*Journal of Reading Behavior*
*1981, Volume XIII, No. 1*

# A LONGITUDINAL STUDY OF THE DEVELOPMENT OF AUTOMATIC RECOGNITION SKILLS IN FIRST GRADERS

**Keith E. Stanovich and Anne E. Cunningham**
*Oakland University, Department of Psychology, Rochester, MI 48063*

**Richard F. West**
*James Madison University, Harrisonburg, VA*

*Abstract*. Experiment 1 was a study in which three times during the school year (in September, February, and April) first graders performed a discrete-trial Stroop task in which they named the colors of stimuli that were either letters, high-frequency words, or low-frequency words. The amount of interference caused by these stimuli was assessed by comparing the naming times to a control condition where the subject named a series of X's. In each testing period the interference caused by letters exceeded that caused by high-frequency words. There was also a nonsignificant tendency for interference caused by high-frequency words to exceed that caused by low-frequency words. There was a marked increase in interference between September and February, but very little change between February and April, indicating that the automaticity function had already flattened out by the end of first grade. There was a tendency for better readers to display more interference and to show interference earlier in the year. Experiment 2 replicated the developmental trends displayed in Experiment 1 and explored the relationship between interference and the speed and accuracy with which subjects named the stimuli. The overall pattern of results in the two experiments was reasonably consistent with the automaticity model of reading developed by LaBerge and Samuels (1974).

The automaticity model of reading developed by LaBerge and Samuels (1974) has had a great impact on research in the experimental and educational psychology of reading. LaBerge and Samuels argued that reading fluency develops because lower-level processes such as letter and word recognition become automated and thus free cognitive capacity for higher-level comprehension processes. An automated process

Portions of the data were presented at the annual meeting of the American Psychological Association, New York, September 1979. The authors would like to thank Karla M. Drayton and Dorothy J. Feeman for their assistance in data collection. Requests for reprints should be sent to Keith E. Stanovich, Department of Psychology, Oakland University, Rochester, Michigan 48063.

is one that can take place while attention is directed elsewhere. To date, two experimental paradigms have been used to assess the extent to which alphanumeric stimuli are recognized automatically. LaBerge (1973) introduced a catch-trial procedure whereby the subject's attention was directed to a particular stimulus set, but he was also required to respond to an unexpected stimulus that appeared on a small number of trials. Response time to the unexpected stimuli was the crucial dependent measure and the difference in reaction time between letters and meaningless lines (which were responded to as fast as the letters when both stimulus classes were expected) was an index of automaticity. Since the paradigm requires that attention be shifted from the expected to the unexpected stimulus, if some amount of processing can take place while the attention shift is occurring, then faster reaction times to letters under unexpected conditions would seem to indicate that the recognition of those stimuli was automatized to a greater extent than the meaningless lines.

The other experimental technique that has been used as an index of automatic processing is the Stroop task (Stroop, 1935). In the Stroop task subjects are asked to name the color of the ink in which a string of stimuli is printed. When the string is a series of letters that spell the name of a conflicting word (e.g., the word "red" written in blue ink) color-naming is much slower than in a control situation where the string consists of nonverbal stimuli. This color-word interference effect is usually explained in terms of the competition between vocal responses to the printed word and the ink color. Since the subjects engaged in a Stroop task are attempting to attend only to the color of the ink, color-word interference is presumably the result of the word having been read automatically (Posner & Snyder, 1975).

In light of the wide influence of the automaticity theory (Science Citation Index lists 41 references for 1976 and 1977 alone) it is indeed surprising that so few attempts have been made to directly test its predictions regarding developmental trends and individual differences. The few studies that have tested the theory have all employed Stroop-type paradigms (suggesting that it might be fruitful to explore the possibilities of the catch-trial procedure). If, as suggested by automaticity theory, skilled readers process words more automatically than less skilled readers, a larger Stroop effect might be expected for the more skilled readers. Evidence supportive of this prediction for adults has been reported by Martin (1978). However, studies using Stroop-type paradigms with children have often failed to find the expected relationship between reading skill and the Stroop effect. For example, Schiller (1966) found that color-word interference was maximal in the second grade and decreased with further development. Using a picture-word interference task that is a variant of the Stroop paradigm, Golinkoff and Rosinski (1976) found that the picture naming times of unskilled third- and fifth-grade readers were delayed by the presence of incongruent words just as long as the times of skilled readers of the same age. In another study, Rosinski, Golinkoff, and Kukish (1975) found no increase in the interference effect from second grade through adulthood. Pace and Golinkoff (1976) also reported results that were not entirely consistent with the automaticity theory. Using a picture-word interference task, they found that when interfering words are difficult, skilled third graders displayed more interference than less-skilled third graders. However, the relationship did not obtain in the times of the fifth grade subjects, and when easy words were used there was no relationship between reading skill and interference in either grade. West and Stanovich (1978) found no increase in Stroop interference between fourth grade and adulthood (compare their Figures 2 and 3). Using a picture-word interference task, Guttentag and Haith (1978) found no increase in interference due to the presence of words from late first grade through adulthood (see also, Guttentag & Haith, 1979). The only significant increase in interference due to words was found between first graders tested in October and first graders tested in May.

