yellow and green). Test-retest reliability and validity data are reviewed for both control and clinical populations. Increased error rates on the Stroop test compared to the COAST were found for the color and color/word interference tasks. These results are discussed in terms of changes in the visual system with increasing age. The implications for better test sensitivity with the COAST for older adult populations are discussed. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <docdelivery@haworthpress.com> Website: <http://www.HaworthPress.com> © 2004 by The Haworth Press, Inc. All rights reserved.]*

Nancy A. Pachana is affiliated with the University of Queensland, Brisbane, Australia.

Larry W. Thompson is affiliated with Pacific Graduate School of Psychology, Palo Alto, CA, USA.

Bernice A. Marcopulos is affiliated with the University of Virginia, Charlottesville, VA, USA.

Ruth Yoash-Gantz is affiliated with Salisbury VAMC, Salisbury, NC, USA.

Test materials and preliminary normative data for the COAST can be obtained from the second author.

Address correspondence to: Larry W. Thompson, 2049 Fallen Leaf Lane, Los Altos, CA 94024.

INTRODUCTION

More than 50 years ago, Stroop (Stroop, 1935) developed the Color-Word Test which has been used to assess selective attention, the ability to shift perceptual set and the ability to inhibit habitual responses. This methodology has proven useful in clinical settings, and over the years a multitude of Stroop-type tests have been introduced. While most of these are labeled as a particular modification of the Stroop test, the numerous versions of this assessment strategy may have varying stimuli, administrative guides and scoring procedures. Some are more commonly used than others, but at the present time, no one type has gained prominence, either clinically or experimentally. One feature that is common among many versions of the Stroop is the task requirement to distinguish between the colors blue and green in order to make an appropriate response. It has been known for some time that the elderly have difficulty in discriminating the blue-green dyad when completing a Stroop-type test (Comalli, 1965), and this factor may heavily influence their performance level. This concern prompted us to design a Stroop-type task more considerate of age-related changes in cognitive processes.

A number of changes in cognitive functioning are associated with increasing age in normal, healthy adults, including changes and/or declines in attentional skills (Salthouse, Toth, Hancock & Woodard, 1997), speed of performance and color/word interference (Filley & Cullum, 1994; Salthouse & Meinz, 1995) and executive functions (Brennan, Welsh & Fisher, 1997). However, substantive decline is not inevitable, even in persons in their ninth and tenth decade (c.f. Benton, Eslinger & Damasio, 1981). Unfavorable premorbid variables such as low educational or intellectual achievement (Bowles & Poon, 1982) may affect later cognitive performance. Likewise, systemic medical conditions (e.g., diabetes), neurogenerative diseases (e.g., Alzheimer's disease) and psychiatric conditions (e.g., depression) may cause either temporary or permanent declines in a number of aspects of cognitive functioning. Although discriminating between likely causes of cognitive dysfunction in the elderly is often challenging, it is important, particularly when competence or forensic issues are raised. Thus in designing the present test, the authors hoped to create an instrument sensitive

to changes in cognitive function, but only minimally influenced by normal aging changes (e.g., color perception, increased vulnerability to fatigue).

Age normative data have been collected for at least two commonly used versions of the Stroop test; the Stroop Color and Word Test (Golden, 1978) and the Victoria version of the Stroop test (Spreen & Strauss, 1998). Yet both require subjects to make blue/green distinctions; one is quite long (100 items) and printed in a very small typeface.

The observation that some older subjects have trouble distinguishing green and blue on the Stroop test had been made as far back as Comalli's series of experiments using the Stroop in an older cohort (Comalli, 1965; Comalli, Krus & Wapner, 1965). Older subjects find it more difficult to discriminate colors at the shorter end of the light spectrum (Botwinick, 1973) secondary to yellowing of the crystalline lens (Bando, Ishii & Nakajina, 1976; Weale, 1965). The presence of cataracts may also hinder accurate distinction of blue from green in older adults due to the yellow cast caused by the cataract itself. Other problems within the visual system, such as retinal degeneration, clouding of intraocular fluid or loss of accommodation may impair visual acuity. Such factors argue that age-related changes in performance on the Stroop test may be due to peripheral factors rather than changes in complex cognitive capacities. Evidence pertaining to this argument could have direct impact on the interpretation of age changes observed on the Stroop test.

Development of the COAST

A revision of the Golden (1978) version of the Stroop was created by the authors using more easily distinguished colors (red, yellow and green), larger type face (pitch #24) and fewer (50) items. It was felt that these changes would minimize color confusion and fatigue in older adults, while still preserving the basic integrity of the test. The revised version in particular was designed to minimize errors due to decreased visual acuity or color-naming difficulty secondary age-related changes to the peripheral visual system.

Conceptually, the COAST is quite similar to other traditional versions of the Stroop test. Administration is similar to the Golden version of the Stroop (Golden, 1978), with the cards presented in the following order:

1. colors;
2. words;
3. color/word interference.

