Memory & Cognition
1981, Vol. 9 (4), 378-388

The development of letter processing efficiency

DANIEL B. KAYE
Yale University, New Haven, Connecticut 06520

SCOTT W. BROWN
University of Connecticut, Storrs, Connecticut 06268

TIM A. POST
University of Pittsburgh, Pittsburgh, Pennsylvania 15260

and

DANA J. PLUDE
Veterans Administration Outpatient Clinic, Boston, Massachusetts 02108

The development of efficiency in letter processing skills was studied using a letter search task. In two experiments, subjects searched for a target letter displayed with items varying in their visual featural or conceptual categorical similarity to the target. Accuracy and reaction time of search were evaluated for evidence of the visual search "category effect." In order to determine if subjects could efficiently use knowledge of stimulus differences to facilitate search, conditions tested search time as a function of the amount of information to be processed both within the visual display and in short-term memory. In the two experiments, subjects of ages 6 years through adulthood showed the category effect; however, efficiency of letter processing was found to be related to the amount of information that had to be processed in memory. While there were drastic changes in search speed with increasing grade level, patterns of processing were consistent, leading to the conclusion that the knowledge required to process the letter information accurately is acquired very early. Results were discussed in terms of the distinctions among accuracy, automaticity, and efficiency of skill development and the relationship of these to general reading and intellectual development.

Young children must be able to identify and discriminate among letters rapidly and accurately in order to read fluently (Gibson, 1969). While theories differ about the relationship between letter processing and higher level processing, most acknowledge the primacy of letter skill acquisition.

In order to study letter processing skills, variants of the visual (letter) search task (Neisser, 1963) have proved useful. If subjects have acquired knowledge about differences between letters and other visual stimuli, then they should be able to detect more accurately and rapidly a target letter presented with visually dissimilar letters (or other visual forms) than one presented with visually similar letters. When these stimuli form perceptual or conceptual categories (e.g., letters vs. digits, straight-line letters vs. curved letters), these subjects should show the "category effect." This effect refers to the facilitation on reaction time (RT) of searching for a letter in a field of dissimilar items (Egeth, Jonides, & Wall, 1972).

The ability to use featural or categorical information allows the subject to detect a target stimulus accurately without complete identification, thereby facilitating RT (Gleitman & Jonides, 1976; Jonides & Gleitman, 1976). If the category effect is found, the usual implication is that subjects have used the featural differences among the stimuli to form conceptual distinctions that aid in detection. If a letter is searched for in a field of digits, the subject need only determine that the target is a letter (rather than the specific letter) to signal target presence or absence. Thus, by varying the degree of similarity between target and distractor stimuli, the

This research was funded in part by the Syracuse University Research and Equipment Fund Grant RE-78-B17, awarded to the first author, and by NIMH Developmental Psychology Training Grant 5 T32 MH 16127, awarded to Yale University. Portions of this paper were presented at the meetings of the American Psychological Association, New York, 1979, and the Eastern Psychological Association, Hartford, Connecticut, 1980. We extend our appreciation to William DeLucia, Don Stebbins, the staff, and the children of Stonehedge Elementary School, Camillus, New York, without whose cooperation this study would not have been possible. Helpful comments on earlier drafts from Janet Caplan, Cathryn Downing, Bob Sternberg, and Lazar Stankov are greatly appreciated. Very special thanks are extended to Tex Garner, whose scrutiny of the research and critical comments vastly improved the final product, although any responsibility for problems with the manuscript lies solely with the authors. Reprint requests may be sent to Daniel B. Kaye, who is now at the Department of Psychology, University of California, 405 Hilgard Avenue, Los Angeles, California 90024.

  0090-502X/81/040378-11$01.35/0

category effect can indicate the extent to which a subject has acquired the knowledge of featural distinctions among letters or other visual stimuli.

That this knowledge is important in reading-fluency development is emphasized by at least one popular hierarchical model of reading (LaBerge & Samuels, 1974). In this model, when letters are processed automatically (without the allocation of conscious attention), more attentional capacity is available for deployment to higher level processing, such as comprehension, than when letter processing requires attention. Stanovich and West (1978) suggested that the category effect may be a by-product of automatic letter processing, since the effect is shown in the absence of conscious scanning strategies. Thus, one might expect the category effect to be found as readers progress beyond the level of letter processing to more fluent stages of reading. This is what guided the work of Stanovich and West, who found the category effect to be present and relatively invariant from 8 years of age until adulthood.

While the category effect may be mediated by an automatic process, it may also be shown prior to the development of automaticity. Furthermore, efficiency of letter processing skills may develop well beyond this automaticity development. As Ehri and Wilce (1979) have conceptualized practice on any task component, subjects first master a task in terms of accuracy and then develop automaticity and efficiency (accurate or effective speed) in sequence. In order to identify a target letter, a subject would first need to have mastered letter processing at the accuracy level. Subsequently, the category effect could be shown, with or without automatic processing. Once automaticity has been achieved, speed in accurate letter processing would continue to develop.

In studies that have investigated automaticity of letters and words with direct measures such as a Stroop-like interference task or a catch-trial procedure, the time course of automaticity development has been rather well-defined. Automatic letter processing develops sometime in the first grade for most children (Guttentag & Haith, 1978, 1979; Stanovich, in press). Thus, the category effect would be expected by the end of first grade at the latest. Beyond that time, development of efficiency of letter processing would be expected to develop and to be related to increased fluency development (cf. Doehring, 1976).

Since the category effect is not a direct test of automaticity of letter processing, the goal of a developmental investigation of the effect is to trace the nature and time course of the efficiency of letter processing. While it is crucial that children acquire knowledge about the distinctions among letters and apply this knowledge accurately and automatically, it is also important that they be able to do this with efficiency (i.e., by accurately processing large amounts of information in a short time span). The decrease in RT with age is a common phenomenon in many information processing tasks (Bisnanz, Danner, & Resnick, 1979; Keating, Keniston, Manis, & Bobbitt, 1980). As yet, we have been unable to explain such changes in theoretically plausible ways. By manipulating the type and amount of processing in the visual search task, we intend to learn more about the developing efficiency of letter processing skills. The results will complement those of studies dealing with automaticity development and provide reading researchers a better understanding of early letter processing skills.