Ehri and Wilce (1979) tested a different prediction of the automaticity theory, using a picture-word interference task with first and second graders. According to the theory, practice at word recognition should lead to automaticity. Thus, more interference should be found for words that have been practiced. Ehri and Wilce (1979) had children complete a picture-word interference task, gave them practice at recognizing the interfering words, and then gave them a posttest on the interference task. Subjects who could not recognize many of the interfering words prior to the practice displayed increased interference subsequent to the word training. In contrast, subjects who could recognize all of the interfering words prior to the word training actually showed a decrease in interference, a finding that is inconsistent with the automaticity theory as conceptualized by LaBerge and Samuels (1974). However, Ehri and Wilce (1979) argued that this finding was not inconsistent with automaticity theory. They hypothesized that, in the latter group of subjects, the word and picture stimuli are processed serially, in a manner such that practice enables the word to move through the central processor in less time and allow the picture to enter sooner. This hypothesis is, however, not consistent with the results of a study by Martin (1978), who found that adult subjects who recognized words faster actually showed larger interference effects.

There are probably several reasons why many developmental studies have failed to find the expected relationship between reading ability and automatic processing. One has to do with the widespread use of the continuous-list procedure, where the subject names a series of items and his score is the total time to name the entire list. The continuous-list procedure involves complex articulatory and sequential-response processes (Proctor, 1978), in addition to making possible nonperceptual strategies for overcoming interference that might change with age (Posnansky & Rayner, 1977). The discrete-trial procedure, where a reaction time to a single stimulus is recorded, is a much more precise method and is therefore preferred over the continuous-list procedure. It is thus relevant to note that two studies that have provided some support for the predictions of automaticity theory (Guttentag & Haith, 1978; West and Stanovich, 1979) have both used the discrete-trial procedure. Differences between the two procedures are also probably the cause of the discrepancy between the results of Martin (1978) and Ehri and Wilce (1979).

Other possible reasons for the lack of support for the automaticity hypothesis (in addition to the possibility that the theory is wrong) involve the reading level of the subjects used in the experiments and the type of words that have been employed. The words used in the Golinkoff and Rosinski (1976) and Rosinski et al. (1975) studies were simple and fairly familiar. The use of highly familiar words, coupled with the fact that even the second graders and unskilled third-grade subjects in these studies were well beyond the initial stages of reading acquisition, may have resulted in a sort of "ceiling effect" whereby the levels of the words and subjects were not in a range where the reading ability-automaticity relationship could be uncovered. This conjecture is supported by research suggesting that the expected relation begins to become apparent when either more difficult words are employed (Pace & Golinkoff, 1976;

West and Stanovich, 1979) or poorer readers are tested (Ehri, 1976; Guttentag & Haith, 1978; West and Stanovich, 1979).

In Experiment 1 an attempt is made to remedy some of the problems discussed above. An age range is investigated in which large changes in automatic processing of written material would be expected. A color-naming task is used in which interfering stimuli are letters, high-frequency words, and low-frequency words (i.e., stimuli that differ in frequency of exposure, and presumably automatization). Thus, changes in the automatic processing of these different types of written items can be observed across the age range under consideration. A ceiling effect would not be expected with low-frequency words. Finally, a discrete-trial presentation procedure is used that is more precise and is less amenable to response strategies than is the continuous-list procedure.

Two recent studies using the discrete-trial procedure (Guttentag & Haith, 1978; West & Stanovich, 1979) have strongly suggested that it is during first grade that the most marked changes in automatic processing occur. However, both studies investigated fairly wide age ranges and neither was able to clearly delineate the development of automaticity during the crucial first-grade period. Thus, Experiment 1 was designed as a longitudinal investigation[1] where the same first-grade subjects were tested at the beginning of the year, mid-year, and at the end of the year. It was hoped that the result would be a more precise picture of how the automatic recognition of letters, high-frequency, and low-frequency words develops during the first grade. Further, neither the Guttentag & Haith (1978) nor the West and Stanovich (1979) study separated skilled from less-skilled readers in their first-grade subjects. This was done in the present study since several investigators (e.g., LaBerge & Samuels, 1974) have suggested that tasks tapping automatic processing might have diagnostic value. It was thus thought important to see if the ability to automatically process written material throughout the first-grade year relates to end of year reading proficiency.

## EXPERIMENT 1

### Method

*Subjects*. The subjects were 24 first-grade children (14 males and 10 females) recruited from a predominantly middle-class elementary school. The children were tested in late September, mid-February, and late April of the school year. At the end of the year (early June) the teacher was asked to rank order the 24 children on the basis of their reading ability. The top 12 readers comprised the skilled group and the bottom 12 comprised the less-skilled group. The children in each group were administered Reading Subtest Level I of the Wide Range Achievement Test (Jastak, Bijou & Jastak, 1965), the Reading subtest (sections A and B, Primary Level 1) of the Stanford Achievement Test, and a short paragraph that was read orally and was timed by the experimenter (two subjects did not complete the latter two measures). An

[1]Prior to the execution of the studies reported here there had not been a longitudinal investigation of the develoment of automaticity in first-graders. After the present studies had been prepared for publication, the research of Guttentag and Haith (1980) appeared. Although their design differed considerably from that of Experiments 1 and 2, their results were largely consistent with ours, and thus served to increase our confidence in the reliability of the findings.

analysis of the WRAT scores indicated that the mean reading ability of the skilled readers was at the 3.0 grade level and the mean reading ability of the less-skilled readers was at the 1.8 grade level. There was no overlap in the WRAT scores of the two groups of children and the difference in means between the two groups was highly significant, $t(22) = 5.05$, $p < .001$. According to the Stanford scores, the mean reading ability of the skilled readers was at the 3.0 grade level and the mean reading ability of the less-skilled readers was at the 1.6 grade level, a difference that was highly significant, $t(20) = 3.78$, $p < .001$. Skilled readers read the paragraph in a mean time of 65 seconds, whereas the mean reading time of the less-skilled readers was 116 seconds, $t(20) = 4.36$, $p < .001$. Not surprisingly, a check on the actual progress the children had made through beginning reading materials also confirmed the teacher's ratings. Although they were not the main focus of the research, a group of 24 second graders were tested in late April in order to insure that the assumption (based, of course, on previous research) that there would be little development in automatic recognition skills after grade one was valid.