The words or colors are read across from left to right. Total time to complete the tasks is recorded, as in the Strickland, D'Elia, James and Stein (1997) administration instructions. Self-corrected errors and total number of errors are also recorded.

During the past few years, several studies using convenience samples have been undertaken to evaluate the role of blue/green confusion in age-related changes, to determine psychometric properties of the COAST, and to examine the performance of normal elders and select patient populations on the test.

Preliminary reliability data on the COAST have been published elsewhere (Pachana, Marcopulos, Yoash-Gantz & Thompson, 1997). For convenience we reproduce general findings on reliability and practice effects here (adapted from Pachana et al., 1997), along with previously unpublished data. This new data has been organized sequentially in several sections to explore the following points:

a. the influence of physiological changes affecting visual acuity and perceptual accuracy on performance declines in the elderly on the Stroop test;
b. issues of reliability and validity in the COAST;
c. the applicability of the COAST in the assessment of select clinical populations, in terms of group differences and absolute levels of impairment.

SUBJECTS, METHODOLOGY AND PROCEDURES

The *California Older Adult Stroop Test (COAST)* was administered to 6 groups of subjects:

a. demented inpatients without major medical problems;
b. non-insulin-dependent Type II diabetic outpatients in good general health;
c. psychiatric inpatients (diagnoses included schizophrenia, affective psychosis and chronic substance abuse);
d. depressed outpatients;
e. caregivers of dementia patients; and
f. age-matched normal community volunteers.

Subjects were obtained from a variety of ongoing research studies conducted at the Palo Alto Veteran's Administration Health Care Service

and the Stanford University Medical Center. A physician saw all subjects and/or reviewed their medical records. A senior psychologist also saw subjects, but testing was conducted in the main by psychology and psychiatry interns, post-doctoral fellows and residents. Persons with chronic medical conditions (other than non-insulin dependent Type II diabetes in the diabetic group), chronic psychiatric conditions (other than the specific diagnostic groups named) or a history of head trauma were excluded from the study.

All subjects in the various experimental groups were administered the COAST, along with a number of other neuropsychological tests which varied depending on the nature of the research study. The Golden version of the Stroop was administered to a select group of healthy diabetics and elderly community volunteers with no diabetes. Three scores were obtained for each task of the COAST: total time to completion, total self-corrected errors, and total uncorrected errors. Similar scores were obtained for the Golden version of the Stroop.

RESULTS

Empirical Data in Support of Blue-Green Confusion on the Stroop Test

In order to examine the hypothesis that performance on the Golden Stroop test is affected by blue-green color confusion in older adults, 33 . diabetics and 18 normal controls were administered both the Golden Stroop and the COAST. The diabetic group was comprised of type II non-insulin-dependent diabetics who were self-reported to be in good health; the normal controls consisted of elderly community volunteers in good health that did not have diabetes. The Stroop and COAST tests were administered approximately three hours apart. The tests were counterbalanced to minimize order effects. The diabetic and normal control groups did not differ on either time to completion or error rates on either the Stroop or COAST. Thus, these two groups were combined for all subsequent analyses.

Of the 51 subjects administered the Stroop, 10 individuals (20% of the sample) were unable to distinguish between the colors blue and green on the color naming task, and 9 of these same individuals were unable to make distinctions between blue and green on the color/word interference task. Difficulty in distinguishing blue from green was ascertained by self-reports from the subjects, either offered by the subjects

themselves or obtained via direct questioning. All of the 51 subjects were able to complete the word reading task of the Stroop, and all tasks of the COAST.

Incorrect items on all three tasks of both the Stroop and the COAST were recorded for each subject. The number of errors per color per task was calculated for both the color and color/word interference tasks of both tests. It was found that subjects made a high proportion of errors on the Stroop on the blue and green items for both the color naming and color/word interference tasks. By contrast, a much lower error rate was found for the red items on the Stroop, and a similar low rate of errors was found across the red, green and yellow items of the COAST (see Table 1).

On the Golden Stroop, relatively high correlations were found between time to completion and total errors on the color task ($r = 0.45$) and time to completion and total errors on the color/word interference task

TABLE 1. Percentage of Errors by Colors on the Stroop and COAST Tests in Type-II Diabetics and Community Volunteers Who Self-Report Good Health

Stroop Color Card (N = 41)			
	Red Items	Green Items	Blue Items
Number of items	33	33	34
M (SD)	0.8 (2.0)	16.9 (16.7)	10.8 (17.3)
Range	0.0-9.1	0.0-63.6	0.0-94.1
Stroop Color/Word Interference Card (N = 42)			
	Red Items	Green Items	Blue Items
Number of items	33	33	34
M (SD)	2.9 (3.1)	9.0 (14.2)	7.8 (10.8)
Range	0.0-9.1	0.0-69.7	0.0-47.1
COAST Color Card (N = 51)			
	Red Items	Green Items	Yellow Items
Number of items	18	16	16
M (SD)	1.5 (3.1)	0.3 (1.4)	0.8 (2.1)
Range	0.0-11.1	0.0-6.3	0.0-6.3
COAST Color/Word Interference Card (N = 51)			
	Red Items	Green Items	Yellow Items
Number of items	18	16	16
M (SD)	3.2 (4.5)	4.5 (5.2)	1.6 (3.7)
Range	0.0-16.7	0.0-18.8	0.0-12.5

