

The development of automatic word recognition and reading skill

MARGARET SCHADLER and DAVID M. THISSEN
University of Kansas, Lawrence, Kansas 66045

The development of automatic word recognition as a function of reading skill was investigated in three experiments using the Stroop task. Reading skill level ranged from nonreaders to readers above the sixth-grade equivalent. Interference with color naming begins to emerge early in the process of learning to read, increases, and then subsequently decreases. Strings of identical letters delayed color naming for children just beginning to learn to read. The interference from words, presumably reflecting semantic processing, began developing early but did not peak until the second- to fourth-grade reading levels. These different sequences of development of interference in the various stimulus conditions suggest that word recognition is the result of a number of component processes that develop as children acquire skill in reading.

Most literate adults cannot or do not ignore words even when it is to their advantage to do so. In the Stroop task, a color name, when printed in ink of a different color, interferes with naming the color of the ink (Stroop, 1935; see also Dyer, 1973; Jensen & Rohwer, 1966, for reviews). Similarly, in the picture naming version (Rosinski, Golinkoff, & Kukish, 1975), the name of another exemplar from the same category printed on a pictured object interferes with naming the picture. These extremely reliable interference effects are usually interpreted as evidence that the reader automatically reads the word, despite the fact that doing so delays and increases the difficulty of executing the naming response (e.g., Posner & Snyder, 1975). These effects are considered evidence of automatic word recognition, and Stroop-like¹ tasks are frequently used to study the development of this skill (e.g., Ehri, 1976; Guttentag & Haith, 1978, 1979; Rosinski, 1977; Rosinski et al., 1975).

LaBerge and Samuels (1974), whose seminal paper directed attention to the concept of automatic processing in reading, considered automaticity to be characteristic of fluent reading and proposed that it developed slowly with extensive practice. However, young children who have had little formal reading instruction evidence interference on Stroop-like tasks. Interference emerges

as early as the first (Ehri & Wilce, 1979; Guttentag & Haith, 1978; Posnansky & Rayner, 1977; Stanovich, Cunningham, & West, Note 1) or second grade (e.g., Ehri, 1976; Rosinski et al., 1975; Schiller, 1966) on both the picture and color naming versions. The word recognition process would seem to become automated much earlier than would be expected from LaBerge and Samuels' model (e.g., Ehri, 1976; Guttentag & Haith, 1979; Rosinski et al., 1975).

Word recognition, while often treated as a unitary process, may be the result of a number of component processes. The amount of delay in the naming response by adults differs with both the graphophonemic and semantic properties of the interfering stimulus. Words from the same category as the picture naming response interfere more than words from other categories (e.g., Guttentag & Haith, 1978; Lupker, Cartman, & Symansky, Note 2). Similarly, words interfere more than nonwords (Lupker et al., Note 2), and pronounceable nonwords interfere more than unpronounceable nonwords (Guttentag & Haith, 1978). Comparable effects have been reported on the Stroop task (see Dyer, 1973). There is some evidence that these graphophonemic and semantic interference effects have different developmental sequences. Both Guttentag and Haith (1978) and Stanovich et al. (Note 1) concluded that interference from letters develops prior to interference from words, although both emerge in the first grade. Guttentag and Haith found that pronounceable and unpronounceable letter strings caused similar amounts of interference with picture naming for first-graders and below-average third-grade readers, whereas pronounceable nonwords caused more interference for above-average third-grade readers and adults. There is also some suggestion of different developmental sequences for graphemic and semantic components of word recognition in Schiller's (1966) results. Posnansky and Rayner (1977) did not find reliable Age by Stimulus interactions when they

This research was supported by National Institute of Education Grants G-77-0010 and G-78-0179 to the first author and by the University of Kansas Graduate Research Fund. We thank Sue Kemper for comments on the manuscript, Laura Baker, Brian Drummond, Donna Fisher, Chris Jordan, Bud Moyer, and Wanda Wilkinson for their help, and the Lawrence schools, children, and parents for their participation. The opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred. Requests for reprints should be addressed to Margaret Schadler, Psychology Department, Fraser Hall, University of Kansas, Lawrence, Kansas 66045.

tested a variety of graphophonemic features of words among a wide age range of children and college students in three experiments. Age may not be a sufficiently precise measure of reading to be sensitive to relatively subtle skill-related changes in interference caused by the various graphic and phonemic features of letter sequences Posnansky and Rayner used. Investigations concerned with developmental changes or differences in process arising from relatively modest stimulus differences should utilize the most direct subject variable available, reading skill in this case.

An important characteristic of the developmental sequence of performance on Stroop-like tasks is that interference appears to decrease with additional reading skill development. In systematic cross-sectional studies using the Stroop task, the greatest interference is for the primary-grade readers, with lesser amounts for older children and adults (Comalli, Wapner, & Werner, 1962; Rand, Wapner, Werner, & McFarland, 1959; Schiller, 1966). Using the picture-word task, Ehri (1976) reported an overall age-related decrease in interference when comparing skilled second-grade readers, fifth-graders, and adults. Less interference for adults than for third-graders is also evident, although not statistically tested, in Guttentag and Haith's (1978) data. Guttentag and Haith (1979) did report decreases in interference between second- and third-graders, as well as between third-grade and college students. Further, both increases and decreases in interference have been observed within groups over relatively short periods of time. In their year-long study, Stanovich et al. (Note 1) found that interference first increased and then decreased during the first grade. Ehri and Wilce (1979) reported that one or two practice sessions of naming the interfering words decreased the amount of interference for children who recognized the words at the outset of the experiment but increased interference for those who learned to recognize the words during the training sessions. Evans (Note 3) found that, for adults, interference from words on the Stroop task gradually declined and then ceased with extensive practice (days).