*Stimuli and Apparatus*. The stimuli consisted of the 26 letters of the alphabet, 26 high-frequency words, 26 low-frequency words, and strings of X's (the control condition). The selection of the words was based on an inspection of the Dolch list, the Kucera and Francis (1967) corpus, and a number of reading primers. The low-frequency words had a mean frequency of 16.0 based on the Kucera and Francis (1967) count and 24.0 based on the Carroll, Davis, and Richman (1971) count of third-grade literature. The high-frequency words had a mean frequency of 326.5 based on Kucera and Francis and 509.2 based on Carroll et al. (1971). The stimuli were typed in IBM Courier 72 font. Only the lowercase was used. Black and white negatives of the stimuli were mounted on slides that were colored either red, yellow, blue, or green with acetate film. The slides were projected onto a white screen by a Kodak Carousel 760H projector. Thus, the stimuli appeared as colored symbols on a white background. Subjects sat approximately 90 cm from the screen, and the size of the projection was such that a five-letter word subtended a horizontal visual angle of approximately 3 degrees. Letter stimuli were centered on the slide. Stimulus onset was controlled by a Vincent Associates Uniblintz Shutter that was positioned over the lens of the projector. When the experimenter pushed a control button, the shutter was electronically opened, and the projected image of the stimulus item appeared. Simultaneously, a Lafayette Instruments electronic clock (Model S4419-A, accurate to the millisecond) was started by the same push of the control button. When the subject verbally responded, a voice activated relay stopped the clock and closed the shutter. The microphone that led to the voice activated relay was held by the subject.

*Procedure*. The first-grade subjects were tested three times during the school year, in late September, mid-February, and late April. The second-grade subjects were tested in late April. The procedure was the same for each testing. Subjects were instructed to name the color of the stimulus as rapidly as possible. The subjects first completed 25 trials (5 practice and 20 experimental trials) in which they named the colors of strings of X's. This condition provided a baseline so that interference effects in the other conditions could be measured. Subjects then completed 78 trials in which they named the colors of letters and words. The block of 78 experimental trials consisted of a random ordering of 26 trials of each of the three stimulus types (letters, high-frequency words, and low-frequency words). The distribution of colors was the same across all three stimulus types.

*Results and Discussion*

Trials on which the subject failed to articulate the color or had a response time more than three standard deviations from the mean of that condition were scored as subject errors and dropped from the analysis. These errors occurred on 8% of the trials during the first testing, 6% of the trials during the second testing, 4% of the trials during the third testing, and were approximately equally frequent across the experimental conditions. Analyses were based on each subject's mean correct reaction time in each of the conditions. The mean reaction time in the baseline condition was 1024 msec for the first testing, 784 msec for the second testing, and 794 msec for the third testing. Interference scores were formed by dividing the subject's mean time in the baseline condition into his mean time in each of the three experimental conditions. These interference ratio scores were used in the analyses reported below. However, it should be noted that virtually identical results were obtained when the analyses were carried out on the difference scores formed by subtracting the mean time in the baseline condition from the mean time in each of the experimental conditions. The results also did not change when median times rather than means were employed.

The interference scores for all 24 subjects as a function of testing period and experimental condition are displayed in Table 1. An analysis of variance on the interference ratios indicated that the effects of time of testing, $F(2, 46) = 18.77$, and experimental condition, $F(2, 46) = 20.66$, were both significant at the .001 level, but that the experimental condition by time of testing interaction did not approach significance. The interference ratios in all conditions were significantly different from unity at the .001 level, except the low- and high-frequency word conditions in the first testing period. However, the high-frequency word condition did approach significance ($p < .10$) in the first testing period. Planned orthogonal comparisons indicated that the interference ratios for letters exceeded the mean interference ratio of the two word conditions in every testing period ($p < .005$, 005, and .025, respectively) and that the two word conditions did not differ significantly from each other. It should be noted, however, that in each testing period the interference ratio for the high-frequency words exceeded that of the low-frequency words.

TABLE 1

Mean Interference Ratios as a Function of Time of Testing and Experimental Conditions (Mean Difference Scores are Indicated in Parentheses)

| | Experimental Condition | | |
|---|---|---|---|
| Testing Period | Letters | High-frequency Words | Low-frequency Words |
| 1. Late September | 1.15* (148) | 1.05 (39) | .98 (-18) |
| 2. Mid February | 1.41* (329) | 1.30* (251) | 1.25* (201) |
| 3. Late April | 1.42* (332) | 1.33* (273) | 1.28* (238) |

* Significantly different from 1.00 at the .001 level.

The interference scores for the skilled and less-skilled readers are presented separately in Table 2. The baseline times for the skilled readers were 998 msec, 737 msec, and 809 msec, for the December, February, and April testing periods respectively. The corresponding baseline times for the less-skilled readers were 1050 msec, 831 msec, and 779 msec. Due to the relatively small number of subjects and the large variability in the response times, these two groups did not differ significantly from each other, nor was there an interaction between reader skill and experimental condition. Nevertheless, a suggestive pattern was apparent. Specifically, in eight of the nine conditions the skilled readers displayed larger interference scores, suggesting that these subjects were experiencing greater interference. Also, for the skilled readers in the first testing period the interference ratio for the high-frequency words was marginally ($p < .10$) different from 1.0, whereas the scores of the less-skilled readers in this condition did not approach significance. For the skilled readers, planned orthogonal contrasts indicated that the mean ratios for letters was significantly higher than the mean of the two word conditions in testing periods one and two but not in testing period three. For the less-skilled readers, this contrast was significant in each testing period. The interference scores of the high-frequency words showed marginal tendencies ($p < .10$) to exceed the scores of the low-frequency words during the third testing period for the less-skilled readers.