EXPERIMENT 1

We have argued that a reasonable test of subjects' awareness of feature differences, and hence, letter identification and discrimination skills, can be constructed with the visual (letter) search task. Thus, we were interested in studying the category effect and related developments of letter processing efficiency from the earliest stages of reading skill. In this first experiment, we had subjects search for the uppercase letter "V," which was displayed along with items varying in their similarity to the target. Distractors were letters that shared the diagonal straight-line feature of the "V," letters that shared the acoustic (rhyme) feature, or digits or Greek letters. To the extent that subjects process the letters on the basis of shared features (with the target) that are well learned, RT should differ across these sets of distractors. It is assumed that the visual distractors should provide the most difficult task for subjects sensitive to letter features. The acoustic set distractors were chosen as a conceptually similar (letters) but perceptually (visual) distinct category of items. Digits have been used often as distractors when the category effect has been found, and they were chosen to be categorically distinct but perhaps perceptually confusable. The Greek letters were conceptually distinct but less well-known, and they were perceptually confusable, since they shared some features with letters (although not necessarily with the target "V").

In order to reject a nontarget item in a display, only partial recognition is required. As soon as a subject detects a feature that does not belong to the target, processing of that item can be terminated. This assumes memory of the target item and efficient use of featural differences between target and distractor stimuli. Thus, an efficient short-term or working memory for the target and relevant and irrelevant features should be operative for successful task performance. If only a single item is displayed on a given trial, the subject's task should be relatively easy, since he (she) need only process a single item, compare it with the target held in memory, and then decide and respond. A more complex display, say, four items, should present a more difficult task. At least, RT for a response would be longer, since there are more items to process and compare. However, it is possible that trials would be processed differently in the two display size conditions. With only one item displayed, the featural or categorical differences between the target and distractor items may not be processed efficiently. These differences cannot be presented visually simultaneously. If more than one instance of a class of stimuli must be processed before category membership will be used to facilitate search, then subjects who do not use the information from short-term memory efficiently would not be expected to show the category effect with a single item displayed. With four items displayed, however, the subject has direct visual access to more of the competing featural or categorical information and needs to depend less on the memory system. On positive trials, the target plus three distractors can be compared visually. On negative trials, the full set of distractors (for these experiments) can be processed visually, and the subject need only compare them with the target item held in memory. While the larger display will result in longer search times since more items need to be processed, it is conceivable that the additional information provided visually will aid the subject in inferring category membership, which in turn can be used to facilitate search. Hence, the category effect would be shown, because processing would then be free of the memory limitation.

Since we were interested in examining the category effect at different levels of experience with letters and letter features, we used children in kindergarten (Grade K) through Grade 3 and college adults as subjects. To investigate whether the efficiency of processing such a low-level perceptual task is related to more general reading skill, individual reading achievement test data were also accumulated for the children. As in other recent attempts to relate information processing abilities to psychometric task performance (cf. Hunt, 1978), general measures of intellectual ability were included. The Peabody Picture Vocabulary Test (Dunn, 1965) was selected as a measure of verbal intelligence because it provides a measure of semantic knowledge that is not confounded by the decoding requirements of word vocabulary tests. Since it is hypothesized that fluent readers (who will have acquired extensive vocabulary skills) have already mastered letter identification processes, we predict that Peabody scores will be related to our measures of letter processing skills (i.e., RT in visual search and evidence of the category effect). The Raven Coloured Progressive Matrices (Raven, 1965) is a test of nonverbal reasoning that ostensibly taps short-term attentional and memorial abilities required in a wide range of problem solving tasks. Since such component abilities are hypothesized to be operative in reading and visual search, the Raven was expected to be related to RT in this task.

Method

Subjects. Eight males and eight females were selected randomly from each of Grades K through 3 of a suburban middle-class elementary school (mean ages in months: Grade K = 73.25, Grade 1 = 85.50, Grade 2 = 95.94, Grade 3 = 109.38; standard deviations in months: Grade K = 4.22, Grade 1 = 4.73, Grade 2 = 2.26, Grade 3 = 3.98). An additional eight males and eight females were selected from the introductory psychology pool at Syracuse University. These subjects received course credit for their participation.

Apparatus. A Polymorphic 8813 microcomputer system was used for all aspects of stimulus presentation and recording of responses. Characters were presented on the cathode-ray tube (CRT) monitor of the system. Each character filled a 5 by 7 font, measuring 3 x 5 mm and subtending a visual angle of less than 1 deg. Four display locations were used (the corners of an imaginary rectangle around the fixation point), subtending a visual angle of less than 3 deg from the central fixation point. Responses were recorded through the keyboard, with RT recorded to the nearest 5 microsec (Post, 1979).

Procedure. Subjects in Grades K through 3 each received the Raven and Peabody tests followed by the letter search task. Subjects were individually tested; however, prior to the running of the actual experiment, they were given a demonstration of the microcomputer system in small groups. This warm-up session was successful in orienting the children to the computer, College students were administered only the RT task.

The reading subtests of the Stanford Achievement Tests (SAT) (Kelley, Madden, Gardner, & Rudman, 1968) were administered in the same month as the experimental tasks. Scaled scores were used for the purposes of intergrade comparisons. For kindergartners, the auditory and visual-motor subtests of the Rosner (Note 1) reading readiness battery were available. Means and standard deviations were well within the average range of performance for such samples. For the Raven, means (and standard deviations) for Grades K through 3 were 15.73 (3.33), 19.31, (5.03), 20.47 (3.20), and 25.56 (4.08), respectively; for the Peabody, they were 62.47 (8.09), 64.31 (5.93), 71.47 (7.86), and 79.69 (9.00). Kindergartners averaged 9.60 and 16.80 (SDs = 3.85 and 5.05) on the fall version of the Rosner visual-motor and auditory tests, respectively; on the spring version, the means (and standard deviations) were 15.73 (3.77) and 25.07 (2.92). Finally, on the SAT reading, for Grades 1, 2, and 3, respectively, means (and standard deviations) were 128.75 (18.60), 136.93 (8.31), and 156.00 (11.93).

The RT task comprised two practice trial blocks and 16 experimental trial blocks, as described below. Subjects were seated in front of the CRT such that their eyes were on a level with the middle of the screen and approximately 60 cm from the screen. Although most subjects understood the task from the warm-up session, the instructions were repeated. Subjects were told to look at the middle of the screen, where a dot (fixation point) would appear, followed by a display of characters. They were to search for the uppercase letter "V" and, on each display, to respond as to whether the target letter was present ("yes" response) or absent ("no" response) by striking one of two predesignated keys on the board. The keys were selected so that even the smallest child could rest his or her forefingers on each key and therefore visually attend only to the screen. Response was then a simple matter of pressing the key. At no time during trials did the subject need to remove the fingers from a resting position on the keys. (Two kindergarten subjects were eliminated for failure to understand the instructions or due to the physical limitation imposed by fingers so small that they fell between the keys.)