($r = 0.66$). In contrast, for the COAST, relatively low correlations were found between time to completion and total errors on the color task ($r = -0.09$) and time to completion and total errors on the color/word interference task ($r = 0.22$). The difference between correlations for the Stroop and the COAST was significant at $p < .01$ (N = 41) for both the color and the color/word interference tasks, using "r-to-z" transformations. Thus, subjects who committed many errors on the color and color/word interference tasks of the Stroop generally took more time to complete these tasks, but a similar relationship was not evident on the COAST.

The tendency for subjects to commit errors on the blue and green items of both the color naming and color/word interference tasks of the Stroop, and to take longer to complete the task if they were committing many errors, could be due to problems in central or peripheral processing, or both. It could be that more impaired performance with respect to errors was due to cognitive impairment. An alternate hypothesis is that a peripheral factor, such as blue-green color confusion, stemming from changes to the visual system with age, could be at work.

Two strategies were used to explore these questions. First, subjects were split into low and high performance groups based on a median split of their numbers of errors on the Stroop color task. Ten subjects were dropped from the analyses due to missing data. The groups were then compared on their respective time to completion on the Stroop and COAST color naming tasks. A two-sample t-test revealed that the high error-rate group took more time to complete the task than the low error-rate group ($t = -2.81$, $df = 30.5$, $p < .05$). In contrast, no significant group differences emerged on time to complete the color task of the COAST ($t = -1.04$, $df = 38.4$, $p < .31$). Similarly, group differences were found on the color/word interference task of the Stroop ($t = -2.21$, $df = 26.8$, $p < .04$), such that the high error-rate group took more time to complete the task than the low error rate group, whereas on the COAST, no significant group differences emerged on time to complete the color/word interference task ($t = -1.96$, $df = 28.7$, $p < .06$). Thus, subjects who took longer on the Stroop, while also making more errors, appeared not to repeat this pattern on the COAST. These findings suggest a peripheral rather than a central mediating factor.

In order to approach this question from yet another perspective, a second analytic strategy was adopted, which involved an attempt to partial out of the relationship other indices of central processing. Other cognitive measures available for all subjects, including the Digit Span and Digit Symbol subscales of the WAIS (Wechsler, 1955); time to

completion on Trail Making Tests (Spreen & Strauss, 1998); and immediate and delayed recall on the Logical Memory and Visual Reproductions of the WMS, Form I (Wechsler, 1945) were utilized. The large number of variables, along with high intercorrelations of these measures suggested collinearity problems, and prompted us to reduce these data by completing a principle components analysis. Two factors, processing speed and learning and memory, accounted for 67% of the variance. Factor scores were calculated and entered as independent variables along with age and number of errors on the Stroop color task in a regression model to predict performance time. Tolerance for the two factors was unacceptably low and diagnostics showed high collinearity between the two factors along with degradation of the processing speed factor (Wilkinson, Blank & Gruber, 1996). This variable was removed, and a sequential regression was completed. Age and the learning and memory factor were entered in model 1 with performance time on the color task as the dependent variable. The color error rate was entered in model 2. The entry of age and memory factors in model 1 was not significant $(F(2,31) = 0.77, p = .47)$. Entry of the color task as the dependent variable in model 2 yielded an R^2 change of .156 $(F(1,30) = 5.88, p = .02)$.

An identical regression analysis was completed, with the performance time on the color/word interference task as the dependent variable and age, memory factor, and error rate on the color/word interference task as the independent variables. Again, the entry of the age and memory factor in model 1 was not significant $(F(2,32) = 2.25, p = .12)$, whereas the R^2 change of .274 when the error rate was entered in model 2 was highly significant $(F(1,30) = 14.10, p = .001)$. The absence of an age and memory effect coupled with evidence of a strong association between performance time and error rate when age and memory effects are accounted for also is consistent with the argument that blue/green color confusion on the Stroop in older adults, who self-report good health, is in large measure a peripheral phenomenon.

Reliability and Practice Effects in an Impaired Population

Test-retest reliability, validity, and practice effects on the COAST have been reported elsewhere (Pachana et al., 1997). In the present study additional Type-II diabetic and community volunteers, who self-reported that they were in good health were added to the earlier data, and some psychometric properties were re-evaluated. Table 2 gives the correlations for the three tasks of the Stroop and the COAST.