TABLE 2

Mean Interference Ratios as a Function of Time of Testing, Experimental Condition, and Reading Ability (Mean Difference Scores are Indicated in Parentheses)

| Testing Period | Letters | High-frequency Words | Low-frequency Words |
|---|---|---|---|
| | Experimental Condition | | |
| **Skilled Readers** | | | |
| 1. Late September | 1.17** (162) | 1.06 (55) | .99 (-7) |
| 2. Mid February | 1.46*** (345) | 1.39*** (299) | 1.31*** (243) |
| 3. Late April | 1.41*** (324) | 1.40*** (315) | 1.36*** (318) |
| **Less-skilled Readers** | | | |
| 1. Late September | 1.13* (134) | 1.03 (24) | .98 (-30) |
| 2. Mid February | 1.36*** (313) | 1.22** (203) | 1.18* (159) |
| 3. Late April | 1.43** (341) | 1.27** (230) | 1.19* (158) |

* Significantly different from 1.00 at the .05 level.
** Significantly different from 1.00 at the .01 level.
*** Significantly different from 1.00 at the .001 level.

Finally, the interference scores of the second-grade subjects are displayed in Table 3. The mean reaction time in the baseline condition was 823 msec for these subjects, and all three interference ratios were significantly different from 1.0 ($p < .001$ in all cases). The interference scores of the second graders were almost identical to those displayed by the skilled first graders during the last testing period, indicating that virtually no further development of automatic recognition skills occurs for these subjects after first-grade.

TABLE 3

Mean Interference Ratios of the Second Graders
As a Function of Experimental Condition
(Mean Difference Scores are Indicated in Parentheses)

| Experimental Condition | | |
|---|---|---|
| Letters | High-frequency Words | Low-frequency Words |
| 1.39 (315) | 1.41 (336) | 1.39 (318) |

As a whole, the results displayed a coherent pattern that was reasonably consistent with automaticity theory. Predictions regarding developmental trends were strongly supported, but the automaticity measures were only weakly related to individual differences in reading skill. The combined data presented in Table 1 indicated that at the beginning of the year the first graders had to some extent automatized the recognition of letters, but not words. Better readers displayed a slight tendency toward automatization of the high-frequency words, but this trend did not quite reach statistical significance. A marked increase in the automatization of all three stimulus types occurred between September and February. By February all three stimulus classes displayed significant interference. There was very little change in interference scores between February and April, indicating that the automaticity function had already flattened out by the end of the first grade. Indeed, the interference scores of the skilled readers during the last testing period were nearly identical to those displayed by the second graders tested late in the year. This finding is consistent with previous research that has indicated that words may become automatized after only a few exposures and that the development of automatic recognition takes place primarily in the first year of reading instruction (Barron & Baron, 1977; Ehri & Wilce, 1979; Guttentag & Haith, 1978, 1979; West & Stanovich, 1979). Finally, the ordering of experimental conditions were consistent with automaticity theory in every testing period.

The analysis of individual differences was mildly supportive of the automaticity theory. The variability inherent in the responses of subjects as young as first graders contributed to the failure to detect a significant difference in interference between the skilled and less-skilled readers. Nevertheless, a pattern consistent with automaticity theory is clearly apparent in the data displayed in Table 2. The skilled readers had higher interference scores in all conditions except one (that condition was the third-testing letter condition, where the performance of both groups should be at asymptote, thus rendering this one reversal less of an exception than it appears). From Table 2 it also appears that by the end of the year the skilled readers had automatized the recognition of the three stimulus types to an equal extent, whereas in the less-skilled readers automatization of words still lagged behind that of letters. Planned contrasts confirmed the statistical reliability of this pattern and it is also consistent with the results of West and Stanovich (1979). Nevertheless, from the standpoint of the individual difference predictions of automaticity theory, the lack of a statistically significant difference in interference between ability groups is troublesome. Whereas speed of word naming is strongly related to reading ability (e.g., Biemiller, 1977–1978; McCormick & Samuels, 1979; Stanovich, 1980, 1981), the relationship between automaticity and reading ability appears to be rather weak. For example, correlational analyses were carried out on some additional data that was collected on the first graders during April, as part of another study. These subjects named twenty words that were presented via discrete-trial procedure. The mean naming times were highly correlated with WRAT scores ($r = -.60$, $p < .005$), Stanford scores ($r = -.51$, $p < .01$), and paragraph reading time ($r = -.71$, $p < .001$). In contrast, the correlations between the interference ratios and the three measures of reading ability (see Table 4), although mostly in the right direction, were small in size and many did not reach statistical significance.