Only one target letter was selected so as to enable selection of appropriate distractor categories: An English letter was desired that had at least four other letters sharing visual features and four that shared acoustic features. As mentioned, the visual feature was the diagonal straight line shared by the target with W, X, Y, and M (Gibson, 1969). The rhyming letters were B, C, P, and T. In pilot testing, "F" was used as an alternative target (after Conrad, 1972) with visual distractors having horizontal and vertical straight lines (E, H, I, T) and acoustic distractors

S, N, M, and X. Subjects were not similarly confused by these acoustic distractors and reported they did not "hear" them rhyme as well as in the "V" condition. Thus, we decided to keep only one target and one set each of visual and acoustic distractors. Digit distractors used were 4, 6, 8, and 9; Greek letters were λ, φ, δ, and ζ. These digits and lowercase Greek letters were the most easily discriminable with the computer's character set and were not easily mistaken for English letters.

The session began with two practice trial blocks. Each block consisted of 16 trials with the target letter "V" and randomly selected English letter distractors. One block used Display Size 1: Either the target or a distractor was presented in one of four screen locations. One half of the trials contained the target; the other half contained a distractor. The other practice block displayed four characters (target and three distractors, or four distractors). Although the experimenter was free to administer more practice if it was felt the subject was not prepared to begin the experiment, this prerogative was exercised in only three cases.

The 16 experimental trial blocks were presented next. The session was divided into two halves: Each condition was presented in both halves to measure effects of practice. Each of the four distractor sets was presented at each display size, for a total of eight trial blocks in each half of the session. Blocks alternated between display sizes, and order of distractor set was balanced across subjects within grade and gender.

Each block consisted of 17 trials; the data were not recorded for the first of these trials in each block. The next 16 were balanced for location of the target and for target presence/absence. On positive trials (target present), the three distractors were selected randomly from the set of four, with location of each also randomized. On negative trials, the same distractors appeared on each trial, with location of each randomized. Beginning display size was counterbalanced across subjects within grade and gender.

For each trial, a fixation point appeared in the center of the screen. After a 500-msec interval, the display was presented. Subjects were required to respond with the right forefinger for target presence and the left forefinger for target absence. No attempt was made to counterbalance for handedness of response.[1] The response was to be made with as much speed and accuracy as possible.

Following each trial block, a 30-sec rest period intervened. Sessions lasted approximately 25 min. Although a total of 304 trials (including practice and warm-up) was administered, the frequent rest periods allowed for reliable testing of even the youngest subjects.

Results

Analyses. Due to the extreme differences in mean level of RT between display sizes and among age groups, all analyses were computed within grade and display size. This analysis strategy avoids the metric problem that would arise if grade or display size were analyzed as independent variables in an analysis of variance design and if interactions of these and other factors were interpreted. Nevertheless, important condition differences and developmental performance patterns were still apparent when the data were analyzed as described.

Two types of analyses are reported. First, analyses of variance of the search task itself are reported, with either number correct or mean median RT in milliseconds (for the correct trials of each block of trials) as the dependent measure. The final section presents the results of correlational analyses testing the hypothesized relationships between the search task and psychometric measures.

Before considering the main findings, let us examine two preliminary sets of analyses. First, gender effects are not henceforth discussed, since few were significant and, when there were significant main effects or interactions, there were no consistent patterns across grade, display size, or response type.

There was no evidence of speed-accuracy tradeoffs of the type described by Pachella (1974). In conditions that resulted in longer RTs (Display Size 4), subjects were not more accurate. On the contrary, in the only two grades in which accuracy was significantly different between the two display sizes, Grades K and 3, subjects were more accurate for the Display Size 1 condition, for which they were also faster (mean RT and proportion correct: Grade K, Size 1 = 929.52 and .933, Size 4 = 1,341.04 and .920; Grade 3, Size 1 = 716.44 and .973, Size 4 = 871.98 and .958).

A second type of speed-accuracy tradeoff involves that across subjects rather than conditions. Correlations between RT and accuracy computed over subjects are evaluated to determine whether subjects who take longer to respond are also more accurate. Positive evidence was limited to the correlations for college subjects: When collapsed across all task factors, $r(14) = -.06$, $-.13$, $-.14$, and .16 for Grades K, 1, 2, and 3, respectively (all $ps > .05$); for college, $r(14) = .55$ ($p < .03$). Those college students who took longer to search for the target letter were also more accurate.

Search task analyses. The main RT results are depicted in Figure 1. Although the speed of letter search increased dramatically from kindergarten to college, there were important, reliable consistencies across grade levels. Most obvious was the nearly constant rank ordering of distractor set difficulty at the display size of four items. In every grade, there were significant differences in RT as a function of distractor set [$Fs(3,45) = 3.77$, 10.37, 10.94, 3.95, and 20.82 for Grades K, 1, 2, 3, and college, respectively; $ps < .02$, .0001, .0001, .02, and .0001, respectively]. There was only one exception to the order of decreasing RT of visual followed by Greek, acoustic, and digit distractor sets (binomial $p < .0001$). In every grade, visually confusable distractors led to the longest RTs and digits led to the shortest (binomial $p < .0001$). In contrast, when only a single item was displayed, only the college subjects could reliably search for the target with differential speed when it appeared with the different distractors (although the rank ordering was similar to that of Display Size 4 for Grades 2 and 3, as well). Reliable differences were found for Display Size 1 only for the college students [$F(3,45) = 3.39$, $p < .03$], with visual and acoustic set times each longer than digit set time (Fisher l.s.d. test, $\alpha = .05$). At Display Size 4, in each grade, visual times were