Stanovich et al. (Note 1), assuming response competition to be the source of interference (cf. Posner & Snyder, 1975), suggested that interference decreases because the speed with which the interfering words are processed continues to increase after the child has learned to automatically process the words (see also Ehri & Wilce, 1979). Stanovich et al. argue that as words are processed faster, there is less overlap between the reading and naming responses and, thus, less competition (see also Ehri & Wilce, 1979). Response competition explanations have been challenged a number of times (e.g., Guttentag & Haith, 1979; Seymour, 1977; Treisman & Fearnley, 1969). Further, a simple quantitative increase in speed of processing does not seem adequate to explain either Evans' (Note 3) results or the range of factors that contribute to decreasing inter-

ference. Subjects may modify the way they process the information rather than the speed with which they do so.

The research to date suggests that Stroop-like tasks may be useful in identifying changes in the way in which words are recognized as a result of learning or development. The patterns emerging in the research are (1) rapid increase of interference, followed by a decrease, and (2) different developmental functions for the component processes. These results, however, must be qualified by the fact that age or grade has been used as the developmental variable when reading skill is actually the crucial factor. Since a few investigators have shown differences in amount of interference between skilled and less skilled readers within the same grade (Ehri, 1976; Ehri & Wilce, 1979; Guttentag & Haith, 1978; Pace & Golinkoff, 1976), research using reading skill as the developmental variable is needed. If we are to argue that the developmental sequence of interference effects is curvilinear, nonreaders, who should show no interference effects, should be tested. Also, a group of older children should be tested to provide a reference for performance at the upper end of the reading scale. The first experiment was designed to investigate the development of some of the components of word recognition as a function of reading comprehension level. The second and third experiments replicated and further explored an unexpected source of interference in the first experiment.

EXPERIMENT 1

Method

Subjects. The subjects were 140 children, grouped according to their comprehension scores on the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968). All children were Lawrence, Kansas, residents randomly selected from the state population lists. At the beginning of the school year, 10 boys and 10 girls from the kindergarten, first, second, and third grades were tested. Midway through the school year, 6 more male and 9 female kindergartners plus 14 male and 11 female first-graders were tested. A group of 10 males and 10 females were tested during the summer just prior to their entering the sixth grade. All subjects had normal or corrected-to-normal sight and normal color vision, and all knew the names of the colors used. One male who showed some evidence of color blindness was excluded from the study. Each child was paid \$2 to come to the campus for testing.

Equipment and materials. The apparatus for the Stroop test was a 35-mm Kodak Carousel projector with a Hunter 120 Klockounter attached. The timer was started by a phototransistor when a stimulus slide was projected, and it was stopped by a voice-operated key when the child named the projected color. The stimulus slides were projected onto a 27.5-cm-square rear-view projection screen placed on a table 90 cm in front of the subject. A Sony two-track audio cassette recorder, used to record all trials, provided an auxiliary timing system. One track was used to record the child's response and the second to record a "beep" as each stimulus slide was projected. If the child prematurely stopped the timer with a nonlinguistic utterance (e.g., "ah"), the trial was hand-timed from the tape with a stopwatch.

The stimulus slides were prepared using Kodalith high-contrast 35-mm film to photograph 20 x 25 cm white cards on which

the stimulus letters were printed in black capital letters. The resulting negatives were inserted between two sheets of red, blue, or green theatrical gelatin. The finished slides had capital letters in red, blue, or green print on a black background. When the slides were projected onto the screen, each letter was approximately 1 x 2 cm.

Stimulus materials were (1) the words RED, BLUE, and GREEN in the congruent color (e.g., the word RED in red print), (2) the same words in incongruent colors (e.g., the word RED in blue or green print), (3) the unpronounceable nonwords UBD, LGNE, and NDRBE, generated from the letters in the color names, in each color, and (4) strings of three, four, and five Xs in each color. Opaque slides were alternated with the test stimuli to block the light flow between trials.

The test items were grouped into six blocks of 12 slides. Each block consisted of three slides from each of the four stimulus types. The slides were randomly ordered within blocks, with the restrictions that no color appear more than three times in succession and no two slides of the same stimulus type appear in succession. All forms of the congruent and the incongruent stimulus slides were presented within two consecutive blocks. Eight additional slides, two examples of each of the four stimulus types, were used as training stimuli.

The first part of the reading test consisted of a list of words, DOG, CAT, BOY, SHOE, HAT, GIRL, HORSE, RED, BLUE, GREEN, YELLOW, BLACK, BROWN, and COLOR, typed in Pica uppercase letters on a sheet of 21.5 x 17 cm white paper. The second part was the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968). It consists of 10 oral reading paragraphs, one per grade level for Grades 1-10. Each paragraph is followed by five comprehension questions. The comprehension score is derived from the number of correctly answered questions adjusted for the difficulty level of the last paragraph read (Gilmore & Gilmore, 1968).

Plates 24, 25, 28, and 29 of Ishihara's (1939) color-blindness test were used to test color vision.

Procedure. A woman tested each child individually in a psychology laboratory room on the university campus. The child sat at a table facing the rear-view projection screen.

The test for color-vision was given first. The child was asked to identify the number embedded in Ishihara (1939) plates 24 and 25. If the child did not know the number, he/she traced it with his/her finger. If the child could not find the number, Plates 28 and 29 were used as a second check. When evidence of color blindness was found, the interference test was not given.

For the Stroop task, each child was asked to identify the color projected on each trial as rapidly as possible and to ignore the letters. A verbal ready signal was given prior to each slide. Verbal feedback and encouragement were given for correct responses. The child was asked to correct any error made before the display was terminated. Between trials, the experimenter recorded the response time, measured from slide outset to the verbal response, and any incorrect response.

The reading test was given after the Stroop task. Errors and total reading time were recorded for the list of words. The child was then given the Gilmore Oral Reading Test, using the standard directions (Gilmore & Gilmore, 1968).