In the total sample, the correlation between the two measures for the color naming task was low, although the correlations for word and color/word interference tasks were acceptable. Examination of the scatterplot indicated that outliers on number of errors were attenuating the relationship, and this was most likely due to color-confusion. Outliers on number of blue-green errors on the Stroop and/or subjects who complained of difficulty in making the distinction were removed and the correlations were obtained on the reduced sample. Table 2 also shows that the correlation between the Stroop and the COAST improved on both tasks requiring blue-green distinctions on the Stroop after these subjects were deleted. In this second set of analyses, the COAST appeared favorably with the Stroop.

Test-retest correlations were obtained for the COAST at three different time points spaced at monthly intervals. These analyses were completed on Type-II diabetics and community volunteers who self-reported that they were in good health. There were no differences between the diabetics and the community volunteers, so the two groups were combined. Table 3 reports the correlations among the three time intervals for time to completion on the three tasks. Since so few errors were made, these data are not reported. The correlations range from .711 to .907 (N = 90) indicating good test-retest reliability for time to completion.

Practice effect was also evaluated in this sample. Means and SDs for the three intervals are reported in Table 4. A two (diabetic vs. volunteer) by three (time of measurement) MANOVA for repeated measures was used to evaluate practice effects in the two groups. No group or group by time interaction effects were found in any of the analyses. Only the color/word interference task showed a significant time effect (Wilks'

TABLE 2. Relationship Between the Stroop and COAST for Time to Completion in Type-II Diabetics and Community Volunteers Who Self-Report Good Health: Correlations Obtained Before and After Outliers for Errors Were Removed

Task	Total Sample (n = 41)	Outliers Removed (n = 31)
Color	.332	.547*
Word	.737**	.727**
Color/Word interference	.887**	.913**

* $p < .05$; ** $p < .001$.

TABLE 3. Test-Retest Correlations for the COAST in Type-II Diabetics and Community Volunteers Who Self-Report Good Health: Completion Time for the Color, Word and Color/Word Interference Tasks (N = 90 for all three tasks)

Time of Testing[1,2]	Time 1	Time 2	Time 3
		Color	
Time 1	–	.843	.788
Time 2		–	.907
Time 3			–
		Word	
Time 1	–	.796	.801
Time 2		–	.852
Time 3			–
		Color/Word interference	
Time 1	–	.838	.711
Time 2		–	.721
Time 3			–

[1] One-month interval between each testing.
[2] All correlations are significant at $p < .001$.

Lambda = .858; F, estimate $(2,87) = 7.02, p = .001$). Univariate analyses showed this effect to be due to the decrease in time to completion from month one to month two $(F(1,88) = 7.79, p = .006)$. The change from month two to month three was not significant $(F(1,88) = 1.48, p = .227)$. Although the decrease in time was statistically significant across a one-month period, the mean change was roughly 2.4 seconds (3.7%), and only 1.3 seconds (2.1%) across the second month. Thus, in general the performance of this sample of elderly subjects appeared to be reasonably stable across time.

Our interest in psychometric properties extends to cognitively impaired populations, given reports of increased color confusion effects in such patients. Fisher, Freed and Corking (1990) found performance on the Golden version of the Stroop test not only declined with age, but also correlated with the presence and severity of dementing illness, especially for the color/word interference task. In their study, 22 percent of demented patients confused the colors blue and green. Cohen, Cronin-Golomb, Growden and Corkin (1988) administered a color vision test to Alzheimer's disease (AD) patients and healthy age-matched controls. More overall errors were made by the dementia patients than the control

TABLE 4. Means and Standard Deviations for the COAST at Three Times of Testing for Type-II Diabetics and Community Volunteers Who Self-Report Good Health (N = 90)

Time of Testing	Time 1 Mean	SD	Time 2 Mean	SD	Time 3 Mean	SD
			Color Task			
Time to completion	32.233	6.104	33.233	6.832	33.022	6.702
Uncorrected errors	0.078	0.308	0.089	0.486	0.033	0.181
Self-corrected errors	0.522	0.782	0.444	0.809	0.622	0.869
			Word Task			
Time to completion	24.344	4.806	24.611	4.663	24.411	4.571
Uncorrected errors	0.056	0.275	0.056	0.275	0.033	0.181
Self-corrected errors	0.178	0.413	0.167	0.431	0.156	0.539
			Color/Word Interference Task			
Time to completion	64.189	16.274	61.822	16.511	60.544	14.157
Uncorrected errors	0.478	1.256	0.311	0.979	0.378	1.023
Self-corrected errors	1.189	1.498	0.800	1.376	0.800	1.351

group. The dementia patients appeared to have particular difficulty distinguishing between blue and green stimuli. This difficulty with color discrimination led to 15% of AD patients making blue-green discrimination errors, with only 6% of controls making similar errors (Cohen et al., 1988).