TABLE 4

Correlations Between Measures of Reading Ability And Interference Ratios

| Testing Period | Experimental Condition: Letters | High-frequency Words | Low-frequency Words |
|---|---|---|---|
| **WRAT** | | | |
| 1. Late September | -.04 | .04 | .16 |
| 2. Mid February | .12 | .45* | .40* |
| 3. Late April | -.17 | .18 | .21 |
| **Stanford** | | | |
| 1. Late September | .17 | .22 | .36 |
| 2. Mid February | .25 | .49** | .38* |
| 3. Late April | -.07 | .11 | .22 |
| **Paragraph Reading Time** | | | |
| 1. Late September | -.18 | -.16 | -.10 |
| 2. Mid February | -.29 | -.50** | -.35 |
| 3. Late April | -.10 | -.27 | -.44* |

* $p < .05$
** $p < .01$

The above results suggest the importance of distinguishing between automaticity and speed, a point emphasized by Ehri and Wilce (1979) who argue that success in recognizing words can be assessed in regard to three criteria: accuracy, automaticity, and speed. One could conceptualize these as three stages (the stages may be overlapping, strict seriality is not to be implied) that are traversed in the development of word recognition skills. First, the reader develops the ability to accurately identify a given word. Next the reader develops the ability to recognize the word without allocating attention to it (i.e., the recognition of the word is automatized). Finally, during the time that automaticity is developing, and even after the word is fully automatized, recognition speed continues to increase. The latter point is often lost in discussions that center on the automaticity theory itself, even though there is ample evidence in the literature that recognition speed continues to increase after words have become fully automatized. Several studies have failed to find an increase in automatic processing after the second- or third-grade reading level has been reached (Golinkoff & Rosinski, 1976; Guttentag & Haith, 1978, 1979; Pace & Golinkoff, 1976; Posnansky & Rayner, 1977; Rosinsky, Golinkoff & Kukish, 1975; West & Stanovich, 1978). In contrast, marked increases in word recognition speed occur as children progress beyond the second-grade level of reading ability (Biemiller, 1977–1978; McCormick & Samuels, 1979; Perfetti, Finger & Hogaboam, 1978; Perfetti & Hogaboam, 1975; Stanovich, 1980; West & Stanovich, 1978). Thus, it is possible that by the end of the first grade the better readers in the present study had fully automatized the recognition of words and their further progress in reading was more dependent on the development of speed rather than automaticity.

## EXPERIMENT 2

It was thought desirable to replicate the trends in the development of automatic recognition skills that were displayed in Experiment 1, since few studies have investigated this crucial age range. Furthermore, Experiment 2 provided the opportunity to make one important change and one important addition to the methods employed in Experiment 1. Schadler and Thissen (Note 1) have suggested that the color-naming baseline condition consisting of a string of X's may be inappropriate, because the X's themselves may be processed as letters and cause interference. Thus, in Experiment 2 a nonverbal stimulus, string of letter-size rectangles, served as the control condition. Also, in order to further examine the issues raised above regarding the relationships of speed and automaticity to reading ability, in Experiment 2 measures of how fast the subjects named the letters and words in isolation were obtained in addition to measures of how much interference was caused by these stimuli. This additional aspect also allowed for an examination of how the ease of processing a word in isolation is related to the amount of interference caused by that word in a Stroop situation, an analysis at the word level that relates importantly to the adequacy of automaticity theory.

### *Method*

*Subjects*. The subjects were 24 first-grade children (12 males and 12 females) recruited from the same school as those in the first study. The children were tested twice during the school year, in mid-December and mid-March. Twelve students were defined as skilled and 12 as less-skilled on the basis of teacher rankings as in Experiment 1. An analysis of the children's scores on the Reading subtest of the Stanford Achievement Test indicated that the mean reading ability of the skilled readers was at the 3.4 grade level and the mean reading ability of the less-skilled readers was at the 1.8 grade level, $t(22) = 5.40$, $p < .001$. The two groups also displayed significantly different scores on the WRAT, $t(22) = 4.18$, $p < .001$, and the Reading Survey Test (Primary 1, Form JS) of the Metropolitan Achievement Tests, $t(22) = 5.70$, $p < .001$. The skilled readers read a short paragraph in 33.9 seconds, whereas the less-skilled readers took 79.1 seconds to read the same paragraph, $t(22) = 3.95$, $p < .005$.

*Stimuli and Apparatus*. The stimuli and apparatus were the same as those employed in Experiment 1 except that a new baseline condition was used and an additional set of 78 stimuli were constructed. The baseline condition consisted of four line drawings of small rectangles, each approximately the size of a letter, which had their lines colored in a manner similar to the letters. These baseline stimuli were thus nonverbal, but had approximately the same amount of colored area as the words. The 78 new stimuli were the same letters and words that were employed in the Stroop condition, but the colored acetate films were not added to the black and white negatives. Thus, these slides would serve as the stimuli for the condition where the subjects named the letters and words in isolation.

*Procedure*. The procedure was the same for each testing. The Stroop condition was administered as in Experiment 1. Approximately one week after completing the Stroop condition, subjects were administered the 78 noncolored stimuli in the same manner as in the Stroop condition, except that the subjects were told to try to name the letters and words as quickly as possible.

### *Results and Discussion*

Trials on which the subject failed to articulate the correct stimulus, had a response time greater than three standard deviations from the mean of that condition, or failed to produce a response within 4000 msec were scored as subject errors and dropped from the analysis. In the Stroop conditions these errors occurred in 4% of the trials during the first testing and 5% of the trials during the second testing. In the word and letter naming conditions these errors were, of course, much more numerous and will be analyzed separately. Analyses were based on each subject's mean correct reaction time in each of the conditions. Interference scores for the Stroop conditions were formed as in Experiment 1. The interference ratio scores were used in the analyses reported below. It was again the case that virtually identical results were obtained when the analyses were carried out on the different scores, and analyses based on medians also did not change the results.

