

**Grant Aram Killian, Ph.D.**

*Assistant Professor of Psychology, Nova University, Fort Lauderdale,  
Florida.*

---

## **STROOP COLOR AND WORD TEST**

*Charles Golden. Chicago, Illinois: Stoelting Company.*

### **Introduction**

The origins of the Stroop Color and Word Test date back to 1883 and the beginnings of experimental psychology when Wilhelm Wundt suggested that his student James Cattell investigate, in his doctoral research, the time required to name objects and colors and read the corresponding words. Although some of the first experimental studies with color-word testing were carried out by Cattell (1886) who recognized that color-naming required more time than word-naming, the test as it is known today originated in the laboratory of Erick Rudolf Jaensch (1929).

The Stroop Test itself was introduced into American psychology by John Stroop when he was working at George Peabody College under the direction of Joseph Peterson. Peterson, who was interested in individual differences in the speed of color-naming and word-reading, stimulated Stroop to do his doctoral dissertation with interference in serial verbal reactions. The Color-Word Interference Test was an outgrowth of that doctoral dissertation (Stroop, 1935; 1938). To date the test remains as inconspicuous as it did over a hundred years ago, despite more than a hundred studies showing its reliability, validity and ease of administration since it was originally developed by Stroop in 1935.

The Stroop Test (1935) and all current variations consist of three cards: a *word* (W) card, a *color* (C) card, and a *color-word* (CW) card. The word card is a list of words for colors (e.g., red, green, yellow, and blue) printed in black ink; the color card is a series of color patches or asterisks in the same colors listed on the word card; and the color-word card is a list of those same color words printed in conflicting colors (e.g., the word "red" might be printed in the color "yellow" or "green" and the subject is requested to ignore the word and state the color of the ink.). The test is given in three parts, and the subject is asked to read all three parts as rapidly as possible. The subject's task on the word card is to read aloud the color words, on the color to name the colored patches or asterisks, and on the color-word to name the colors of the ink, ignoring the printed color word.

Currently, there is a standardized version (available from Stoelting Co.), which has been carefully refined by Golden (1978). However, several versions are used within the field that have differed on 1) the colors used, 2) the number of colors used, 3) the type of stimuli used to present the color patches, 4) the reading of the items across rows or down columns, and 5) the method of scoring (for an excellent review, see Jensen & Rohwer, 1966).

The original Stroop Test (1935) consisted of the three basic cards using five colors: red, blue, green, brown, and purple. The words were printed in 14-point

Franklin lower-case black ink on white paper and were arranged in a 10" x 10" matrix of evenly spaced rows and columns. The words were arranged randomly and each of the five colors occurred twice in a column and row, but no word was allowed to follow itself. This version is now out of print.

Thurstone and Mellinger (1953) developed a modification of the original Stroop and also used the three basic cards. However, Thurstone used a black background overlaid by photostatic negatives instead of white paper and only four colors: red, green, yellow, and blue. The colors and color-words were tinted in photographic watercolors, and the color patches were  $\frac{5}{16}$ " circular dots, which were more striking over the black background than the original white background. The words and dots, like Stroop's, were arranged in a 10" x 10" matrix, and each card had a single row of ten practice items to insure that the subject understood the test requirements.

Another modification used by Lazarus et al. (1957), Broverman (1960) and Comalli et al. (1962), according to Jensen and Rohwer (1966), was developed at Clark University. In this modification, the items were in a 10" x 10" matrix; the color card consisted of rectangular patches  $\frac{5}{16}$ " x  $\frac{3}{16}$ "; and only three colors—red, blue and green—were printed on a white background. Like Thurstone's version, at the top of each card was one row of practice items.

Gardner et al. (1959) developed a version similar to Thurstone's except that the words and colors were printed on a white background and the color cards consisted of a set of colored asterisks.

Jensen (1965) developed a set of cards that could be placed on an easel at eye level. The cards were 18" x 25" with colored dots  $\frac{5}{8}$ " in diameter, the letters were  $\frac{5}{16}$ " high, and five colors—red, green, orange, blue, and yellow—were used.

The current standardized version revised by Golden (1975) consists of three cards listing 100 items each presented in a 5" x 20" matrix. Golden (1975) compared the performance of subjects, using three, four, or five colors and found no significant differences between the three versions. Consequently, because all versions were equally reliable the simplest version using only red, green, and blue over a white  $8\frac{1}{2}$ " x 11" background was adopted. Unfortunately, in this version the colors BLUE and GREEN are not clearly distinct and tend to blend into a blue-green color, which may contaminate the interference score and lower the color (C) score. This problem could have been easily avoided if a bolder print or more ink was used by the printer.