To further examine test-retest reliability of the COAST with an impaired population, 15 individuals with degenerative dementia (age range, 73 to 81), who had been residing in a long-term care ward for at least one year were administered the COAST on two consecutive days. The subjects were counter-balanced, with half seen first at 10:00 AM, then at 2:00 PM on the following day for the second testing, and the other half seen first at 2:00 PM, then at 10:00 AM on the following day, resulting in a 30 or 20 hour testing interval respectively. There were no differences due to time of testing, so the two groups were combined. Correlations for time to completion, between the administrations of the COAST at Time 1 and Time 2 were quite high, ranging from .82 to .95 (see Table 5). Correlations for uncorrected errors, and self-corrected errors ranged from .47 to 1.00 on the three tasks. There was no evidence of a practice effect in these data. Although the sample is small, these data suggest adequate test-retest reliability in dementia patients whose cognitive function is stabilized.

TABLE 5. Test-Retest Correlations in Dementia Inpatients: Completion Time and Errors (N = 15)

Condition[1]	Trials		
	Completion Time	Self-Corrected Errors	Uncorrected Errors
Color AM vs PM	.95	.78	.64
Word AM vs PM	.82	.87	1.00
Interfer. AM vs PM	.85	.47	.89

[1] Correlations significant at $p < .01$.

Performance of Clinical Populations on the COAST

The suitability of this measure for use with older populations is supported by data presented in the previous sections. The purpose of this section is to ascertain whether subjects from a variety of clinical diagnostic groups differ in their performance on this measure. Five distinct groups of clinical patients and a group of healthy elderly volunteers were selected for further evaluation with this measure. Diabetic subjects were recruited from outpatients who were participating in a range of clinical trials evaluating new treatment regimes. The depressed group was comprised of outpatients with a diagnosis of major depressive disorder seeking treatment for their condition. Neuropsychiatric inpatients with a diagnosis of schizophrenia or bipolar disorder with psychotic features comprised the psychiatric group. The dementia sample was recruited from an inpatient dementia ward population, and met DSM-IV criteria for a dementing disorder. Community volunteers were recruited by advertisements for participation in clinical research studies, and were self-reported to be in good health. A group of high functioning elderly female caregivers who reported that they were experiencing stress related to their caregiving duties was also recruited in the study.

Mean and standard deviations for these six subject groups on demographic variables, and their scores on selected neuropsychological tests are presented in Table 6. Means and standard deviations for the groups on total time to completion, number of self-corrected errors, and number of total errors for each of the three tasks of the COAST (color-naming, word-reading, and color/word interference) are given in Table 7.

One-way ANOVA revealed overall significant differences among the six groups for age ($F(5,404) = 20.293$, $p < .000$ and education

TABLE 6. Means and Standard Deviations of Sociodemographic and Select Cognitive Variables for Experimental Groups

	Experimental Groups											
	Type-II Diabetics		Depressed Outpatients		Psychiatric Inpatients		Demented Inpatients		Community Volunteers		Family Caregivers	
N	124		39		38		51		86		72	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Age	67.78	5.5	66.74	6.5	69.08	5.6	77.00	7.7	69.44	5.5	63.43	12.5
Education	14.44	2.9	14.23	2.5	11.26	2.7	10.41	3.1	15.29	2.9	13.96	3.2
Occupation	3.37	1.6	3.10	1.1	4.82	1.3	4.88	1.2	3.22	1.4	—	—
Digit Span (WAIS-R)	9.15	2.4	8.92	2.3	5.77	2.2	5.25	1.0	9.72	2.7	—	—
Vocabulary (WAIS-R)	11.45	2.6	11.53	3.3	7.53	3.0	6.00	2.3	12.83	3.0	—	—
Similarities (WAIS-R)	9.21	2.4	9.74	2.3	6.50	2.2	4.67	2.1	10.05	3.7	—	—
Digit Symbol (WAIS-R)	6.84	2.1	7.42	2.3	3.20	0.9	3.50	2.1	7.52	2.2	8.1	2.4
Trials B (time)	98.68	50.6	107.95	71.9	231.84	77.7	292.77	57.4	96.08	44.2	—	—
Logical Memory (WMS)	17.66	5.9	21.05	7.3	12.81	6.0	9.61	5.4	18.92	6.5	—	—
Logical Memory Delayed (WMS)	13.02	6.2	16.21	7.7	7.20	6.0	4.22	3.8	14.01	6.6	—	—
Visual Reproduction (WMS)	11.78	8.3	8.33	3.1	12.24	7.9	6.73	7.2	11.17	7.3	—	—
Visual Repro. Delayed (WMS)	10.77	7.3	6.49	3.4	4.18	4.9	0.60	1.1	10.05	6.60	—	—
Global Cognitive Functioning	—	—	—	—	125.20	6.0[a]	107.72	14.1[a]	—	—	29.0	1.1[b]

[a] Mattis Dementia Rating Scale scores.   [b] Mini-Mental State Examination score.