Design. The design was a 6 (comprehension level) by 4 (stimulus condition) by 3 (response color) by 2 (sex of subject) mixed factorial. The six comprehension levels were: (1) nonreaders, who showed no ability to read any of the 14 list words or the Gilmore Oral Reading Test (mean age = 5.6 years), (2) beginning readers, who were able to read 1 or more of the 14 list words but obtained a comprehension score of less than 10 on the Gilmore Oral Reading Test (mean age = 6.1 years), (3) first-grade level readers, who obtained comprehension scores of 10-14 on the Gilmore Oral Reading Test, a grade equivalency range of 1.3-1.9 (mean age = 6.8 years), (4) second-grade level readers, who obtained comprehension scores of 15-20, a grade equivalency of 2.1-3.4 (mean age = 7.2 years), (5) fourth-grade level readers, who obtained comprehension scores of 21-27, a grade equivalency of 3.8-5.8 (mean age = 7.6 years), and (6) the pre-sixth-grade children, all of whom attained reading comprehension scores above 27 (grade equivalency 6.2 or higher; mean age = 11.2 years). A group of 20 first-, second-, and third-graders who obtained reading comprehension scores of 27 or higher (mean age = 7.8 years) were excluded from these statistical analyses.

The 4 (stimulus condition) by 3 (response color) within-subjects factorial completed the mixed design. Stimulus conditions were (1) incongruent color words, (2) congruent color words, (3) colored Xs, and (4) colored nonwords; the response colors were red, blue, and green. Each person made six responses to each of the 12 stimulus-by-color conditions; the median of these six response times was the dependent variable.

Results and Discussion

The distribution of errors according to reading skill and interference condition is shown in Table 1. The pattern is obvious; errors in the incongruent condition tended to increase with reading skill through the second-grade level, whereas there was a slight tendency for them to decrease with reading skill in the other conditions. Since there were few errors in any condition, no statistical analyses were conducted.

Response times to the stimulus conditions and reading comprehension levels were analyzed using orthogonal polynomial trend analysis (Bock, 1975). For this analysis, the reading comprehension groups were taken to be equally spaced, as no specific interpretation of the polynomial effects was expected in any event. For the planned orthogonal comparisons, the response

Table 1
Means and Standard Deviations for Errors on the Stroop Task in Experiment 1

Reading Level	Number of Subjects	Stimulus Condition							
		Incongruous		Nonwords		Xs		Congruous	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Nonreaders	18	.56	.86	.56	.92	.72	.83	.61	.85
Beginning	19	.95	1.03	.58	.77	.42	.61	.32	.48
First Grade	18	1.89	1.71	.67	.69	.22	.55	.39	.85
Second Grade	23	3.04	2.70	.17	.39	.30	.56	.39	.58
Fourth Grade	22	2.91	2.65	.27	.63	.14	.35	.32	.48
Pre-Sixth Grade	20	2.25	1.65	.35	.59	.00	.00	.15	.37
Sixth Grade	20	2.70	1.66	.10	.31	.05	.22	.45	.60

time for the incongruent color-word condition was set against the mean of the other three conditions, the mean for the congruent color-word condition was set against the mean of the control Xs and nonsense conditions, and the X condition was contrasted with the nonword condition. Preliminary analysis showed no main effects of sex and no significant interactions of sex with any other variable or variables. Therefore, sex was dropped from the analysis.

The trend analysis indicated that the response times to the four types of interfering stimuli (incongruent words, congruent words, nonwords, and Xs) followed different curves as reading comprehension level increased. There were no significant higher order (quartic and quintic) trends; therefore, the fitted cubic functions are used to represent the data, as shown in Figure 1.

The response time functions for both the nonword and the Xs conditions were essentially concave downward. They differed primarily in that the response time curve for the Xs condition accelerated more quickly to an earlier peak than did the curve for the nonword condition; the nonword condition curve peaked later and then decelerated more rapidly. These differences are reflected in the significant difference between the cubic components of the trends for response times in the Xs and nonword conditions [$F(1,116) = 7.45$, $p = .007$]. The initial upward trend of these curves is evidence that both the Xs and the nonwords interfered with the color naming response for children learning to read. Interference from nonword strings has been found previously (e.g., Guttentag & Haith, 1978) and was expected. Strings of Xs, however, are frequently used as the control condition and are assumed not to interfere with color naming. This result is further explored in the second and third experiments.

In contrast to the convex trends across increasing reading comprehension for the Xs and nonword conditions, mean response times in the congruent condition exhibited an essentially linear decrease across the entire range. The difference between the quadratic com-

ponents, reflecting curvature, for the mean response times to the congruent stimuli vs. those for Xs and nonwords was highly significant [$F(1,116) = 27.26$, $p < .0001$].

There are two factors potentially contributing to the reading skill-related decrement in the congruent condition. One is the typical age-related improvement in response times, which is usually held to be a nonspecific effect of maturation and experience. The second is a facilitative effect of the congruent word on naming the color. Since the curve for response times in the Xs condition reflects the development of interference rather than neutrality, the relative contributions of these two factors cannot be determined for these data (but see Experiment 2).

Finally, the trend for the incongruent color words was concave downward, as were the curves for the Xs and nonwords, with peak response times for the Stroop stimuli occurring between the second- and fourth-grade reading equivalent levels. The incongruent color-word trend had a different orientation from the other curves: The mean response times for the incongruent stimuli increased as reading comprehension increased from nonreaders to about the fourth-grade equivalent and then it decreased, but the pre-sixth-graders' mean response time was no faster than that of the nonreaders. In contrast, the response times for the Xs, nonword, and congruent word conditions all subsequently decreased after the initial increase, such that these response times for the pre-sixth-grade group were substantially faster than those for the nonreaders. This difference between the essentially zero linear component for the Stroop condition and the negative slope components for the other three curves was significant [$F(1,116) = 52.92$, $p < .001$]. The high level of interference in this condition has been found previously on Stroop-like tasks (e.g., Guttentag & Haith, 1978; Rosinski, 1977; Schiller, 1966). Both semantic and the graphophonemic qualities of the incongruent words probably contribute to the magnitude of the interference (see also Dyer, 1973).