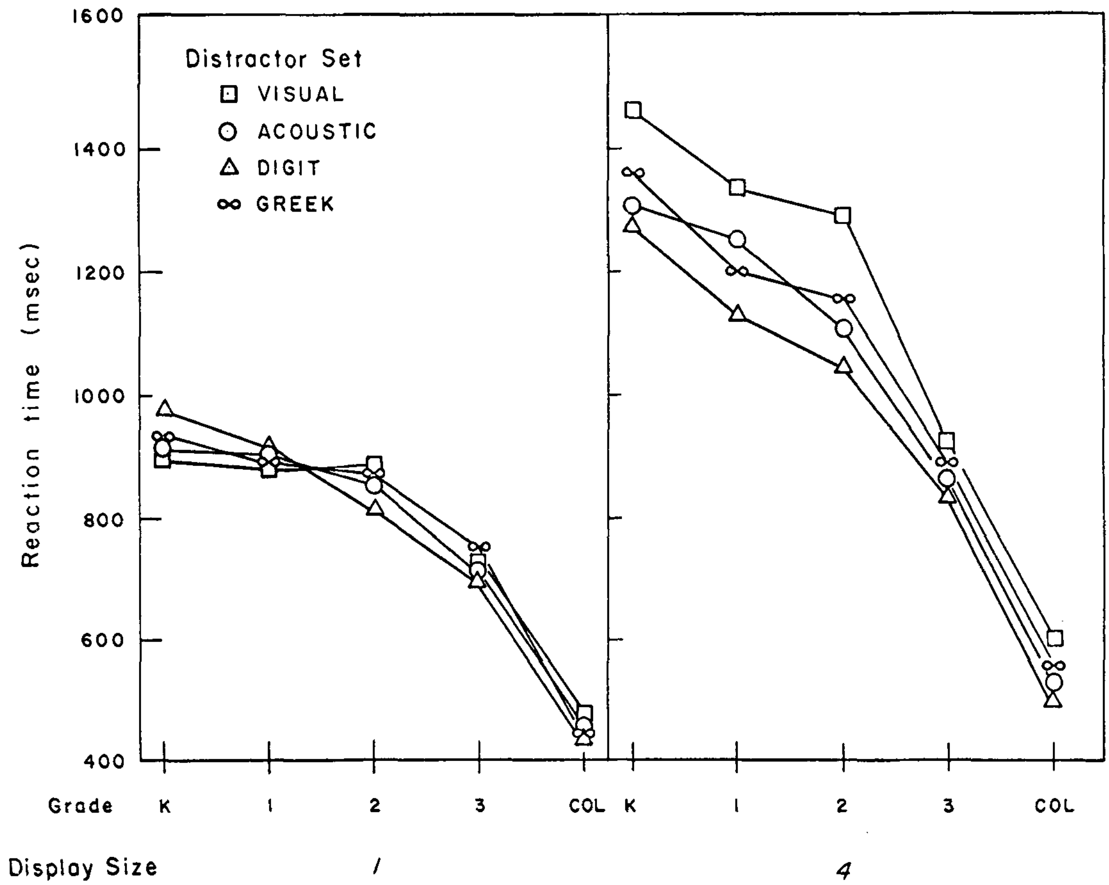

Figure 1. Mean median reaction time as a function of distractor set for each grade and display size (Experiment 1).

significantly longer than digit times. In Grades K, 2, 3, and college, visual times were longer than acoustic times, and in Grades 1, 2, and college, visual times were longer than Greek times. Acoustic times were longer than digit times in Grades 1 and college, and Greek times were longer than digit times in Grades 2 and college.

The task factors not included in Figure 1 also exerted consistent effects. Subjects in all grades showed practice effects. From the first to the second presentation of each condition, RTs decreased. As usual, "no" trials took longer than "yes" trials.

Correlational results. Our examination of within-grade correlations between RT and psychometric measures is not presented as a general test of the correlations among these measures. Of course, these RT and psychometric measures are related. Speed increased with age, as did psychometric task performance. The correlations for the entire sample of children (Grades K-3) support this point [for the Raven and RT, $r(62) = -.41$, $p < .001$; for the Peabody and RT, $r(62) = -.44$, $p < .0002$; and for reading achievement and RT, $r(46) = -.47$, $p < .0007$].

What we were examining with the correlational analyses separated by grade was whether any additional patterns of importance would emerge. While there were several significant correlations between RT and both the reading achievement and Peabody scores (but none with the Raven scores), no notable patterns of correlations emerged. This does not lend strong support to the hypothesis that general intellectual ability is related to processes reflecting the developing efficiency of letter processing skills when examined at each developmental level.

Discussion

The results of this first experiment were clear. From 6 years of age through young adulthood, subjects were able to capitalize upon perceptual or conceptual differences among alphanumeric stimuli to facilitate visual search performance. The findings of previous studies (Lefton & Fisher, 1976; Stanovich & West, 1978) have been extended by showing that the category effect is produced by children as young as 6 years of age. The patterns of the category effect suggest strongly that the acquisition of featural knowledge relevant to the present stimuli is accomplished rather early and before subjects have experienced much formal instruction with such stimuli. While it is remotely conceivable that such knowledge is acquired very early in life or is innate in some sense, we think that Gibson's (1969) position on perceptual learning of such discriminations is more reasonable. While some feature differences have been found to be ontogenetically primitive (Piaget & Inhelder, 1956), those most relevant to letter processing are detected later, beginning at approximately 4 years of age (Gibson, 1969).

However, before making too strong a case for the early acquisition and refinement of letter identification and discrimination abilities, let us examine the data more carefully. First, we have accepted a negative finding of no developmental changes in our conclusions. We did not analyze the Age by Category Effect interaction, which would certainly have been significant, and we argued that such a test would have capitalized on gross metric differences across age and display size. But RT drastically decreased with age, as it does in most RT studies. Are we justified in ignoring this developmental change and accepting instead the similarity in patterns across age and condition? One source for caution comes from the findings with the small display size. While subjects at all grade levels experienced differential difficulty with the various distractor sets with four items displayed, only the college students did so reliably with a single item displayed. So, the lack of a category effect with a display size of one was not the result of a floor effect, since college subjects' RTs were much shorter. As mentioned in the introduction, in this condition, subjects would have to depend more heavily on short-term memory, or perhaps even a strategy change. It is possible that our task failed to impose enough short-term memory load to detect developmental differences in the larger display condition in the extent to which letter processing has become efficient or even automatic.

The nature of the search task is such that letter processing may operate relatively independently of other processes necessary in actual reading situations, whether they be such things as use of orthographic knowledge or use of higher order syntactic or semantic knowledge. Since there was evidence that efficiency in letter processing was facilitated by providing competing featural information or by presenting the entire set of distractors in a single visual display, rather than having the subject depend on short-term memory, it is reasonable that memory load factors might provide a process limitation on the efficiency of letter processing skills. This possibility was pursued in Experiment 2.

EXPERIMENT 2

Thus, further manipulations were included in this second experiment to determine if children and adults would exhibit similar category effects when processing the stimuli under greater amounts of cognitive load. Furthermore, several procedural modifications characterized Experiment 2: (1) Whereas the first experiment used only a single target letter ("V") that may have had special characteristics not common to other letters, multiple targets were used in this experiment by sampling the required number of targets in each condition from a fixed set of letters. (2) Distractor sets consisted of visually confusable letters, acoustically confusable letters, or digits. (The Greek letters were omitted.) (3) Distractor sets were varied between subjects to avoid possible contamination effects from within-subjects strategy changes and to allow for the first manipulation above. (4) Display sizes of one, two, and four were used for a more complete analysis of this factor. (5) A memory set of targets to be searched of one, two, or three items was included to study the effects of carrying a greater memory load. While other techniques have been used to investigate automatic processing, such as requiring subjects to remember irrelevant information or interpolating another memory task (cf. Logan, 1979), we used this memory set factor to study efficiency of letter processing. (6) Subjects from Grades 1, 2, and 3 were tested approximately 5 months earlier than their grade counterparts in Experiment 1. When considered together with the results of the previous experiment, a rather complete developmental picture from ages 6-9 years should thus emerge. (7) The digit span forward subtest of the Wechsler Intelligence Scale for Children (Wechsler, 1974) was included to examine relationships between memory span and letter search measures.