Card I (W) consists of 100 words (RED, GREEN, and BLUE) printed in black ink on a white sheet and arranged randomly with no word allowed to follow itself in a column.

Card II (C) consists of 100 colors (written as XXXX) printed in either red, green, or blue on a white sheet with no color allowed to follow itself in a column or match the corresponding word on card I. That is, none of the words on card I match in position the colors of card II.

Card III (CW) consists of 100 colored-words on a white sheet. On this card the order of the words from card I are printed in the order of the colors from card II. For example, word 1 on card I is printed in color 1 on card II to produce the color-word 1 on card III. Using this procedure, no word for a color matches that particular color.

During the test the examiner's participation is minimal and involves only reading the brief instructions and stopping the subject after 45 seconds has elapsed from the start of each card.

Although no study has offered definite norms for children on the Stroop, Golden (1978) has provided preliminary age norms. Essentially any literate child who can recognize the words "red," "green," and "blue" can be administered the test. At all ages above seven years the raw color-word score should be lower than the raw color score that should be lower than the raw word score. To date, the Stroop has been used with children as young as six years (Rand et al., 1963) and with adults up to age 80 (Broverman, 1960). Because the difficulty level is minimal the test is appropriate for ages 6 and over and can be used with all populations except the visually impaired or those unable to recognize the words for the colors. Numerous researchers have investigated Stroop scores with subjects having significantly below-average intelligence, and results of the scores seem to depend on the reading level of the retarded group (Das, 1969, 1970; Leisman, 1971; and Wolitzky et al., 1972).

In most studies females have a slight advantage in color naming (Brown, 1915; Ligon, 1932; Stroop, 1935; Jensen, 1965), however, this efficiency seems to disappear when the interference score (the time required to read the third card) is used (Golden, 1974). Consequently, Golden (1978) concluded that "the differences between the groups are slight, even when significant, and are generally of no importance in clinical or experimental work" (p. 8).

The Stroop Color and Word Test requires few materials and minimal space; any quiet setting with adequate illumination where the subject can be comfortably seated with sufficient room to read on a flat surface is appropriate. The test can be administered in an individual or group format.

For individual administrations the subjects are given all three pages with card W on top, followed by card C, then card CW. The three cards are placed in front of the subjects on a flat surface. They are allowed to rotate the cards only 45° in either direction but may not lift the cards off of the flat surface. They are given 45 seconds to read as many items on the three cards yielding three basic scores. After the subjects are given the three cards, the examiner reads the following instructions to card W (Golden, 1978):

This is a test of how fast you can read the words on this page. After I say begin, you are to read down the columns starting with the first one until you complete it and then continue without stopping down the remaining columns in order. If you finish all the columns before I say "Stop", then return to the first column and begin again. Remember, do not stop reading until I say "Stop" and read out loud as quickly as you can. If you make a mistake, I will say "No" to you. Correct your error and continue without stopping. Are there any questions? Ready? Then begin. (After 45 seconds) Stop. Circle the item you are on. If you finished the entire page and began again, put a one by your circle. Turn the page" (p. 4).

For card C the instructions are identical except that the subjects are told to name the colors on the page. For card CW the subjects are instructed "to name the color of the ink the words are printed in, ignoring the word that is printed in each item" (p. 4) and given examples.

For group testing the same materials are used as with individual testing. Empirical studies are scant comparing the two methods (Jensen, 1966) and in most cases investigators have used different forms for both individual and group administrations. Consequently, comparisons of the two methods from study to study are extremely limited (Kipnis & Glickman, 1959, 1962; Uhlmann, 1962; Podell, 1963). In the current standardized version, Golden (1978) recommends that the group administration and the individual administration be similar, except that the words "out loud" be replaced by "to yourself." As with the individual administration, instructions should be given for each page to insure understanding and should not be utilized with psychiatric or brain-damaged populations.

Problems seldom occur with the administration of the Stroop Color and Word Test. However, sometimes a subject will cover up all but one letter on card CW in order to read the color easily. This strategy, as well as rotations of the cards in excesses of 45° should be stopped immediately.

The testing procedure is simple and straightforward; a trained psychologist is not necessary to administer the separate phases of the test and either a secretary, aide, teacher, or mental health counselor can be quickly trained to administer the test. In general, because subjects are given only 45 seconds per card the entire administration rarely requires more than 5 minutes to complete, even when administered to psychiatric populations.