TABLE 7. Means and Standard Deviations on the COAST for Experimental Groups

	Type-II Diabetics 124		Depressed Outpatients 39		Psychiatric inpatients 38		Demented Inpatients 51		Community Volunteers 86		Family Caregivers 72	
N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Color-time	33.55	6.8	37.80	17.1	53.50	23.5	106.55	72.1	31.70	6.3	31.36	9.0
Color-uncorrected errors	0.09	0.4	0.10	0.4	0.68	1.4	3.51	6.5	0.14	0.4	0.08	0.4
Color-self-corrected errors	0.43	0.7	0.39	0.7	1.00	1.5	1.71	1.9	0.38	0.7	0.20	0.5
Color-uncorrected errors	0.09	0.4	0.10	0.4	0.68	1.4	3.51	6.5	0.14	0.4	0.08	0.4
Word-time	24.44	4.3	26.08	8.5	36.21	14.9	63.43	34.3	23.66	4.8	23.9	5.6
Word-uncorrected errors	0.06	0.3	0.03	0.2	0.13	0.4	1.02	1.8	0.02	0.2	0.01	0.1
Word-self-corrected errors	0.20	0.5	0.10	0.3	0.29	0.6	0.43	1.0	0.06	0.2	0.12	0.4
Interfer-time	71.44	23.5	72.77	26.4	107.40	42.8	179.10	70.6	65.56	21.1	58.50	20.3
Interfer-uncorrected errors	0.84	1.9	0.44	0.8	7.11	9.8	17.80	14.2	-0.59	1.2	1.78	2.1
Interfer-self-corrected errors	1.30	1.6	0.92	1.2	3.29	2.8	2.86	3.1	1.02	1.6	0.88	1.1

$(F(5,404) = 24.975, p < .000$. Occupation was compared for all groups except caregivers. ANOVA revealed a significant overall difference $(F(4,333) = 20.698, p < .000$. Post hoc tests using Tukey's HSD were completed to evaluate differences among all the groups. As shown in Table 3 the overall difference in age is due to the fact that the dementia group is significantly older and the caregiver group is significantly younger than the remaining subject groups $(p < .001$ in all comparisons that were significant). The overall effect in education was due to the fact that the dementia group and the inpatient psychiatry group had less education than the other four groups $(p < .001$ for all significant comparisons). There was no difference between the dementia group and the inpatient psychiatry group, or among the other four groups. Table 6 shows that the occupation difference was due to the fact that the dementia group and the inpatient psychiatry group had lower occupational levels than the other three groups $(p < .001$ for all comparisons). No other comparisons were statistically significant.

These data reflect substantive differences among the groups on sociodemographic variables that are often related to performance on Stroop-type tests. For example, in our study the relationship between performance time and education was highly significant for all three tasks $(r(410) = .361, .416$ and $.439$ for color naming, word reading and color/word interference respectively; $p < .000$ for all three correlations). A similar pattern of significant correlations is evident for age and occupation, although the values are slightly smaller than those obtained between performance time and education. Thus, adjustment for sociodemographic factors was called for in making subsequent group comparisons on COAST measures. Because the correlation between education and occupation was high $(r(338) = .57; p < .000$) and occupation level was missing for the caregiver group, we decided to drop it in subsequent analyses.

A MANOVA was run with group as the independent variable and age and education as covariates. The dependent measures included performance time, number of self-corrected errors and total number of errors for the color naming, word reading and color/word interference tasks. The estimate of Wilks' lambda indicated a significant overall group effect on COAST performance after adjustments were made for the effects of age and education $(F(9,398) = 2.486, p = .009$). Univariate analyses revealed a significant group effect for performance time on the color naming $(F(1,406) = 6.484, p = .011$), word reading $(F(1,406) = 6.388, p = .012$), and color/word interference $(F(1,406) = 6.034, p = .014$). The total number of errors made on the color/word interference

task was also highly significant $(F(1,406) = 16.869, p < .001)$. The remaining dependent variables showed no evidence of a group effect.

Post hoc univariate comparisons, using Tukey's HSD, were made among the groups for these four variables. As shown in Table 7 the dementia group took substantially longer to complete the color naming task than all of the other groups $(p < .000$ for all comparisons). The inpatient psychiatry group also required significantly more time to complete the color naming task than the diabetic, community volunteer and caregiver groups $(p < .001$ for all three comparisons), but still took less time than the dementia group $(p < .000)$. The pattern of group differences for performance time on the word reading and color/word interference tasks was identical to the color reading task. Turning to the error rates, only the total number of errors on the color/word interference task was still significant after adjusting for age and education. As can be seen in Table 4, the pattern of differences among the groups is similar to that for performance time. The dementia group and the inpatient psychiatry group made substantially more errors than the other four groups $(p < .001$ for all comparisons), and the dementia group made significantly more errors than the inpatient psychiatry group $(p < .001)$. None of the comparisons among the other four groups was significant. Thus, after adjustments for age and education, the dementia group and the inpatient psychiatry group were substantially slower than the other groups and made considerably more errors on the color/word interference task, but the inpatient psychiatry group was still faster and made fewer errors than the dementia group.