Method

Subjects. Eighteen males and 18 females were selected from each of Grades 1, 2, and 3 of the same school used in the first experiment (mean ages in months: Grade 1 = 80.76, Grade 2 = 90.83, Grade 3 = 103.16; standard deviations in months: Grade 1 = 5.12, Grade 2 = 3.47, Grade 3 = 5.40). An additional 18 males and 18 females were selected from the undergraduate population of Syracuse University.

Apparatus. The apparatus used was the same as that in Experiment 1.

Procedure. Subjects in Grades 1, 2, and 3 each received the Peabody Picture Vocabulary Test, the Raven Coloured Progressive Matrices Test, and the digit span forward followed by administration of the letter search task, in two separate sessions. Subjects were again tested individually following group warm-up sessions. College subjects received only the RT task.

Due to circumstances beyond our control, the experimental data were gathered 6 months prior to the collection of the achievement data. Hence, we will be examining predictive rather than concurrent validity of these data. The SAT was available for the children in Grades 1, 2, and 3. Again, the scores on these measures were typical. For Grades 1, 2, and 3, respectively, the means (and standard deviations) on the Peabody were 62.83 (5.77), 68.56 (7.38), and 76.78 (7.63); for the Raven, they were 17.94 (3.75), 21.06 (5.00), and 21.70 (5.52); for the digit span, scores were 4.64 (.94), 4.75 (.91), and 5.08 (.89); for the reading total score, they were 131.21 (20.48), 145.47 (15.15), and 155.64 (13.70); and for the math total score, they were 125.52 (9.04), 136.64 (9.90), and 153.94 (13.38).

The search task comprised nine practice blocks of 5 trials each and nine experimental blocks of 21 trials each. Each subject was presented one of three sets of target and distractors (visually confusable letters, acoustically confusable letters, or digits). The visual set contained the uppercase letters A, M, V, W, X, Y, and Z; the acoustic set contained B, C, D, E, G, P, and T; and the digit set contained all single-digit numerals except 1, 2, and 6. For each subject in each of the two letter distractor conditions, three targets were selected randomly from the total set, with the remainder used as distractors. In the digit condition, half the subjects received visually confusable targets and half received acoustically confusable targets. Four of the seven digits were selected randomly on each trial to be used as distractors, and location of all stimuli on each trial was randomized.

The number of targets to be searched, or memory set size, was always presented in an ascending sequence; thus, the first three experimental blocks were for Memory Set Size 1, the next three for Memory Set Size 2, and the last three for Memory Set Size 3. Within each memory set size, a different display size (one, two, or four) was used for each block of trials. The order of display sizes within memory set size was varied across memory set sizes within and between subjects. For Display Size 2, only adjacent corners of the imaginary rectangle described for Experiment 1 were used. Each of the six possible orders of display size was used an equal number of times. Additionally, the change in memory set size from one to two to three was accomplished by adding an additional member of the original target set of three to the target selected initially on a random basis. Thus, if the target set was B, C, D, the first memory set might have been C, the second B and C, and the third B, C, and D. This procedure was used to guard against the possibility that subjects might be operating on a letter from a previous but not current memory set. (However, specific effects of practice with a certain letter across memory sets were not assessed.) Finally, half of the trials in each block contained a single target, and half contained no target at all.

Prior to each set of three blocks at each memory set size, a set of three "miniblocks" of five trials of practice for each display size was administered. These were given to ensure that subjects had memorized the targets, understood the entire procedure, and were capable of performing the task. Practice was repeated if necessary. During the experimental session, subjects could retrieve the memory set by pressing a predesignated key. However, subjects showed the need to do so rarely. Other aspects of procedure were the same as in Experiment 1.

In summary, each subject was given five practice trials for each memory set size/display size combination, for a minimum total of 45 trials. The experimental blocks consisted of 21 trials (1 warm-up) for each memory set size/display size combination, for a total of 180 trials for which data were recorded. Although this total of 234 trials is far below the number used in the first experiment, we believe it was the optimal number for a single session, given the difficulty level and time needed for completing the task (approximately 20 min). Subjects were allowed to rest whenever they needed; however, even most first-graders needed only short rest periods between trial blocks.

Results

Analyses. The analytical strategies used in Experiment 1 were employed here as well.