#### **Practical Applications/Uses**

Research on the Stroop Color and Word Test has established that the test assesses psychological processes and functions that affect cognition in normal, neuropsychological, and psychiatric populations. Stroop scores have been associated with cognitive flexibility, attention deployment, resistance to interference from outside stimuli, creativity, defense structures, and cognitive style and complexity. Moreover, the Stroop is a reliable, efficient and effective clinical test for evaluating psychopathology and brain dysfunction and can be utilized as a screening test or as part of a general test battery for making a differential diagnosis. Its short administration time, reliability, validity, and ease of administration make the Stroop a valuable test which can be successfully used in numerous settings. For example, a junior high school counselor could use the Stroop to assess efficiency of attention deployment and cognitive flexibility, a psychopathologist could use the Stroop to assess responsiveness to psychotropic medication (Killian et al., 1984), and a private practitioner could use the test to evaluate brain dysfunction.

The Stroop has not yet become part of the psychological test battery, but could easily serve as an introductory test in the battery by providing a minimally threatening and maximally absorbing beginning to psychological assessment. Serving as an easy bridge to the psychological examination with minimal contact with the examiner, the subject is not intimidated by specific questions concerning intelligence or threatening inkblots, which may disturb unconscious conflicts.

As a screening instrument for the detection of brain dysfunction, the Stroop has numerous advantages: it requires only elementary education, can be translated into foreign languages without difficulty, and requires only 5 minutes to administer. Either simple cutoff points or patterns of Stroop scores can be identified and

be diagnostically useful (e.g., normal W, low C, and CW, or all low scores) in brain dysfunction (Perret; 1974, Golden, 1978).

In terms of psychopathology, Wapner and Krus (1960) found that schizophrenic patients had a significant loss of speed on all cards of the Stroop. Along similar lines, Weiss and Sherman (1962) found a significant correlation between the Manifest Anxiety Scale and poor performance on the Stroop among chronic schizophrenics but not among acute schizophrenics. Peixotto and Rowe (1969) compared schizophrenics with normals and found significant differences between the groups on all three cards. Golden (1976) and Killian et al. (1984) also found significant differences between normals and psychiatric patients on all Stroop measures.

Lichtenstein (1961) was the first to find that on the Minnesota Multiphasic Personality Inventory (MMPI) constricted subjects were more conventional and defensive. In an attempt to determine the specific aspects of personality and psychopathology that were affected by the Stroop, Golden and Golden (1975) administered the Stroop to three groups of normal subjects, categorized by high scores obtained on particular MMPI scales: 1) high scores on D and Sc; 2) high scores on Hs, Hy, Pt and Ma; and 3) high scores on Pd, Mf, Pa and Si. A one-way analysis of variance revealed significant differences among the three groups: group 1 scored worst, group 3 showed the best performance, and group 2 fell in between.

Bush (1975), also using the MMPI in normal males and females, examined the relationship between psychoticism (i.e., scales F, Sc, and Pz) in males and defensive rigidity (i.e., scales L, R, and Ec-5) in females. The Stroop significantly correlated with the psychoticism index for females ( $p = .01$ ) and defensive rigidity for males ( $p = .01$ ). Killian (1981) also found that of 19 cognitive measures only the Stroop inference score (CW) successfully discriminated between those schizophrenic and depressed patients who responded to psychotropic drugs from those who were classified as nonresponders.

Traditionally, there have been many scoring formulas since Stroop's (1935) original work. Jensen and Rohwer (1966) reviewed 16 scores derived from the three basic time scores on cards W, C and CW, and presented a comprehensive analysis of scores into three classes: 1) basic time scores and derived scores, 2) error scores, and 3) serial scoring. The complexities of these formulas are avoided in the revised Stroop manual which clearly and simply presents four scores based on the three cards. The Word Score (W), the Color Score (C), the Color-Word score (CW), and a predicted CW score are all based on the items completed on each card. Errors are not counted because the subject is made to repeat each incorrect response, lowering the overall score. In order to reduce the redundancies of the derived scores and assess which formulas yielded additional unique information, Golden (1978) performed a factor analysis on eight major scores which yielded three basic factors: speech, color difficulty, and interference. The presence and item loadings on these three factors were consistent with Jensen and Rohwer (1966). The one exception occurred on the speed factor in which Jensen (1965) found that the W score correlated .97 while Golden (1978) found that only  $W \div C$  correlated .98 with the speed factor and W correlated  $-.67$ .

