DOCUMENT RESUME

ED 034 663 24 RE 002 325

AUTHOR	Levin, Harry; And Others
TITLE	The Analysis of Reading Skill: A Program of Basic and Applied Research. Final Report.
INSTITUTION	Cornell Univ., Ithaca, N.Y.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-5-1213
PUB DATE	Dec 68
CONTRACT	OEC-6-10-156
NOTE	390p.
EDRS PRICE	EDRS Price MF-$1.50 HC-$19.60
DESCRIPTORS	Audiolingual Skills, Auditory Discrimination, Beginning Reading, Eye Voice Span, *Investigations, Linguistic Theory, *Oral Reading, Reading Diagnosis, *Reading Processes, *Reading Research, *Reading Skills, Spelling, Structural Grammar, Theories, Visual Discrimination

ABSTRACT

With the contention that increased concern over the status of reading requires that the experimental testing of reading hypotheses be conducted with the most modern methods of behavioral science and be based on a theoretical analysis of the reading process, both psychological and linguistic, this final report from a group of investigators pursues various ideas with individual reports. Three are on aspects of visual and/or auditory relationships to reading, and five deal with specific investigations of graphic discrimination. A majority of the papers are grouped under "Studies of Oral Reading." Seven of them cover a broad scope of the relation of spelling to reading ability--three of these dealing with homographs; eye-voice span is considered in six--four of these relating eye-voice span to syntax. Three articles deal with beginning reading in terms of grammatical context, visual perception, and scribbling and drawing. The abstracts to five related theses written by research assistants working on the project are included. Each paper includes its own references. Most of them include abstracts and tables. (BT)

ED034663

BR 5-1213
PA 24

FINAL REPORT

BR 5-1213
PA-24
OE/BR

Project No. 5-1213

Contract No. OE 6-10-156

THE ANALYSIS OF READING SKILL: A PROGRAM OF BASIC AND APPLIED RESEARCH

Harry Levin

Eleanor J. Gibson James J. Gibson

Department of Psychology
Cornell University
Ithaca, New York 14850

December, 1968

U. S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

FINAL REPORT

Project No. 5-1213

Contract No. OE 6-10-156

THE ANALYSIS OF READING SKILL:

A PROGRAM OF BASIC AND APPLIED RESEARCH

Harry Levin
Eleanor J. Gibson James J. Gibson

Department of Psychology
Cornell University
Ithaca, New York 14850

December, 1968

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Table of Contents

STUDIES OF ORAL READING:

Introduction

When work on this program was begun, we held a strong bias as regards to the kind of work which we were equipped to do and which, we thought, was needed. Our bias was the conviction that our research should be basic; that it should stem from a theoretical analysis of the reading process, both psychological and linguistic, and that the studies developing from the analysis should be carried out with the most modern methods of behavioral science. The time was ripe, we thought, to contribute to the understanding of reading from the point of view of behavioral science, and to contribute to behavioral science by a study of the reading process. The search was to be for theories and correct principles, and the methods to be experimental techniques and careful description.

The number of publications bewailing the present status of reading instruction which have appeared during the course of this project strengthened our faith that a new approach, based on theory and culminating in experimental tests of our hypotheses, was a good choice. We have been heartened, also, by the interest of psychologists and educators alike as soon as we had material ready to report at meetings and in preliminary written form. The research must stand on its own, but we feel that some new directions have been indicated. If we have helped lure the educator into the laboratory and the scientist into an educational problem, something at least will have been achieved.

The organization of the research group was a loose federation with each of the principal investigators pursuing his own ideas and taking sole responsibility for them. The completed studies are presented in the report as individual papers, arranged in groups ordered according to the principal investigators and the work which each one took responsibility for. Some of the sub-projects were presented as dissertations by research assistants working on the project, and in these cases a large share of the responsibility was assumed by the candidate. Much that was creative and new came from these dissertations. The original group of investigators acknowledges a debt of gratitude not only for assistance from these people but for ideas as well.

A Developmental Study of the Effects of Visual and Auditory Interference on a Visual Scanning Task[1,2]

Eleanor J. Gibson and Albert Yonas

Abstract

The effects of visual and auditory interference on a visual scanning task were compared with children from the third grade and college sophomores. A highly confusable visual context significantly reduced scanning rate for both children and adults, but a highly confusable auditory context, played over earphones, had no effect on either group. There was a significant age interaction with interfering visual context. It seems likely that theories assuming auditory encoding of visually-presented graphic items have little predictive value for a scanning task.

In a previous experiment, Gibson & Yonas (1966) found a significant age difference in the time required to find a visual target, a letter, embedded in a list of other letters. The scanning task was adapted from one devised by Neisser (1963). When the context letters were made highly confusable with the target letter, search time increased

1. This research was supported by NIH Grant MH-07226-02 and by Grant OE6-10-156 from the U. S. Office of Education.

2. We are grateful to Mr. Hart, Principal of the Northeast School of Ithaca, N. Y., for his cooperation in providing laboratory space in the school.

at all age levels, showing a strong interfering effect of visual context. In considering the possible reasons for the longer latencies in children, the hypothesis occurred to us that children in the early stages of reading skill tend to articulate vocally to a greater extent than adults, particularly when difficulties are encountered. It might be that in the present task, children were processing the letters during the search with more explicit vocal responses, or naming the target in order to keep attention on it. If this were the case, one might expect that introduction of a confusing auditory background of letter names coming in through earphones would interfere and increase scanning time for children as compared to adults.