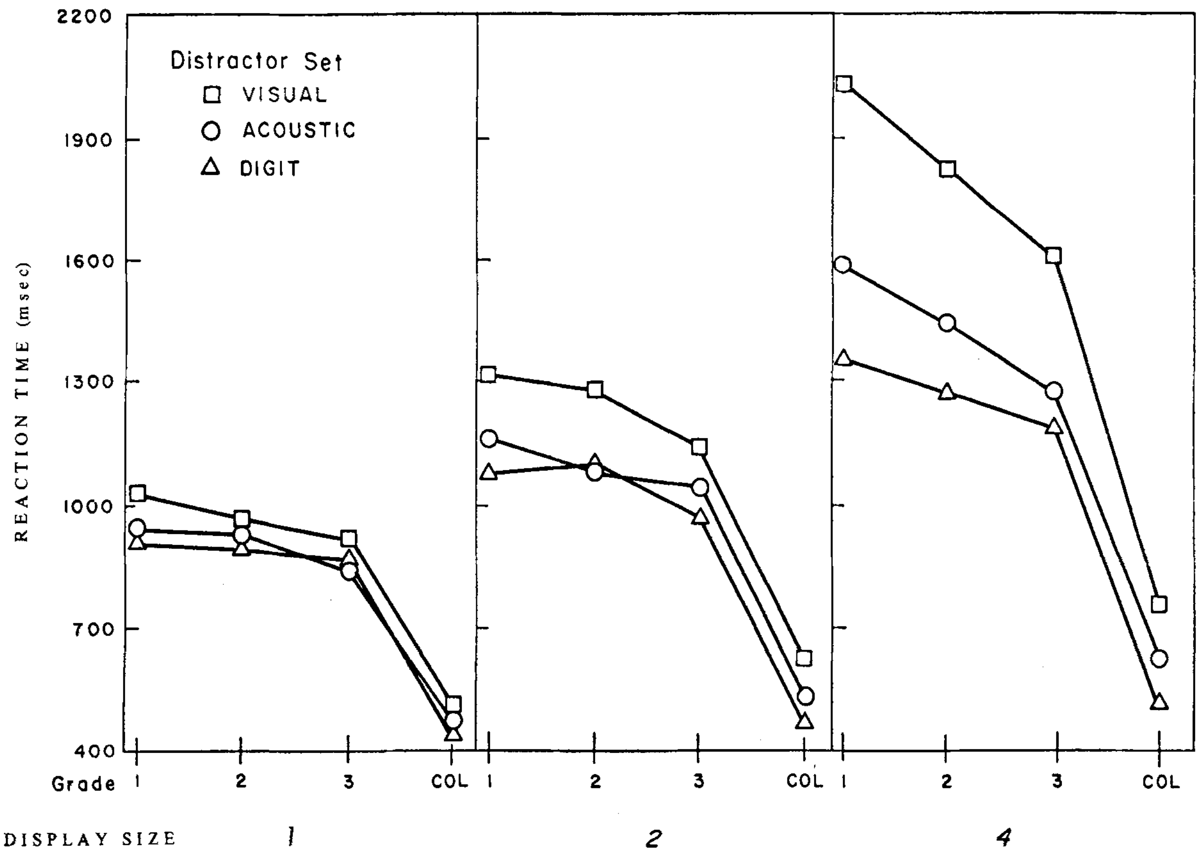

Figure 2. Mean median reaction time as a function of distractor set for each grade and display size collapsed across memory set size (Experiment 2).

As in Experiment 1, few gender effects were significant and there were no consistent patterns of these effects. Hence, gender was not considered in further analyses.

Evidence was lacking for speed-accuracy tradeoffs over conditions. While accuracy did change reliably from the Display Size 1 to the Display Size 4 conditions, it was a decrease in accuracy that was related to the increase in RT. Thus, the Display Size 4 conditions were more difficult, but subjects did not adopt a tradeoff strategy.

The speed-accuracy tradeoff for subjects found for college subjects in Experiment 1 was replicated [for Grades 1, 2, and 3, $r(34) = -.19$, $-.24$, and $-.28$, all $ps > .05$; for college, $r(34) = .39$, $p < .02$].

Search task. In order to evaluate the hypotheses generated from the results of the first experiment, consider first the results of Figure 2, which contains results analogous to those of Figure 1. These are the mean median RTs for each grade, display size, and distractor set, collapsed across memory set size. Although the ages are somewhat different, a comparison of the results for Grades 1, 2, 3, and college at Display Sizes 1 and 4 reveals that these conditions resulted in consistent levels and patterns of RT across very different experimental procedures when compared with the results of Experiment 1. While the highest level of RT in this experiment exceeds the level for the same condition in Experiment 1 by a considerable margin (Grade 1, visual set, Display Size 4), it should be noted that the differences were due to the memory set size effect. As we shall see, this factor exerted a potent influence.

Turning again to Figure 2, in all 12 possible comparisons of distractor sets, the visual distractors resulted in the longest RTs (binomial $p < .0001$). In 10 of the 12, digit times were shortest (binomial $p < .0001$), with acoustic distractors leading to time savings of 7 and 8 msec over digit distractors in the remaining 2 comparisons (both $ps > .05$ for Fisher's l.s.d. t tests). While there were no significant distractor set effects at Display Size 1, there were at Display Size 2 in Grades 1 and college [$Fs(2,33) = 4.52$ and 5.11, respectively; both $ps < .02$], with visual times in both grades longer than digit times (Fisher's l.s.d. test, both $ps < .01$). At Display Size 4, $Fs(2,33) = 6.50$, 11.79, and 5.19 ($ps < .005$, .0001, .002, and .02), for Grades 1, 2, 3, and college, respectively. The t tests confirmed that visual set times were longer than both acoustic and digit times in Grades 1, 2, and 3 and longer than only digit in college (all $ps < .01$).

The factor of memory set size was expected to add a short-term memory load in addition to that accomplished by display size. Subjects were given one, two, or three targets to remember, only one of which could be displayed on any one trial. Examination of Figure 3 reveals the results concerning this factor. It is evident that the distractor set effect was dependent on the memory set size. The greater the number of targets to be held in memory, the larger was the distractor set effect. While there were no significant effects for Memory Set Size 1, the distractor set effect was significant in all grades for Memory Set Sizes 2 and 3 [for Size 2, $Fs(2,33) = 6.76$, 5.20, 5.68, and 5.22 ($ps < .004$, .01, .008, and .01) for Grades 1, 2, 3, and college, respectively; for Size 3, $Fs(2,33) = 5.76$, 8.71, 5.57, and 5.43 ($ps < .008$, .0009, .008, and .009), respectively]. Fisher's l.s.d. t tests revealed that in every grade, at Memory Set Sizes 2 and 3, visual set RTs exceeded digit set times. Visual times were longer than acoustic times, as well, in Grades 1, 2, and 3 at Memory Set Size 3 (all $ps < .01$).

Response type, the lone remaining task factor, once again showed its main effect in an advantage in speed for "yes" trials over "no" trials.

Correlational analyses. Compared with Experiment 1, the psychometric measures and RT were not correlated as strongly for the entire sample of children [for Raven, $r(103) = -.10$, $p > .10$; for Peabody, $r(103) = -.14$, $p > .10$; for digit span, $r(103) = -.16$, $p > .10$; for reading achievement, $r(103) = -.20$, $p < .05$; and for math achievement, $r(103) = -.36$, $p < .0002$]. To some extent, lower correlations were to be expected, since the range in the measures was restricted in comparison with the sample from Experiment 1, which included kindergartners.

Correlations were computed between RT measures and the psychometric tasks within grade. While significance patterns were weak in Grades 1 and 2, there was one pattern of note in first grade. This concerned the relationship between the digit span and RT with digit distractors and a single item displayed. For Memory Set Sizes 1, 2, and 3, respectively, $r(10) = -.60$, $-.75$, and $-.77$ (all $ps < .01$). These correlations are consistent with the hypothesis that the absence of the category effect overall for children at this small display size was due to a working memory limitation. At least in the first-grade sample, children who were able to detect rapidly these stimulus differences had higher digit span scores and, presumably, better working memories.

In Grade 3, RT was correlated with the Raven scores [$r(34) = -.33$, $p < .03$] and the conceptual mathematics achievement test [$r(34) = -.53$, $p < .01$]. Breaking these down further, the conceptual math test scores were related to RT uniformly across most task conditions. Of note were the correlations for the display sizes of one, two, and four [$rs(34) = -.46$, $-.61$, and $-.40$ ($ps < .005$, .0001, and .02), respectively]. The Raven scores were not as strongly related generally to RT; however, the strongest correlations were for the visually confusable letters at all three display sizes with a single target to be searched [$rs(10) = -.54$, $-.64$, and $-.65$; all $ps < .05$].