The interference scores as a function of testing period and experimental condition are displayed in Table 5. The mean reaction time in the baseline condition was 840 msec for the first testing and 760 msec for the second testing. The interference ratios in all of the conditions were all significantly different from unity at the .001 level. An analysis of variance on the interference ratios indicated that the effects of testing period, $F(1, 23) = 12.62$, $p < .005$; experimental condition, $F(2, 46) = 7.96$, $p < .005$; and the testing period by experimental condition interaction, $F(2, 46) = 5.62$, $p < .01$, were all significant. Thus, all of the developmental trends displayed in

Experiment 1 were replicated. Automaticity increased with time. Letters were more automatized than high-frequency words, which were in turn more automatized than the low-frequency words. The significant interaction indicated that the interference for words increased more markedly than that for letters. By the March testing, the interference ratios for the three stimulus types were nearly equivalent, indicating, as did the results of Experiment 1, that the subjects were near asymptote. Thus, the general pattern of the developmental results confirmed automaticity theory.

TABLE 5

Mean Interference Ratios as a Function of Time of Testing And Experimental Conditions (Mean Difference Scores are Indicated in Parentheses)

| | Experimental Condition | | |
|---|---|---|---|
| Testing Period | Letters | High-frequency Words | Low-frequency Words |
| 1. December | 1.57 (470) | 1.44 (351) | 1.38 (312) |
| 2. March | 1.71 (530) | 1.72 (536) | 1.83 (477) |

Table 6 displays a breakdown of the interference scores by reading ability. In the December testing the baseline time was 833 msec for the skilled readers and 848 msec for the less-skilled readers. In the March testing the baseline time was 761 msec for the skilled readers and 760 msec for the less-skilled readers. These data replicated the pattern of results in Experiment 1. The difference between the interference scores of the two ability groups did not reach statistical significance. However, in the word conditions, at each of the two testing periods, the skilled readers displayed larger interference scores. In addition, the reader ability by experimental condition interaction approached significance ($F(2, 44) = 3.06$, $p < .06$; the analysis of variance on the difference scores indicated $p < .04$). As is clear from Table 6, the interaction results because the interference scores of the skilled readers were more nearly equivalent across stimuli type, whereas the scores of the less-skilled readers showed greater differences across stimuli type. This pattern suggests that the skilled readers had automatized the three stimulus types to almost an equal degree and were probably near asymptote. The less-skilled readers, however, had not developed the ability to automatically recognize words to an asymptotic level. Although the pattern of results regarding reading ability is consistent with automaticity theory, the trends, as in Experiment 1, were rather weak. A consideration of the correlational data also confirms this conclusion and the results of Experiment 1. Although most of the correlations between the interference ratios and the four measures of reading ability were in the expected direction, very few reached statistical significance.

TABLE 6

Mean Interference Ratios as a Function of Time of Testing, Experimental Condition, and Reading Ability (Mean Difference Scores are Indicated in Parentheses)

| | Experimental Condition | | |
|---|---|---|---|
| Testing Period | Letters | High-frequency Words | Low-frequency Words |
| **Skilled Readers** | | | |
| 1. December | 1.57 (456) | 1.50 (396) | 1.47 (378) |
| 2. March | 1.70 (523) | 1.76 (561) | 1.69 (514) |
| **Less-skilled Readers** | | | |
| 1. December | 1.58 (484) | 1.37 (308) | 1.30 (246) |
| 2. March | 1.71 (537) | 1.68 (511) | 1.58 (439) |

The mean times to name the stimuli in isolation and the mean number of errors are displayed in Table 7 as a function of testing period and reading ability. An analysis of variance on the naming times indicated that the effects of testing period, experimental condition, and reading ability were all significant ($p < .001$, $p < .001$,

TABLE 7

Mean Naming Times as a Function of Time of Testing, Experimental Condition, and Reading Ability (Mean Number of Errors are Indicated in Parentheses)

| | Experimental Condition | | |
|---|---|---|---|
| Testing Period | Letters | High-frequency Words | Low-frequency Words |
| **Skilled Readers** | | | |
| 1. December | 1054 (1.7) | 1521 (8.8) | 2171 (21.1) |
| 2. March | 1017 (1.7) | 1189 (3.1) | 1618 (13.4) |
| **Less-skilled Readers** | | | |
| 1. December | 1102 (2.0) | 1763 (20.4) | 3782 (25.0) |
| 2. March | 991 (2.3) | 1457 (10.2) | 2292 (22.5) |

and $p < .01$, respectively). The experimental condition by testing period and the experimental condition by reading ability interactions were both significant ($p < .001$), as was the three-way interaction ($p < .025$). An analysis of variance on the number of errors revealed trends parallel to those in the naming time analysis. The effects of testing period, experimental condition, reading ability, the experimental condition by testing period, the experimental condition by reading ability, and the three-way interaction were all significant at the .001 level. The two ability groups differed markedly in the ability to name the stimuli accurately and rapidly. Interestingly, there was no difference between the two groups in their ability to name letters, indicating that the large differences in word naming were not due to an inability on the part of the less-skilled readers to deal with the visual components of words. The large differences in word naming accuracy and speed were, of course, consistent with much