The appendix in the test manual provides the user with age corrected scores,

and T-scores for W, C, CW and the predicted Interference score. Because these scores are derived from the number of items completed on each card, are easily calculated, and take only a few minutes to learn, computer scoring and transparencies are unnecessary. Interpretation of these four scores is objective and not based on subjective clinical judgment. The test manual, however, is scant concerning rules or criteria for interpreting either raw- or T-score patterns and profiles. The actual meaning of certain configurations is thus left to the experience of the users and their familiarity with previous Stroop research.

#### Technical Aspects

Jensen (1965) found that test-retest reliabilities of basic and derived scores with intervals of three minutes, one day, and one week showed no appreciable differences. Moreover, reliability of the Stroop scores is highly consistent across different versions of the test (Golden, 1978). However, Jensen (1965) found that derived scores that utilize differences and ratios have somewhat lower reliabilities. For the three basic raw scores W, C, and CW Jensen (1965) reported reliabilities of .88, .79, and .71 respectively (N = 436). For the same scores Golden (1978) reported reliabilities of .89, .84, and .73 respectively (N = 450) for group administrations; .86, .82, and .73 respectively (N = 30) for individual administrations; and .85, .81, and .69 respectively (N = 60) for subjects administered both the individual and group forms.

Hollingworth (1915) was the first to assess the effects of practice on color-naming and found that during 100 administrations over a period of 10 to 40 days 19 subjects had a 30% improvement in speed of color-naming (C). Stroop (1935) found that with practice effects of 8 trials subjects had the most improvement on card CW and the least improvement on card W. Jensen (1965) found similar practice effects that were in agreement with both Hollingworth (1915) and Stroop (1935) and found when subjects (N = 50) were tested every day for 10 days they had a 36% improvement in speed of color-word-naming (CW), a 23% improvement in speed of color-naming (C), and a 15% improvement in the speed of word-naming (W) from the first trial (day 1) to the last trial (day 10). Most of the practice effects (Jensen, 1965) occur within the first four trials for CW (13%, 10%, 5%, and 3% improvement), within the first three trials for C (7%, 7% and 3% improvement), and within the first three trials for W (3%, 4% and 1% improvement). Smith and Nyman (1959) also found that the performance on the three basic scores became asymptotic after five trials. Despite improvement with practice, individual differences do not seem to interact with practice effects, thus, subjects tend to maintain the same rank order during stages of improvement (Gates, 1922).

#### Critique

The Stroop is a highly reliable and valid objective test which seems to have relevance for the practicing clinician and the experimental psychopathologist. Unlike some tests, used in the psychological test battery, that may lack reliability and fail to differentiate processes and functions, (e.g., House-Tree-Person, Killian, in press), the Stroop provides reliable and significant information about

various disorders (Golden, 1976; Killian, 1984). The Stroop, which is as inconspicuous today as it was during its development, is of considerable importance and interest for the following reasons: 1) it provides reliable and stable measures on apparently three simple basic aspects of cognitive-perceptual functioning; 2) despite individual differences, rank orders are maintained even with significant practice effects (Jensen, 1966); and 3) it has been significantly correlated with a variety of instruments, populations, traits, and disorders. What is most unfortunate is that this simple reliable test with many uses has not been included in the standard psychological test battery. To quote Golden (1978), "We strongly believe that we have only touched on a few of the possible uses of the Stroop in our investigations to date, and that future research will identify more important areas (p. i)."

### References

- Broverman, D. M. (1960). Dimensions of cognitive style. *Journal of Personality, 28*, 167-185.
- Brown, W. (1915). Practice in associating color-names with colors. *Psychological Review, 22*, 45-55.
- Bush, M. (1975). Relationship between color-word test interference and MMPI indices of psychoticism and defense rigidity in normal males and females. *Journal of Consulting and Clinical Psychology, 43*, (6), 926.
- Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*, 63-65.
- Comalli, P. E., Wapner, S., & Werner, H. (1962). Interference effects of the Stroop Color Word Test in childhood, adult and aging. *Journal of Genetic Psychology, 100*, 47-53.
- Das, J. P. (1969). Development of verbal abilities in retarded and normal children as measured by the Stroop test. *British Journal of Social and Clinical Psychology, 8*, 59-66.
- Das, J. P. (1970). Changes in Stroop test response as a function of mental age. *British Journal of Social and Clinical Psychology, 9*, 68-73.
- Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. B., & Spence, D. P. (1959). Cognitive Control: A study of individual consistencies in cognitive behavior. *Psychological Issues, 1*, 1-185.
- Gates, G. S. (1922). Individual differences as affected by practice. *Archives of Psychology, 8*, (58), 1-74.
- Golden, C. J. (1974). Sex differences in performance on the Stroop color and word test. *Perceptual and Motor Skills, 39*, 1067-1070.
- Golden, C. J. (1975). A group form of the Stroop color and word test. *Journal of Personality Assessment, 39*, 386-388.
- Golden, C. J. (1976). The diagnosis of brain damage by the Stroop test. *Journal of Clinical Psychology, 32*, 654-658.
- Golden, C. J. (1978). *Stroop Color and Word Test: A Manual for Clinical and Experimental Uses*, Chicago: Stoelting Co.
- Golden, C. J., Golden, E. E. (1975). Resistance to cognitive interference as a function of MMPI profile. *Journal of Consulting and Clinical Psychology, 43*, 749.
- Hollingworth, H. L. (1915). Articulation and association. *Journal of Educational Psychology, 6*, 99-105.
- Jaensch, E. R. (1929). Grundformen menschlichen Seins. Mit Berücksichtigung ihrer Beziehungen zu Biologie und Medizin, zu Kulturphilosophie und Pädagogik. Berlin: Otto Elsner.
- Jensen, A. R. (1965). Scoring the Stroop test. *Acta Psychologica, 24*, 398-408.
- Jensen, A. R., & Rohwer, W. D. (1966). The Stroop color word test: A review. *Acta Psychologica, 25*, 36-93.