The mean response times for the most advanced group of young readers are represented in Figure 1 by the set of underlined symbols. Inspection of these data shows two points. First, the response times were slow. This group responded more slowly in every stimulus condition than did the fourth-grade equivalent readers, but the differential was greatest in the least interfering conditions. The problem, of course, is that we cannot determine here whether we obtained an unrepresentative sample or whether slow responding is associated with early attainment of a high level of reading skill. The data for this group were consequently excluded from the trend analysis and subsequent curve fitting because the slow response times would have distorted the shape of the interference curves and would have done so differentially across conditions and skill levels. With the developmental course established, however, it is apparent that the pattern of interference for the sixth-grade

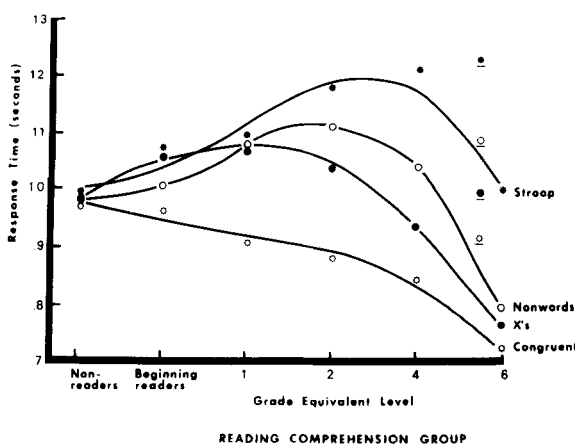


Figure 1. Mean response times in each stimulus condition as a function of reading skill in Experiment 1.

equivalent readers was similar to that of the fourth-grade equivalent group. Difference in response times between conditions were quite similar for both groups. We will return to this point in discussing the results of the second experiment.

There was a significant interaction between response color and stimulus condition [$F(6,111)=2.84, p<.01$]. Responses to congruent color-word pairs, Xs, and nonwords were all slowest for green letters and fastest for red; however, responses to the green color were intermediate to the red and blue colors for the congruent pairs.

Dyer (1973) reported a main effect of color; however, color effects are rarely mentioned and may not be tested in most experiments. The most likely source of color effects is differences in brightness of the three gels. Brighter colors are perceived slightly faster and may also have triggered the phototransistor used to start the timer somewhat sooner than did the dimmer colors. If part of the interference effect is due to response competition (e.g., Posner & Snyder, 1975), then relative speeds of naming each color could interact with stimulus condition. However, the effects were comparatively small and did not differ reliably as a function of reading comprehension level.

In summary, interference with color naming in all conditions but the congruent words began to emerge during the early phases of reading skill acquisition. Each stimulus condition, however, showed a different sequence of development. Understanding this relatively complex Condition by Skill interaction depends, in part, upon understanding the nature of the interference in the Xs condition as it relates to reading skill. The reliability of this unexpected source of interference is explored in the next two studies, and then the development of Stroop interference and word recognition will be discussed.

EXPERIMENT 2

Strings of Xs are frequently used as the neutral or control condition, particularly when the discrete-trial procedure is employed (e.g., Dalrymple-Alford, 1972). Since we found that interference in this condition peaked at the first-grade reading level, which is usually the youngest group given Stroop-like tests, there would be little reason to suspect that the Xs delayed color naming. Age- and grade-related decreases in response times in this condition could be attributed to improvement in reaction time with age. The only evidence in the literature that letters might cause interference in the Stroop task is Stanovich et al.'s (Note 1) recent finding that single letters interfere with color naming for first-graders, and this effect was relative to strings of Xs, their control condition. The Xs have been regarded simply as placeholders, a means of presenting approximately the same amount of color in a neutral context as for a potentially interfering stimulus. A major purpose of Experiment 2 was to determine if strings of Xs would

interfere with color naming relative to rectangular patches, the other common control condition in Stroop tasks (e.g., Comalli et al., 1962; Rand et al., 1959; Schiller, 1966).

One consequence of finding interference in the Xs condition is that we were unable to evaluate development of possible facilitation effects in the congruent color-word condition, particularly among the older readers, whose response times would be expected to be faster than those of the young children. The second purpose of this experiment was to investigate the development of facilitation from the presence of the congruent word. Interest in facilitation in the Stroop task is relatively recent. Dalrymple-Alford and Budayr (1966) first presented congruent words when they introduced the discrete-trial procedure. Congruent words have not been used in the Stroop task, to our knowledge, with children. The comparison condition for the congruent words has been strings of Xs; thus, the effect, which has been interpreted as facilitation from color-word congruence, may actually be due to interference from the Xs in the control condition.

In the present study, we included rectangular patches as a control or noninterfering condition. While all the stimulus conditions used in the first study were repeated in this one, we were primarily concerned with comparing color naming latencies in the congruent and Xs conditions with those in the rectangles condition. Comparable color naming response times in the rectangle and congruent conditions, in conjunction with reliably slower responses in the Xs condition, would be evidence that the congruent word does not facilitate color naming but that Xs do interfere. Alternatively, skilled readers may not show interference in the Xs condition if the Xs are processed as a vehicle for the color dimension rather than as letters or meaningful symbols. If so, there should be no difference in response times for the rectangles and Xs, but colors of congruent words should be named faster than either. The third possibility is that response times for naming the colors of the rectangles are intermediate to those for the Xs and congruent conditions, providing evidence for both facilitation and interference.

Method

Subjects. The subjects were 71 children classified as reading at the second-, fourth-, or sixth-grade equivalent level. All were Lawrence, Kansas, elementary school children, randomly selected from the state population list. Testing was done during the summer prior to the children's entering the second, third, or fourth grade. All participants had normal or corrected-to-normal sight and color vision. Data from four additional children who did not meet the minimum criterion of a raw score of 14 on the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968) were excluded from the analysis. Each child was paid \$2 to come to the campus for testing.

Equipment and materials. The equipment used in the first experiment was also used in this one. The slides were prepared in the same manner as for the first experiment, with the exception that lowercase letters were used. In addition, a set of red, green, and blue rectangles was prepared. When projected, the rectangles

were half the height and the same length as the four-letter strings. The same stimulus orders used in Experiment 1 were used here, except that three rectangles, one of each color, were inserted into each block of trials. Insertion was random, with the restriction that no color appear more than three times in succession. The test then consisted of 9 practice trials and 6 blocks of 15 test trials each.