TABLE 8

Correlations of Reading Ability with Naming Time and Errors for each Experimental Condition

| Testing Period | Letters | | High-frequency Words | | Low-frequency Words | |
|---|---|---|---|---|---|---|
| | Naming Time | Errors | Naming Time | Errors | Naming Time | Errors |
| **WRAT** | | | | | | |
| December | -.18 | -.12 | -.51** | -.77*** | -.56** | -.78*** |
| March | -.15 | -.10 | -.44* | -.66*** | -.42* | -.85*** |
| **Stanford** | | | | | | |
| December | -.21 | .21 | -.41* | -.73*** | -.60*** | -.46* |
| March | -.04 | -.21 | -.41* | -.83*** | -.75*** | -.67*** |
| **Metropolitan** | | | | | | |
| December | -.11 | .13 | -.37* | -.68*** | -.63*** | -.42* |
| March | -.02 | -.37* | -.44* | -.87*** | -.58*** | -.72*** |
| **Paragraph Reading Time** | | | | | | |
| December | .12 | -.26 | .44* | .69*** | .55** | .40* |
| March | .17 | .13 | .51** | .87*** | .71*** | .59*** |

Experimental Condition (spanning header over Letters, High-frequency Words, Low-frequency Words)

*$p < .05$
**$p < .01$

previous research (Biemiller, 1977–1978; McCormick & Samuels, 1979; Perfetti & Hogaboam, 1975; Perfetti, Finger & Hogaboam, 1978; Shankweiler & Liberman, 1972; Stanovich, 1980). It should also be noted that ability differences in word naming speed were apparent even though all words that were not known to the subject or that produced unusually long reaction times were eliminated from the naming time analysis. Table 8 displays the correlations between the four measures of reading ability and the naming times and errors. Consistent with the analysis above, letter naming did not correlate with reading ability, although word naming times and errors displayed strong relationships.

The relationships between the speed and accuracy of word naming and interference is interesting. Apparently, a considerable amount of interference can be caused by words that the subject cannot even name correctly (compare the low-frequency results in Tables 6 and 7). Presumably, this interference is caused by the components of words (letters and familiar letter clusters) rather than the word name itself. These relationships can be examined by correlating the speed and accuracy with which a subject names the stimuli and the amount of interference experienced by that subject. Table 9 displays these correlations as a function of testing period and stimulus type. No correlations approached significance for the letter stimuli, perhaps due to ceiling effects. For words, particularly high-frequency words, there was a mild tendency for subjects who named the words more rapidly and accurately to display greater interference, a trend predicted by automaticity theory. Another way to examine these relationships is in terms of stimuli rather than subjects. That is, do stimuli that are named more rapidly and accurately produce more interference? In order to address this question the data were collapsed across subjects in order to form a mean naming time, accuracy, and interference score for each stimulus. The interference scores for the stimuli were then correlated with naming time and accuracy. These correlations are displayed in Table 10 as a function of testing period and reading ability. The prediction of automaticity theory, that stimuli are named rapidly and accurately will display more interference, was confirmed in the data of the less-skilled, but not the skilled readers.

TABLE 9

Correlations of Interference with Naming Time and Errors for Each Experimental Condition (Data Collapsed Across Stimuli)

| Testing Period | Experimental Condition | Naming Time | Errors |
|---|---|---|---|
| December | Letters | .15 | .22 |
| December | High-frequency words | −.32* | −.46** |
| December | Low-frequency words | −.12 | −.02 |
| March | Letters | .01 | .10 |
| March | High-frequency words | −.28* | −.37** |
| March | Low-frequency words | −.13 | −.36*** |

* $p < .10$
** $p < .05$

TABLE 10

Correlations of Interference with Naming Time and Errors as a Function of Testing Period and Reading Ability (Data Collapsed Across Subjects)

| Testing Period | Naming Time | Errors |
|---|---|---|
| | Skilled Readers | |
| December | .03 | -.20* |
| March | .11 | -.12 |
| | Less-skilled Readers | |
| December | -.36** | -.61*** |
| March | -.21* | -.24* |

*$p < .05$
**$p < .01$
***$p < .001$

## CONCLUSION

Experiments 1 and 2 provided valuable data regarding developmental changes in automatic recognition. It was possible to trace the development of automaticity in some detail because the crucial first-grade period (which previous research had identified as important) was the focus of the research effort. Based on the results reported here (and previous work) it appears that, indeed, a sharp increase in automaticity occurs during the first grade, but that by the end of the year the development of automaticity has begun to level off. This trend is particularly true for skilled readers, who appear to have automatized the recognition of letters, high-frequency words, and some low-frequency words to an equal extent. An analysis of the relation between reading ability and automaticity was only modestly successful in establishing the diagnostic value of indicators of automatic processing. Nevertheless, the pattern of results was reasonably consistent with the automaticity theory of LaBerge and Samuels (1974). This is an important outcome since, as noted in the introduction, there have been remarkably few empirical tests of the theory's developmental predictions that have been unambiguously supportive.

An important implication of the research reported here is that word recognition speed continues to increase even after recognition has become automatized. Thus, it is crucial that the distinction between speed and automaticity be made by reading theorists. The results are consistent with the conclusions of Ehri and Wilce (1979), who argued that beginners need only a moderate amount of practice before recognition becomes automatized. All of the first-grade readers in Experiment 2 had automatized recognition to a certain extent, but only the skilled readers recognized words rapidly. Thus, the results lend support to limited-capacity models (e.g., Perfetti & Lesgold, 1977) that argue that slow word recognition strains short-term memory and impairs reading, regardless of whether or not the word was recognized automatically.

Finally, since first grade appears to be the critical period for the development of automaticity, it would be of interest to relate this processing characteristic to the three first grade reading stages defined by Biemiller (1970) in his study of oral reading errors.

## REFERENCE NOTE

1. SCHADLER, M., & THISSEN, D. *The development of automatic word recognition and reading skills: A longitudinal analysis.* Paper presented at the annual meeting of the Midwestern Psychological Association, St. Louis, May 1980.