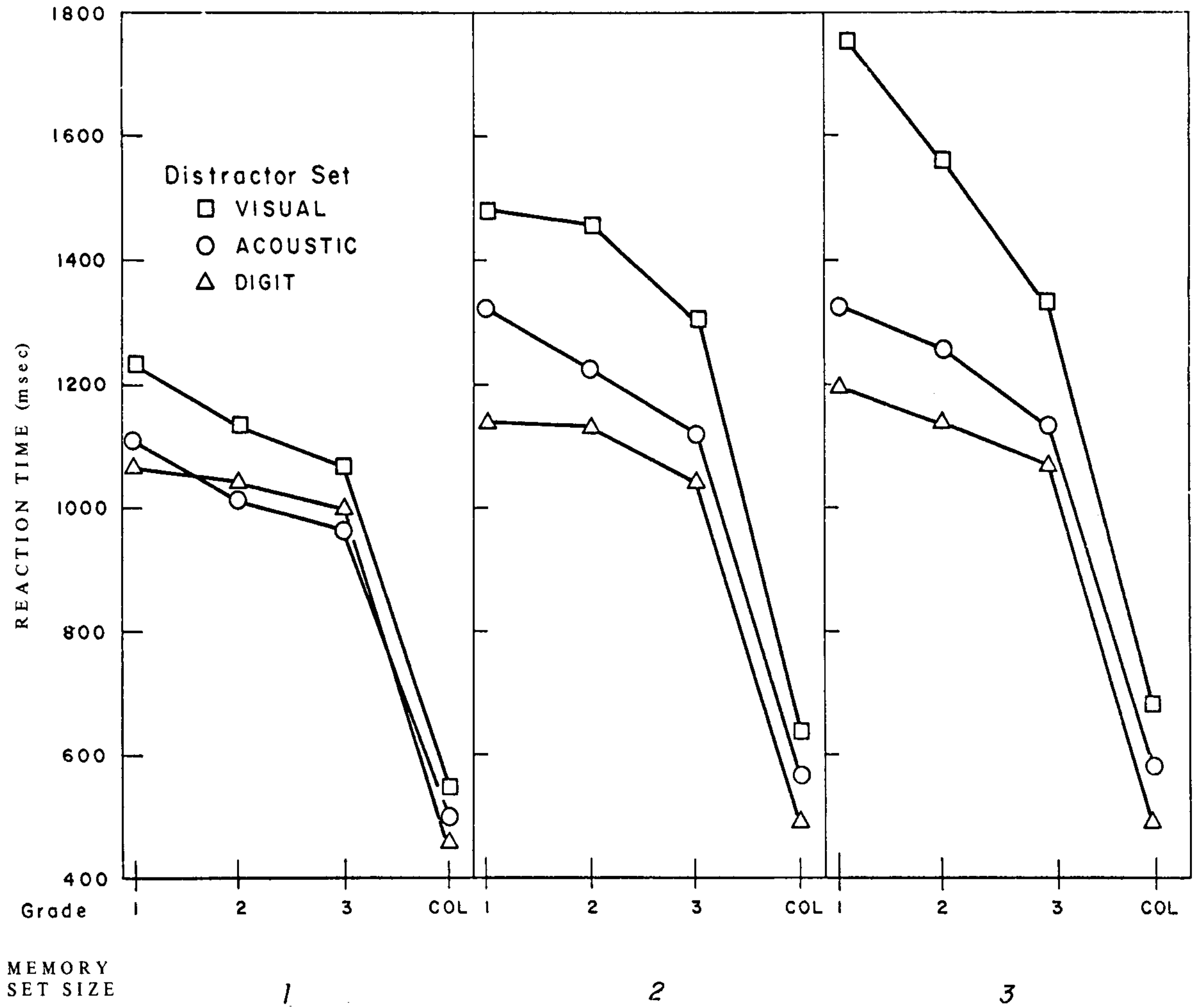

Figure 3. Mean median reaction time as a function of distractor set for each grade and memory set size collapsed across display size (Experiment 2).

Discussion

Essentially, the results replicated the important findings of Experiment 1. With only a single item displayed, subjects could not reliably take advantage of target-distractor dissimilarity to facilitate RT. However, with four items displayed, subjects in all grades showed the category effect in the same manner. Highly confusable distractors led to long RTs; low confusability was capitalized upon to produce shorter RTs. At least with the alphanumeric stimuli used, even 6-year-olds demonstrated that they had learned featural or conceptual differences among the stimuli sufficiently to execute accurate visual search.

Given the procedural changes, we can safely conclude that this phenomenon held across a wide range of letters. The differences between results from a between-subjects and a within-subjects design were not large enough to conclude that this procedural modification is related to important strategic changes.

The addition of an intermediate display size (i.e., two) condition served to emphasize that, with more information about stimulus differences present in the visual field, subjects were able to conduct more efficient search and that this efficiency increased as yet more information was provided visually.

The memory set size factor was included to determine whether the lack of important developmental changes in the first task may have been due to a failure to tax subjects' processing load. Obviously, searching for multiple targets made the task more difficult, as did increasing the number of items displayed. That is, both factors led to increases in RT. However, the increased processing load did not limit subjects' ability to detect stimulus differences, as shown in the category effect. Even when the load was greatest, the category effect was revealed in every grade.

GENERAL DISCUSSION

These two studies converge in finding that children as young as 6 years of age can capitalize upon featural or categorical distinctions among letters and other graphic symbols to facilitate letter search performance. While the degree to which efficiency of such processing increased with age, the similarities in processing were clear-cut. As more of the competing featural information was provided in the visual field, the RTs to the target letter displayed with the various distractors became increasingly differentiated. When this category effect was shown, the patterns of RTs were remarkably consistent across grade, suggesting that subjects of all ages were attending to and processing the stimuli on a similar basis.

There was some evidence in Experiment 1 that college subjects were able to detect some of these differences more efficiently, since the category effect was shown at Display Size 1 for this group. Rather than concluding that college subjects were more sensitive to the featural or categorical differences among the stimuli, we considered the hypothesis that children's performance was limited with a single item displayed, due to an inability to maintain the stimulus differences in working memory adequately. When this load was relieved by increasing the display size, the category effect was shown by subjects of all grade levels. If this memory factor were not limiting performance, then we would have expected category effects to be evidenced at Display Size 1, since there was less information to process. We would not have expected a floor effect for young children, since there was no evidence of such for college subjects. An alternative hypothesis is that children adopted a different strategy when viewing only a single item. They may not have chosen to use or been able to use categorical differences among distractor items to facilitate their decision in this task, and they may have instead compared the target item with whichever item was displayed, without considering the full set of possible stimuli. If this were the case, then no category effect would be expected. Again, this possibility does not preclude a conscious strategy change implemented because of a memory limitation. This type of memory limitation effect has been found often in the RT literature (e.g., Garner, 1974, 1978). At any rate, besides the general increase in speed across age, which may have accounted for the decreases in magnitude of the category effect across age, the Display Size 1 result was the only evidence of a qualitative developmental shift.