Procedure. Scheduling of subjects was arranged by a female college student; all testing was conducted by a male. The actual test procedure was identical to that used in the first experiment. Following testing, the data were categorized on the basis of the Gilmore and Gilmore (1968) comprehension test scores into the three highest reading equivalent groups used in the first experiment. There were 25 children (mean age = 8.3 years) in the second-grade equivalent, 28 children (mean age = 8.4 years) in the fourth-grade equivalent, and 18 children (mean age = 8.8 years) reading at the sixth-grade equivalent or higher.

Design. The design was a 3 (comprehension level) by 5 (stimulus condition) by 3 (response color) mixed factorial. The reading groups were the grade equivalents of second, fourth, and sixth, as defined in the previous experiment. Stimulus conditions were incongruent color words, congruent color words, colored Xs, nonwords, and rectangles. The response colors were red, blue, and green. Each subject made 6 responses in each of the 15 stimulus by color conditions; the median response latency for each of these 6 trials was the dependent variable.

Results

The error data are summarized in Table 2. Most of the errors were again made in the incongruent condition, with few made in any other. No statistical analyses were performed on these data.

As in the previous experiment, the response time data were analyzed using the orthogonal polynomial trend analysis; reading comprehension groups were again taken to be equally spaced. For the planned comparisons, each of the other four stimulus conditions was set against the colored rectangles condition. The trend analysis of response times across the three levels of reading comprehension indicated that the five types of stimuli had similar curves and that only the constant components were significantly different. The differences between levels for each of the letter stimulus conditions and the rectangle condition were all reliable at or beyond the .01 level [for the incongruent words vs. rectangles, $F(1,68) = 166.10$; for nonwords vs. rectangles, $F(1,68) = 126.62$; for Xs vs. rectangles, $F(1,68) = 65.55$; and for congruent words vs. rectangles, $F(1,68) = 6.35$]. Response times decreased linearly with increased reading skill; the data are shown in Figure 2. The interaction of the linear component with stimulus condition approached significance only for the difference between the Xs and

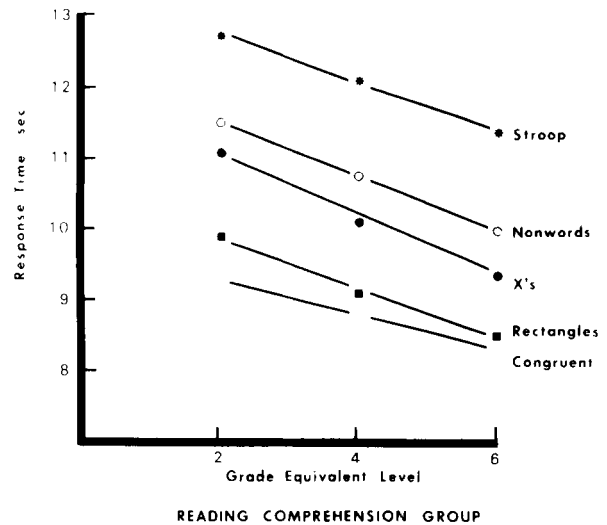


Figure 2. Mean response times in each stimulus condition as a function of reading skill in Experiment 2.

rectangles [$F(1,68) = 3.13$, $p = .08$], indicating that the decrease in mean response time for the Xs was slightly greater than that for the rectangles.

There was a main effect of color. Naming the red colors was somewhat faster than naming the blues and greens [$F(1,68) = 56.6$, $p < .01$]. The difference between response times for blue and green was not reliable, nor were the interactions of color with stimulus condition and reading group.

Discussion

There were both a facilitative effect on color naming from the congruent words and interference from the Xs. The differences between mean response times for the Xs condition and the rectangles and between the rectangles and the congruent pairs were reliable. Although we cannot assume that the linear relationships found among the stimulus conditions in this experiment are maintained as age and/or reading skill increases, there is no theoretical basis for predicting that the facilitative effect found for these reading skill groups should disappear in adulthood. Rather, the relative advantage for the congruent pairs over the rectangles should be maintained for college students, whose word processing should be quite efficient.

Since there is some evidence of greater deceleration in the response time curves for the Xs condition than for

Table 2
Means and Standard Deviations for Errors on the Stroop Task in Experiment 2

Reading Level	Number of Subjects	Stimulus Condition									
		Incongruous		Nonwords		Xs		Rectangles		Congruous	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Second Grade	25	4.66	2.78	.28	.53	.24	.51	.00	.00	.14	.58
Fourth Grade	28	4.96	2.77	.25	.52	.14	.45	.07	.26	.04	.19
Sixth Grade	18	4.44	2.20	.22	.55	.17	.38	.00	.00	.11	.32

the rectangles condition in this experiment and in the congruent condition in both experiments, it is possible that Xs may not reliably interfere in color naming by older and more highly skilled readers. However, the effect can and should be tested empirically, particularly considering the amount of Stroop research conducted with college students.

The source of the facilitative effect of the congruent word is not well understood. Color naming requires identifying the hue presented and accessing its name. Reading the color label involves accessing it. When the color and word are congruent, the name is accessed when the word is read, so the color name may have only to be checked with the name for accuracy rather than accessed, thus facilitating the color identification process.

Performances of the groups of younger children reading at the fourth- and sixth-grade equivalent levels in these two experiments were quite similar to each other. The consistency provides some evidence that, while the slow mean response time for the sixth-grade equivalent group in the first experiment may have been spurious, the pattern of interference was not. The sixth-grade equivalent readers' performance is more like that of the fourth-grade equivalent readers than like that of the older, pre-sixth-grade group. The initial temptation might be to infer that age is the major factor. The average age of the pre-sixth-grade children was 2.5 and 3.5 years older than the sixth-grade equivalent groups. Age in this case might represent some relevant factor such as overlearning of letter and word recognition skills.