## REFERENCES

BARRON, R. W., & BARON, J. How children get meaning from printed words. *Child Development*, 1977, *48*, 587–594.

BIEMILLER, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, 1970, *6*, 75–96.

BIEMILLER, A. Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. *Reading Research Quarterly*, 1977–1978, *13*, 223–253.

CARROLL, J. B., DAVIS, P., & RICHMAN, B. *Word frequency book*. New York: American Heritage, 1971.

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.

GOLINKOFF, R. M., & ROSINSKI, R. R. Decoding, semantic processing, and reading comprehension skill. *Child Development*, 1976, *47*, 252–258.

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.

GUTTENTAG, R. E., & HAITH, M. M. A longitudinal study of word processing by first-grade children. *Journal of Educational Psychology*, 1980, *72*, 701–705.

JASTAK, J. F., BIJOU, S. W., & JASTAK, S. R. *Wide range achievement test*. Wilmington, DE: Guidance Associates, 1965.

KUCERA, H., & FRANCIS, W. M. *Computer analysis of present-day American English*. Providence, RI: Brown University Press, 1967.

LABERGE, D. Attention and the measurement of perceptual learning. *Memory and Cognition*, 1973, *1*, 268–276.

LABERGE, D., & SAMUELS, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

MARTIN, M. Speech recoding in silent reading. *Memory and Cognition*, 1978, *6*, 108–114.

MCCORMICK, C., & SAMUELS, S. J. Word recognition by second graders: The unit of perception and interrelationsips among accuracy, latency, and comprehension. *Journal of Reading Behavior*, 1979, *11*, 107–118.

PACE, 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.

Perfetti & Lesgold, 1977) that argue that slow word recognition strains short-term memory and impairs reading, regardless of whether or not the word was recognized automatically.

Finally, since first grade appears to be the critical period for the development of automaticity, it would be of interest to relate this processing characteristic to the three first grade reading stages defined by Biemiller (1970) in his study of oral reading errors.

## REFERENCE NOTE

1. SCHADLER, M., & THISSEN, D. *The development of automatic word recognition and reading skills: A longitudinal analysis.* Paper presented at the annual meeting of the Midwestern Psychological Association, St. Louis, May 1980.

## REFERENCES

BARRON, R. W., & BARON, J. How children get meaning from printed words. *Child Development*, 1977, *48*, 587–594.

BIEMILLER, A. The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly*, 1970, *6*, 75–96.

BIEMILLER, A. Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. *Reading Research Quarterly*, 1977–1978, *13*, 223–253.

CARROLL, J. B., DAVIS, P., & RICHMAN, B. *Word frequency book*. New York: American Heritage, 1971.

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.

GOLINKOFF, R. M., & ROSINSKI, R. R. Decoding, semantic processing, and reading comprehension skill. *Child Development*, 1976, *47*, 252–258.

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.

GUTTENTAG, R. E., & HAITH, M. M. A longitudinal study of word processing by first-grade children. *Journal of Educational Psychology*, 1980, *72*, 701–705.

JASTAK, J. F., BIJOU, S. W., & JASTAK, S. R. *Wide range achievement test*. Wilmington, DE: Guidance Associates, 1965.

KUCERA, H., & FRANCIS, W. M. *Computer analysis of present-day American English*. Providence, RI: Brown University Press, 1967.

LABERGE, D. Attention and the measurement of perceptual learning. *Memory and Cognition*, 1973, *1*, 268–276.

LABERGE, D., & SAMUELS, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293–323.

MARTIN, M. Speech recoding in silent reading. *Memory and Cognition*, 1978, *6*, 108–114.

MCCORMICK, C., & SAMUELS, S. J. Word recognition by second graders: The unit of perception and interrelationsips among accuracy, latency, and comprehension. *Journal of Reading Behavior*, 1979, *11*, 107–118.

PACE, 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.

PERFETTI, C. A., FINGER, E., & HOGABOAM, T. Sources of vocalization latency differences between skilled and less skilled young readers. *Journal of Educational Psychology*, 1978, *70*, 730–739.

PERFETTI, C. A., & HOGABOAM, T. Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, 1975, *67*, 461–469.

PERFETTI, C. A., & LESGOLD, A. M. Discourse comprehension and sources of individual differences. In M. Just, & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1977.

POSNANSKY, C. J., & RAYNER, K. Visual-feature and response components in a picture-word interference task with beginning and skilled readers. *Journal of Experimental Child Psychology*, 1977, *24*, 440–460.

POSNER, M. I., & SNYDER, C. R. R. Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1975.

PROCTOR, R. W. Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, 1978, *23*, 413–419.

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. H. Developmental study of color-word interference. *Journal of Experimental Psychology*, 1966, 72, 105–108.

SHANKWEILER, D., & LIBERMAN, I. Y. Misreading: A search for causes. In J. Kavanagh, & I. Mattingly (Eds.), *Language by ear and eye*. Cambridge: MIT Press, 1972.

STANOVICH, K. E. Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 1980, *16*, 32–71.

STANOVICH, K. E. Relationships between word decoding speed, general name-retrieval ability, and reading progress in first-grade children. *Journal of Educational Psychology*, 1981, in press.

STROOP, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, *18*, 643–661.

WEST, R. F., & STANOVICH, K. E. Automatic contextual facilitation in readers of three ages. *Child Development*, 1978, *49*, 717–727.

WEST, R. F., & STANOVICH, K. E. The development of automatic word recognition skills. *Journal of Reading Behavior*, 1979, 11, 211–219.