Post-hoc analysis using Tukey (HSD) pairwise comparison method revealed the main effects to be due primarily to differences between the performance of the dementia group and the other five groups.

DISCUSSION

Blue Green Color Confusion

Both increased time to completion and increased numbers of errors on the Golden Stroop appeared linked to difficulties in distinguishing blue from green. This suggests that a peripheral rather than a central factor (namely, changes in the crystalline lens with age) are responsible in large part for this impaired performance. At multiple points in the analysis process, this peripheral effect of vision appeared more strongly on the pure color naming task than the color/word interference task. We

postulate that the increased cognitive load required to successfully complete the color/word interference task was responsible for this shift.

All subjects were able to complete the COAST version of the Stroop test, regardless of diagnostic status, in contrast to the Golden version of the Stroop, which some members of some groups did not complete. Moreover, the error rates on the COAST appeared unaffected by any color confusion. Indeed, the high standard deviations on the Stroop tasks, seen at several points during the analyses, particularly with respect to the color naming task, point to a wide range of ability to cope effectively in the face of inability to process the stimuli adequately. Fisher et al. (1990) recommend caution in interpretation of Stroop results with older populations secondary to possible color confusion. It is hoped that use of the COAST might make results more interpretable with an older population.

Psychometric Properties of the COAST

Earlier work demonstrated high test-retest reliability on the COAST for healthy elderly and controlled Type-II diabetes, with some evidence of a practice effect over a one-month period for the time to completion in the color/word interference task. High test-retest reliability was observed in a larger combined sample of Type-II diabetics and community volunteers, and in demented inpatients in the present study. Since no differences were evidenced between the community volunteers and the diabetics, it seemed reasonable to combine the two groups to address the questions in this paper. A slight practice effect was evident only for time to completion on the color/word interference task in this group, and there was no evidence of a practice effect on any of the COAST measures in the dementia patients. Relatively high correlations between the COAST and Stroop reported in this and in earlier work suggest that the two tests are similar in terms of measurement of time to complete the three tasks of the tests. Correlations between error rates on the two tests were lower and likely reflect the problem of the blue/green confusion on the Stroop. In general, reliability and validity for time to completion indices appear acceptable.

Performance of Selected Patient Populations on the COAST

Age and education of our sample had an impact on all variables of interest over the three tasks of the COAST. This finding of decrement in performance secondary to age is consistent with the findings of other re-

searchers (Comalli et al., 1965; Comalli, Wapner & Werner, 1962; Cohn, Dustman & Bradford, 1984).

No gender differences were noted on any variables of interest when age and education were taken into account. The absence of gender differences on the color/word interference task, regardless of subject age, is a consistent finding in the literature (MacLeod, 1991). However, some researchers (cf., Mekarski, Cutmore & Suboski, 1996) have found gender differences over the color/word interference task. Others (cf., Strickland et al., 1996) have found gender differences only on the color naming or word reading portions of the test. Generally women have been found to take less time to name colors and read color names on the Stroop test (MacLeod, 1991).

The performance of demented subjects differed significantly from the other five diagnostic groups on time to completion for each of the three tasks of the COAST. This is consistent with the work of other researchers who have found that patients with dementia perform more poorly than controls on the Stroop tasks (Koss, Ober, Delis & Friedland, 1984; Fisher et al., 1990; Spieler, Balota & Faust, 1996), as well as other visual attention tasks (Wright, Geffen & Geffen, 1998).

Since the data reported in this study were obtained from a variety of convenience samples, caution is needed in drawing firm conclusions. Yet it does appear that color confusion is minimized in this instrument, and there is preliminary evidence indicating that the test may provide a valid index of sustained attentional and set-shifting capabilities. Furthermore, the fact that there were minimal differences between elderly patient groups experiencing stress due to chronic medical and psychosocial factors and elderly community volunteers self-reported to be in good physical and mental health, while at the same time robust differences were evident between these groups and more cognitively impaired individuals, even after age and education adjustments had been made, suggests that this measure may have potential as an assessment tool in both clinical and research settings.

Further research using the COAST modification of the Stroop test is needed, in terms of assessing performance of different types of dementia patients (e.g., frontotemporal dementia versus AD) and different disease stages (i.e., early-versus late-onset AD). Koss et al. (1984) have suggested that errors on the color/word interference task of the Stroop increase with dementia severity. Comparative data on the COAST with different ethnic populations would also be useful. This new task has been specifically designed with an elderly population in mind, and it is hoped it will provide both clinicians and researchers with a new tool, which may measure attentional and executive functions more accurately in an older population.

REFERENCES

Bando, M., Ishii,Y., & Nakajina, A. (1976). Changes in blue fluorescence intensity and coloration of human lens protein with normal lens aging and nuclear cataract. *Ophthalmic Research, 8*, 456-463.