In Experiment 2, we sought to explore this phenomenon further. By presenting one, two, or four items in a single display, we hoped to have a task more sensitive to the effect of decreasing the demand to maintain information in memory that would be useful for efficient performance. Results confirmed our original hypothesis. The more information provided in the visual field, the easier it was for subjects to respond with differential speed to the different sets of items. Additionally, we added a memory load by having subjects search for one, two, or three letters. This manipulation would serve to increase the memory load if these memory set items were considered to be discrete items. In this case, we would expect that increasing memory load would decrease the category effect, as we hypothesized for Display Size 1 in Experiment 1. On the other hand, if subjects capitalized upon the similarity of the multiple memory set items, this might effectively have reduced the memory load by giving subjects additional cues as to stimulus differences. For example, when a subject is searching for an A, V, or X in a field of B, C, D, or P distractors, he (she) need only note the presence or absence of a diagonal straight line to respond rapidly and accurately. It need not mean that additional search through the memory set and display is required to compare each target with each displayed item. The results were again straightforward. An increase in memory set size did not have an adverse effect on the category effect. On the contrary, the effect increased in magnitude for all grades. The impact of this factor was similar to that of increasing display size in both experiments.[2]

In summary, the consistencies across grades suggest that the ability to use information about stimulus differences (across a wide range of stimuli within the domain tested) is acquired early, before much formal practice. Subsequent to the initial development of accuracy, it is primarily the effective speed (efficiency) of processing this information that continues to develop. As noted, other studies have pointed to the first grade as the level at which automatic processing of letters is achieved. While no doubt an important landmark in reading development, automatic processing is obviously not the culmination of skill development.

The generally disappointing evidence for external validity can be viewed in the following manner. When considering correlations among tasks indicative of a single "stage" of development, it is imperative that tasks be administered during skill refinement, after the initial period of skill development but prior to mastery (Flavell & Wohlwill, 1969). In these latter two substages, no intertask correlations are expected, due to lack of meaningful variance. In the present case, we have argued that accuracy development was achieved prior to 6 years of age, whereas automaticity and efficiency development were probably operative during the age period sampled. Thus, the paucity of significant intertask correlations could be explained by asserting that reading and cognitive development, as tested, are related only to the acquisition of the knowledge required to discriminate and identify letters. This is not to argue that automaticity and efficiency are in no way related to reading. However, these processes have more to do with deployment of attention to multiple tasks that impose time constraints, and other tasks besides the present letter search task would be required to uncover process commonalities.

In conclusion, we have ascertained that letter processing skill is well developed at an early age. While the letter search task proved a valuable tool in investigating such early developments, we maintain that other paradigms should be used to understand the long-term development of efficiency (and automaticity) of such skills as they relate to cognitive development in general and reading in particular.

REFERENCE NOTE

1. Rosner, J. *The development and validation of an individualized perceptual-skills curriculum* (Publication 1972/7). Pittsburgh, Penn: University of Pittsburgh, Learning Research and Development Center, 1972.

REFERENCES

Bisnanz, J., Danner, F., & Resnick, L. B. Changes in age in measures of processing efficiency. *Child Development*, 1979, **50**, 132-141.

Conrad, R. Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye: The relationship between speech and reading.* Cambridge, Mass: M.I.T. Press, 1972.

Doehring, D. G. Acquisition of rapid reading responses. *Monographs of the Society for Research in Child Development*, 1976, **41**(2), 1-54.

Dunn, J. M. *Expanded manual for the Peabody Picture Vocabulary Test.* Circle Pines, Minn: American Guidance Service, 1965.

Egeth, H., Jonides, J., & Wall, S. Parallel processing of multielement displays. *Cognitive Psychology*, 1972, **3**, 674-698.

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.

Flavell, J. H., & Wolhwill, J. F. Formal and functional aspects of cognitive development. In D. Elkind & J. H. Flavell (Eds.), *Studies in cognitive development: Essays in honor of Jean Piaget.* New York: Oxford University Press, 1969.

Garner, W. R. *The processing of information and structure.* Potomac, Md: Erlbaum, 1974.

Garner, W. R. Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General*, 1978, **107**, 287-308.

Gibson, E. J. *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts, 1969.

Gleitman, H., & Jonides, J. The cost of categorization in visual search: Incomplete processing of targets and field items. *Perception & Psychophysics*, 1976, **20**, 281-288.

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.

Hunt, E. B. Mechanics of verbal ability. *Psychological Review*, 1978, **85**, 109-130.

Jonides, J., & Gleitman, H. The benefit of categorization in visual search: Target location without identification. *Perception & Psychophysics*, 1976, **20**, 289-298.

Keating, D. P., Keniston, A. H., Manis, F. R., & Bobbitt, B. L. Development of the search-processing parameter. *Child Development*, 1980, **51**, 39-44.

Kelley, T. L., Madden, R., Gardner, E. F., & Rudman, H. C. *Stanford achievement test.* New York: Harcourt, Brace, Jovanovich, 1968.

LaBerge, D., & Samuels, S. J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, **6**, 293-323.

Lefton, L. A., & Fisher, D. F. Information extraction during visual search: A developmental progression. *Journal of Experimental Child Psychology*, 1976, **22**, 346-361.

Logan, G. D. On the use of a concurrent memory load to measure attention and automaticity. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 189-207.

Neisser, U. Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology*, 1963, **76**, 376-385.

Pachella, R. G. The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Hillsdale, N.J: Erlbaum, 1974.

Piaget, J., & Inhelder, B. *The child's conception of space.* New York: Humanities Press, 1956.

Post, T. A. Software control of reaction time studies. *Behavior Research Methods & Instrumentation*, 1979, **11**, 208-211.

Raven, J. C. *Guide to using the coloured progressive matrices.* Dumfries, Scotland: Grieve, 1965.

Stanovich, K. E. Toward an interactive compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, in press.

Stanovich, K. E., & West, R. F. A developmental study of the category effect in visual search. *Child Development*, 1978, **49**, 1223-1226.

Wechsler, D. *Manual for the Wechsler Intelligence Scale for Children—Revised.* New York: Psychological Corporation, 1974.

NOTES

1. Pilot testing failed to reveal an effect due to handedness on target presence/absence. This was most likely a function of the ease of responding. No movement to a key was necessary; the actual keypress was attained easily by even kindergartners.

2. As we noted, there was an unfortunate confounding between order of memory set size and practice. However, we do not believe that any practice effect mediated these results, since in Experiment 1 the practice effect was responsible for main effects only.

(Received for publication October 23, 1980;
revision accepted February 17, 1981.)