However, it is at least equally plausible that the similarity of the sixth-grade equivalent readers to the fourth-grade equivalent group is a result of measurement error (as defined in classical test theory) in the reading comprehension test. The Gilmore and Gilmore (1968) test may provide unbiased measurement of the reading ability of the sample as a whole, but not necessarily of a subgroup of subjects selected because its members scored extremely high on the test. To the extent that the test is not perfectly reliable, the true reading ability of that extreme group is overestimated. Many subjects in that group are misclassified, in that their true reading ability (were it known) would put them in a lower group. The probability of misclassification of this sort is much greater in the extreme groups, such as the sixth-grade equivalent readers, than in more central groups. In most cases, there should be equal difficulty with misclassification in the lowest group, but in the first experiment, assignment to the lowest group was on the basis of an extremely reliable test (inability to read any of a list of simple words), and so the lower group was probably accurately placed.

We cannot distinguish between these two alternatives with our current data. The possibility of measurement error necessitates caution in interpreting results of

studies such as these but should not negate their use. Age and grade are more familiar measures, but not, *a priori*, better.

EXPERIMENT 3

The third experiment was designed to explore the development and the generality of the interference found in the Xs condition. If the rapid emergence of interference were due to the symbolic value of the Xs, then other letters used as symbols in children's games and literature, such as "O" and "Z," might also delay color naming for beginning readers. If familiarity were the major factor, then frequently used letters that were introduced early in the instructional process should also cause interference, whereas less frequently occurring letters that were introduced late in the curriculum should be less likely to do so. In order to evaluate the development of interference due to these characteristics of letters, we included colored rectangles and pronounceable and unpronounceable nonwords. Contrary to the tradition of the Stroop task, all words were excluded, in order to eliminate any possible effect that their presence on some trials might have on the other stimuli. If beginning readers can ignore letters, they should be most likely to do so when there are no words in the entire test array.

Method

Subjects. Subjects were 32 kindergarten children, tested in the late fall, plus 14 girls and boys tested within 1 month of entering the first grade. All children were Lawrence, Kansas, residents, had normal or corrected-to-normal sight and normal color vision, and could name the colors used. Each child was paid \$2 to come to the campus for testing. Data from four additional children were discarded because of erratic responding and/or equipment failure.

Equipment and materials. The equipment was the same as that used in the first two experiments, except that a solid state digital timer was used to time responses. Stimulus slides were prepared using the same photographic procedure as that used in the first two experiments. The symbolic letters were x, o, and z. The common letters were e, s, and t, and the uncommon letters were k, p, and v. All letters were presented in strings of four. The pronounceable nonwords were mafe, stib, and hund, and the unpronounceable nonwords were sbti, mhua, and dnfe. Common letters were selected from those in the upper third of the most frequently used letters in the alphabet (Kahn, 1967) and those also introduced in the first half of the reading readiness curriculum used in Lawrence kindergartens. Similarly, the uncommon letters were selected from the lower half of the most frequently used letters and from the last half of the reading readiness curriculum. The test items were grouped into 3 blocks of 18 trials each. Each block consisted of three slides, one of each color, for each of the six stimulus types, randomly ordered with the restriction that no color appear more than three times in succession and no two slides from the same interference condition appear more than twice in succession. Each letter and each nonword appeared only once in each color during the entire set of test stimuli. One example of each of the six types was included in the training trials.

The reading test consisted of three parts: a list of the letters

used as stimuli, the list of words used in the previous experiments, and the Gilmore Oral Reading Test (Gilmore & Gilmore, 1968). These were given following the color naming task. The same Ishihara color plates were used to test color vision.

Procedure. The kindergarten children were tested by one of two women in a room at their school. The first-graders were tested by one of two female experimenters, in a room in the psychology department of the University of Kansas. The remainder of the procedure and the instructions were the same as those in the first two experiments.

Design. The design was a 3 (comprehension level) by 6 (stimulus condition) factorial. The reading comprehension levels used in this experiment were (1) ability to name most of the letters but none of the words, (2) reading time for the word list of greater than 20 sec, and (3) a word list reading time of less than 20 sec. There were 15 children in the letters-only group (mean age = 5.6 years), 22 slow word readers (mean age = 6.0 years), and 9 in the fast word reading group (mean age = 6.4 years). The use of the letter and word reading lists enabled us to use the same measure to classify all groups.

Each child made three responses to each of the 18 color by stimulus conditions; the median of these three response times was the dependent variable.

Results

The mean number of errors ranged from .0 to .76, with no discernible pattern across condition or reading level. Only one child made as many as three errors in any condition. No statistical analyses were performed on the error trials.

As in the first two experiments, polynomial trend analysis was used to analyze the data. None of the main effects and interactions involving color was significant. The first planned comparison set the symbols against the common and uncommon strings. The overall means differed significantly [$F(1,43) = 5.29, p < .05$], but neither linear nor quadratic trends were significant. As can be seen in Figure 3, the greater interference from symbolic letters was limited to the group who could

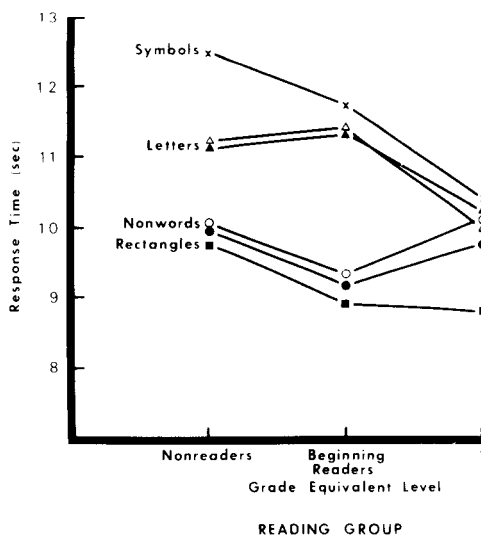


Figure 3. Mean response times in each stimulus condition as a function of reading skill in Experiment 3. Open triangles are uncommon letters; filled are common. Open circles are pronounceable nonwords; filled are unpronounceable.

name letters but could not read words. Letters having both symbolic and alphabetic value may be more meaningful to young children and may perhaps be recognized earlier than other letters. However, the advantage is quite temporary, and the shapes of the functions for symbols and letters do not differ.