Benton, A.L., Eslinger, P.J., & Damasio, A.R. (1981). Normative observations on neuropsychological test performances in old age. *Journal of Clinical Neuropsychology, 3*, 33-42.

Botwinick, J. (1973). *Aging and behavior: A comprehensive integration of research findings.* New York: Springer.

Bowles, N.L., & Poon, L.W. (1982). An analysis of the effect of aging on recognition memory. *Journals of Gerontology, 37*, 212-219.

Brennan, M., Welsh, M.C., & Fisher, C.B. (1997). Aging and executive function skills: An examination of a community-dwelling older adult population. *Perceptual & Motor Skills, 84*(3, Pt 2), 1187-1197.

Cohen, J., Cronin-Golomb, A., Growdon, J.H., & Corkin, S. (1988). Color vision deficits in Alzheimer's Disease. *Abstracts for the Society for Neuroscience Annual Meeting, 14*, 219.

Cohn, N., Dustman, R., & Bradford, D. (1984). Age-related decrements in Stroop color test performance. *Journal of Clinical Psychology, 40*, 1244-1250.

Comalli, P.E. Jr. (1965). Cognitive functioning in a group of 80-90 year old men. *Journal of Gerontology, 20*, 14-17.

Comalli, P.E. Jr., Krus, D.N., & Wapner, S. (1965). Cognitive functioning in two groups of aged: One institutionalized, the other living in the community. *Journal of Gerontology, 20*, 9-13.

Comalli, P.E. Jr., Wapner, S. & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood and aging. *Journal of Genetic Psychology, 100*, 47-53.

Filley, C.M. & Cullum, C. M. (1994). Attention and vigilance functions in normal aging. *Applied Neuropsychology, 1*(1-2), 29-32.

Fisher, L.M., Freed, D.M., & Corkin, S. (1990). Stroop color-word test performance in patients with Alzheimer's Disease. *Journal of Clinical and Experimental Neuropsychology, 12*, 745-758.

Golden, J.C. (1978). *Stroop Color and Word Test.* Chicago, IL: Stoelting.

Koss, E., Ober, B.A., Delis, D.C. & Friedland, R.P. (1984). The Stroop Color-Word Test: Indicator of Dementia Severity. *International Journal of Neuroscience, 24*, 53-61.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review.

Mekarski, J.E., Cutmore, T.R.H. & Suboski, W. (1996). Gender differences during processing of the Stroop task. *Perceptual & Motor Skills, 83*, 563-568.

Pachana, N.A., Marcopulos, B., Yoash-Gantz, R. & Thompson, L.W. (1995). California Older Adult Stroop Test (COAST). *The Gerontologist, 35* (Special Issue I), 219.

Pachana, N.A., Marcopulos, B.A., Yoash-Gantz, R. & Thompson, L.W. (1997). California Older Adult Stroop Test (COAST): Pilot test results for subjects with non-insulin-dependent Type II diabetes mellitus and age-matched normal controls. In Habermann, G.M. (Ed) et al. *Looking back and moving forward: 50 years of New*

Zealand psychology. Wellington, New Zealand: New Zealand Psychological Society, Inc. pp. 171-178.

Salthouse, T.A., Toth, J.P., Hancock, H.E., & Woodard, J.L. (1997). Controlled and automatic forms of memory and attention: Process purity and the uniqueness of age-related influences. *Journals of Gerontology Series B-Psychological Sciences & Social Sciences, 52B*, P216-P228.

Salthouse, T.A., & Meinz, E.J. (1995). Aging, inhibition, wording memory, and speed. *Journals of Gerontology Series B-Psychological Sciences & Social Sciences, 50B*, 297-306.

Spieler, D.H., Balota, D.A., & Faust, M.E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 461-479.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests* (2nd ed.). New York: Oxford University Press.

Strickland, T.L., D'Elia, L.F., James, R., & Stein, R. (1997). Stroop color-word performance of African Americans. *Clinical Neuropsychologist, 11*, 87-90.

Stroop, J.R. (1935). Studies in Color/Word interference in serial verbal reaction. *Journal of Experimental Psychology, 18*, 643-662.

Weale, R.A. (1965). On the eye. In A.T. Welford & J.E. Birren (Eds.), *Behavior, aging and the nervous system*. Springfield, IL: Charles C. Thomas.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19*, 87-95.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale. Manual.* New York, NY: Psychological Corporation.

Wilkinson, L., Blank, G., & Gruber, C. (1996). *Desktop data analysis with SYSTAT.* Upper Saddle River, NJ: Prentice Hall.

Wright, M., Geffen, G., & Geffen, L. (1997). Comparative effects of ageing and dementia of the Alzheimer type on orientation of visual attention. *Dementia & Geriatric Cognitive Disorders, 8*, 366-375.