- Killian, G. A. (1981). *The Effects of psychotropic medication on cognitive control measures*. Unpublished doctoral dissertation, University of Chicago.
- Killian, G. A., Holzman, P. S., Davis, J. M., & Gibbons, R. (1984). Effects of psychotropic medication on selected cognitive and perceptual measures. *Journal of Abnormal Psychology*, 93, (1), 58-70.
- Killian, G. A. (in press). The House-Tree-Person (H-T-P) Technique. In D. Keyser & R. Sweetland, (Eds.), *Test Critiques* (Vol. 1, pp. 338-352). Kansas City, MO: Test Corporation.
- Kipnis, D., & Glickman, A. S. (1959). Validity of non-cognitive tests at nuclear power school. *U.S.N. Bur. Naval Personnel Technical Bulletin*, No. 59-6.
- Kipnis, D. & Glickman, A. S. (1962). The prediction of job performance. *Journal of Applied Psychology*, 46, 50-56.
- Lazarus, R. S., Baker, R. W., Broverman, D. M., & Mayer, J. (1957). Personality and psychological stress. *Journal of Personality*, 25, 559-577.
- Leisman, G. (1971). Cognitive interference in spastic-hemiplegic children on the Stroop color word test. *British Journal of Social and Clinical Psychology*, 10, 379-382.
- Lichtenstein, E. (1961). The relation of three cognitive controls to some perceptual and personality variables. *Dissertation Abstracts*, 22, 2467.
- Ligon, E. M. (1932). A genetic study of color naming and word reading. *American Journal of Psychology*, 44, 103-121.
- Peixotto, H., & Rowe, A. (1969). Effects of cognitive interference on performance in relation to psychopathology. *Perceptual and Motor Skills*, 29, 523-527.
- Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behavior. *Neuropsychologia*, 12, (3), 323-330.
- Podell, H. A. (1963). Note on successive dimensional analysis applied to affective, cognitive, and personality traits. *Psychological Report*, 13, 813-814.
- Rand, G., Wapner, W., Werner, H., & McFarland, H. (1963). Age differences in performance on the Stroop color word test. *Journal of Personality*, 31, 534-558.
- Smith, G. J. W., & Nyman, G. E. (1959). Psychopathologic behavior in a serial experiment. *Lunds Universitets Arsskrift*, N.F. Avd. 2, 56, 5. Lund: Gleerup.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643-662.
- Stroop, J. R. (1938). Factors affecting speed in serial verbal reactions. *Psychological Monograph*, 50, (5), 38-48.
- Thurstone, L. L., & Mellinger, J. J. (1953). The Stroop test. *The Psychometric Laboratory*, University of North Carolina, No. 3.
- Uhlmann, F. W. (1962). Test of color recognition (Form DE x-27-61). Detroit: The Detroit Edison Co.
- Wapner, S., & Krus, D. M. (1960). Effects of lysergic acid diethylamide and differences between normals and schizophrenics on the Stroop color word test. *Journal of Neuro-psychiatry*, 2, 76-81.
- Weiss, R., & Sherman, M. (1962). Anxiety and interfering responses in college students and psychiatric patients. *Newsletter of Research Psychology*, 4, 35-40.
- Wolitzky, D. L., Hofer, R., & Shapiro, R. (1972). Cognitive controls and mental retardation. *Journal of Abnormal Psychology*, 79, 296-302.