The mean response times for the three letter types, symbolic, common, and uncommon, were tested against the means for the pronounceable and unpronounceable nonwords. More interference was observed for the letters than for the nonwords, as indicated by the significant mean differences [$F(1,43) = 112.31, p < .0001$]. The difference between linear trends was also reliable [$F(1,43) = 5.51, p < .05$], indicating that the slopes of the two functions were different. Response times for the letters tended to decrease across reading levels, whereas there was no overall linear change in response times for the nonwords as a function of reading level. Most important, the shape of the curve for the letter strings was concave downward, whereas the shape of the curves for the nonwords was concave upward. This difference was reliable, as indicated by the significant difference in the quadratic trends between groups [$F(1,43) = 15.70, p < .001$]. The interference from letters both developed earlier than and decreased prior to the comparable changes for nonwords. This difference is comparable to the results of the first experiment, in which the interference for the Xs both emerged and declined at lower levels of reading skill than did interference from the nonwords.

When the response times for all of the letter and nonword strings were combined and compared against the response times for the rectangle, the latter was significantly faster [$F(1,43) = 57.05, p < .0001$]. There are two important inferences to be drawn from this difference. First, automatic letter recognition as indicated by Stroop-like interference from letter strings develops very early in the process of learning to read. Second, inclusion of the traditional incongruent color words in the test array is not necessary to produce Stroop-like interference.

Discussion

One reason for using strings of Xs as the control condition has been that approximately the same area of color is projected for the Xs as for other letter strings. The rectangular patches obviously have a larger area, which could facilitate performance in this condition. The upper limit on any potential advantage from difference in amount of color may be estimated by the difference between mean response times to nonwords and rectangles for the letter naming and slow word reading groups. It appears to be a relatively constant, small difference of about 25 msec for this age group, and it may be even less for older subjects. This advantage, attributable to area differences, appears to be much smaller even for very slow responders than is the amount of interference attributable to strings of Xs. Further, the

shape of the function for response times for the Xs changes with reading skill, whereas we would not expect such a change for the rectangles.

Beginning readers do name colors more slowly when the interference is a string of identical letters than when it is a nonword or a rectangle, and letters that have symbolic value produce more interference with color naming by nonreaders. We have no idea how much exposure to letters, how much familiarity, is necessary to produce interference in color naming, but it would seem to be much less than LaBerge and Samuels (1974) originally implied would be necessary for automaticity.

Why do individual letters produce interference for the child at this low level of reading skill? For some period in the beginning of reading instruction, individual letters may be the most salient entities for the child. The alphabet is permanently displayed in most primary classrooms. Letters are named, isolated, blended, combined, written, recombined, and rewritten. Although some words may become highly familiar and thus automatically processed equally early in the course of learning to read, the letter is the common element. For the beginning reader, the letter is the smallest element to be identified and, usually, the largest to be processed as a unit; thus, it is distinctive.

GENERAL DISCUSSION

The results of these three experiments are remarkably consistent, despite differences in samples, minor procedural variations, and planned differences in stimulus conditions. The results of Experiment 2 are largely consistent with the performances of the second-, fourth-, and sixth-grade equivalent groups in the first experiment. Similarly, there is a great deal of consistency between the results of Experiment 3 and those for the first three reading groups in Experiment 1.

There are differences in the data also. The nonreaders in Experiment 1 showed no differences in mean color naming response times across stimulus conditions; those in Experiment 3 responded more slowly when the identical letters were the interferents. The latter group could name letters; the letter identification skills of the former were not tested. In the first experiment, interference in the Stroop condition peaked about the fourth-grade level; the amount of interference within all conditions remained relatively stable in the second experiment, in which the average age was 1-1.5 years older than in the first. We could only speculate on the basis for these differences, and they seem less important than the similarities.

Skilled or automatic word recognition is a complex operation that takes a long time to develop fully. Words and letters are highly salient to adults, and little knowledge seems necessary for them to become salient to children. Symbolic letters, particularly, interfered with color naming for children who could not read the simple

words on our list. All identical letter strings caused interference for the children who read the word list slowly.

The finding that interference from identical letter strings develops prior to interference from nonwords, which are essentially strings of mixed or different letters, suggests that the entire stimulus configuration determines whether or not attention will be drawn from the color naming task. If the presence of a single familiar letter, say the first, in a string had been sufficient to attract attention, then interference from the mixed (nonword) and identical letter strings should have emerged at approximately the same reading skill level. The identical string, however, is simpler and easier to parse and process than is the mixed string.

When interference from mixed letter strings first emerges, there seems to be little difference among words, pseudowords, and unpronounceable nonwords (see also Guttentag & Haith, 1978). This result suggests that the letter strings themselves are the source of the interference; the semantic properties of the words are not yet a significant factor.

By the second-grade reading level, the incongruent words caused more interference than did the nonwords. This difference, as well as the difference in amount of interference between intra- and extracategory words noted in the picture-word task (e.g., Guttentag & Haith, 1978; Rosinski, 1977), indicates that words come to be processed to a greater depth and their semantic content is accessed as reading skill increases. Pace and Golinkoff (1976) suggested that word decoding must occur easily if meaning is to be accessed in single-word recognition tasks. Thus we would expect the interference in Stroop-like tasks attributable to a semantic component to increase until decoding skills reach some asymptotic level and the meaning of familiar single words is reliably accessed. Such a peak is suggested at about the fourth-grade reading level in the first experiment but may have occurred by the second-grade level in Experiment 2.

We suggest that the increasing portions of the Stroop interference curves are due to increasing familiarity with and organization of the components of word recognition. As such knowledge develops, the depth to which the interfering word can be efficiently processed increases. Deeper processing places more demand on the cognitive system, and the result is more interference in the color naming task.

These same factors of increasing organization and processing efficiency may also contribute to decrements in interference as reading skill develops. More skilled readers should take less time to process the interfering stimuli than do less skilled readers. Greater control over cognitive processes also occurs with development. Thus the skilled reader may more easily disregard interfering stimuli, or, possibly, learn to process interferent and primary stimuli independently; some investigators have found only small and nondifferential amounts of inter-

ference for letter strings and words other than incongruent (intracategory) words for adults (Hintzman, Carre, Eskridge, Owens, Schaff, & Sparks, 1972; Lupker et al., Note 2). Similarly, the most reasonable explanation of the decrement in interference from color words for the adults in Evans' (Note 3) study is that his subjects learned to ignore the interfering words and process only the color.

The issue of decrements in interference in Stroop-like tasks with increases in reading skill merits further investigation. There is some evidence for decreasing interference in the letter and nonword conditions. However, there is little evidence for decreases in the actual Stroop condition within the reading skill range we tested. Most of the evidence for decreasing interference in the color-word task has been in studies utilizing the continuous list procedure (e.g., Comalli et al., 1962; Rand et al., 1959; Schiller, 1966), which confounds interference from the incongruent word with that from other words and color patches. Guttentag and Haith (1978, 1979) have evidence for decreasing interference in the picture-word task, as does Ehri (1976), who also used the continuous list procedure. The two interference tasks may differ in rate or in the actual course of development. In view of the importance of changes in interference among more skilled readers for models of word recognition and automatic processing, the issue needs to be explored further.

REFERENCE NOTES

1. Stanovich, K. E., Cunningham, A. E., & West, R. *Automatic recognition skills and reading ability in first-graders*. Paper presented at the annual meeting of the American Psychological Association, New York, September 1979.
2. Lupker, S. T., Cartman, E., & Symansky, J. *Picture-word interference: The role of response competition*. Paper presented at the annual meeting of the American Psychological Association, Toronto, Ontario, August 1978.
3. Evans, R. W. *Situational determination of processing strategy in the Stroop task*. Unpublished manuscript, Cornell University, 1979.

REFERENCES

- BOCK, R. D. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.
- COMALLI, P. E., WAPNER, S., & WERNER, H. Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, 1962, **100**, 47-53.
- DALRYMPLE-ALFORD, E. C. Associative facilitation and interference in the Stroop color-word task. *Perception & Psychophysics*, 1972, **11**, 274-276.
- DALRYMPLE-ALFORD, E. C., & BUDAYR, B. Examination of some aspects of the Stroop color-word test. *Perceptual and Motor Skills*, 1966, **23**, 1211-1214.
- DYER, F. N. The Stroop phenomenon and its use in the study of perceptual, cognitive, and response processes. *Memory & Cognition*, 1973, **1**, 106-120.

- EHRI, L. C. Do words really interfere in naming pictures? *Child Development*, 1976, **47**, 502-505.
- EHRI, L. C., & WILCE, L. S. Does word training increase or decrease interference in a Stroop task? *Journal of Experimental Child Psychology*, 1979, **27**, 352-364.
- GILMORE, J. V., & GILMORE, E. C. *Gilmore oral reading test*. New York: Harcourt Brace Jovanovich, 1968.
- GUTTENTAG, R. E., & HAITH, M. M. Automatic processing as a function of age and reading ability. *Child Development*, 1978, **49**, 707-716.
- GUTTENTAG, R. E., & HAITH, M. M. A developmental study of automatic word processing in a picture classification task. *Child Development*, 1979, **50**, 894-896.
- HINTZMAN, D. L., CARRE, F. A., ESKRIDGE, V. L., OWENS, A. M., SHAFF, S. S., & SPARKS, M. E. "Stroop" effect: Input or output phenomenon? *Journal of Experimental Psychology*, 1972, **95**, 458-459.
- ISHIHARA, S. *Tests for colour-blindness* (8th ed.). Chicago: Stoelting, 1939.
- JENSEN, A. R., & ROHWER, W. D. The Stroop color-word test: A review. *Acta Psychologica*, 1966, **25**, 36-93.
- KAHN, D. *The code breakers: The story of secret writing*. New York: Macmillan, 1967.
- LABERGE, D., & SAMUELS, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, **6**, 293-323.
- PAGE, A. J., & GOLINKOFF, R. M. Relationship between word difficulty and access of single-word meaning by skilled and less skilled readers. *Journal of Educational Psychology*, 1976, **68**, 760-767.
- POSNANSKY, C. J., & RAYNER, K. Visual-feature and response components in a picture-word interference task with beginning and skilled readers. *Journal of Experimental Psychology*, 1977, **24**, 440-460.
- POSNER, M. I., & SNYDER, C. R. R. Attention and cognitive control. In R. C. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J.: Erlbaum, 1975.
- RAND, G., WAPNER, S., WERNER, H., & MCFARLAND, J. H. Age differences in performance on the Stroop color-word test. *Journal of Psychology*, 1959, **47**, 259-266.
- ROSINSKI, R. R. Picture-word interference is semantically based. *Child Development*, 1977, **48**, 643-647.
- ROSINSKI, R. R., GOLINKOFF, R. M., & KUKISH, K. S. Automatic semantic processing in a picture-word interference task. *Child Development*, 1975, **46**, 247-253.
- SCHILLER, P. Developmental study of color-word interference. *Journal of Experimental Psychology*, 1966, **72**, 105-108.
- SEYMOUR, P. H. K. Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, 1977, **29**, 245-265.
- STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643-662.
- TREISMAN, A., & FEARNLEY, S. The Stroop test: Selective attention to colours and words. *Nature*, 1969, **222**, 437-439.

NOTE

1. "Stroop-like" will be used as a generic term to include the color and picture naming tasks. Stroop refers only to the color naming task.

(Received for publication April 10, 1980;
revision accepted September 4, 1980.)