Furthermore, recent studies of short-term memory and recognition of visually presented material have held the implication that the visual material is encoded to an acoustic representation (Conrad, 1964; Sperling, 1963). This raises the question whether visual processing of letters involves such encoding even in a scanning stage. Sperling proposed, for example, that one source of information for auditory storage might be the scanning process; it might be that "observers hear themselves make a verbal response as they scan" (p. 27). If this transformation actually occurs in scanning, it might be expected that even adult Ss would show interference effects from a confusing auditory background. In fact, one would surmise that the effects would be at least as great as with visual confusability.

The present experiment thus sought to compare the interfering effects of a highly confusable auditory context with a highly confusable visual one, and also to compare these effects in children and adults.

Method

The basic task was the same as in the previous experiment, scanning down a list of 30 letter-strings of four letters each, looking for a designated target letter. The letters were ¼ in.-high capitals typed with an IBM sign typewriter. The typed list was 10 in. long. The S was instructed to start at the top of the list and scan downward until he found the target letter, proceeding as rapidly as possible. He was urged to scan downward, skipping no lines. He held a push button in his hand which he pressed when ready to begin scanning. This caused a list to appear and started a Hunter clock-counter. The S immediately started scanning down the list. When he found the target, he pressed the button again, stopping the clock and turning off the lamps. The E then recorded the time and placed another list in the display box. If S had failed to find the target, the list was rerun at the end of the series. Five practice trials were given.

Auditory input accompanied the scanning in all conditions. The S wore earphones over which he heard a male voice naming letters at the rate of two per sec. The sound came on as he pressed the starting switch and ceased when he found the target letter and pressed the switch to stop his scan.

The design of the experiment incorporated three conditions, each including 20 trials with 20 different lists. Condition I had a low-visual, low-auditory confusability of context. The target letter (G) was embedded in a context of letters selected for low confusability in the previous experiment. Its position was randomized across and down the 20 lists, with the restriction that it appeared equally often in the four quadrants of the list from top to bottom and never in the same line more than once. The letters heard over the earphones were selected for low acoustic confusability with the target letter G. The values were taken from an auditory confusion matrix for letter names obtained by Conrad (1964).

Condition II had low-visual, high-auditory confusability. It was similar to I, with the same target letter and context letters. Twenty new lists were made up as before. However, the letters heard over the earphones in this condition were six letters of high-auditory confusability with G.

Condition III had high-visual, low-auditory confusability. The letters heard over the earphones were the same as those in Condition I, but the context letters had high visual confusability, the same lists as those run in this condition of the earlier experiment.

The Ss were 18 children who had just completed third grade, and 18 college sophomores. The experiment was run during the summer, making it possible to run each S in all three conditions. The order in which the conditions were run was counterbalanced over the subjects of each age group to balance practice effects.

Results

The time-scores were transformed to take account of position of the target in a list, and averaged to determine scanning rate. As in the earlier experiment, linearity of slope over the 30 positions was not perfect, owing to an initial lag in the first few list positions. It was decided, as before, to include all scores in the analysis since the slope was similar for conditions and ages. It may be seen in

Fig. 1 that scanning rate was faster for the adults as compared with the children in all three conditions. Analysis of variance showed that the age difference was significant ($p < .001$).

Condition III (high-visual, low-auditory confusability) was more difficult than both the other conditions ($p < .001$), lowering the scanning rate by .36 sec. for adults and by .55 sec. for children as compared with the control condition of low-visual and low-auditory confusability. On the other hand, the condition of high-auditory confusability did not reduce the rate of scan for either the adults or the children. For the children, in fact, speed was slightly (but not significantly) greater with high-auditory confusability.

The interaction of condition with age was significant at the .05 level of confidence. All the variance in this interaction was contributed by the condition of high-visual confusability, which reduced rate of scan relatively more for the children than for the adults. We did not find a similar interaction in the earlier experiment, but since the condition of high-visual confusability always had the benefit of practice in that experiment, it is possible that a small interaction might have been masked.

It is of some interest whether noise as such acted as a distraction, to slow the visual scanning rate. The comparison of the present experiment with the previous one without auditory noise is not perfect, since practice was not balanced in that experiment, the high-confusion condition always getting the benefit of practice. The rate of scan for adults in the condition of low visual confusability was faster in the present experiment (.14 sec. as compared with .20 sec.), probably due to greater benefit of practice in the present experiment. The rate was very similar in the conditions of high-visual confusability (.51 in the earlier experiment and .50 in the present one), where practice was comparable. There is thus no evidence that hearing letters spoken while scanning was distracting at all. There was no exactly comparable age group for children in the two experiments.

Discussion

The strong effect of confusing visual context was confirmed in the present experiment. But attempts to explain the effect on the basis of interfering acoustic representations are discouraged by the total lack of interference from high auditory confusability. These results definitely weaken the hypothesis that visually perceived letters are encoded to acoustic representations as they are scanned. It is possible that such encoding may occur in a recall task, when rehearsal is attempted, but no such strategy appears to be taking place when the task is one of detection.

That information coming in from auditory and visual channels does not combine easily was suggested by an experiment of Broadbent & Gregory (1965). They presented three visual digits in succession, with a blank interval of 2/3 sec. between each digit and the next. During each of the blank intervals, a spoken digit was transmitted over a sound channel. Reproduction of the six digits was considerable poorer than when all six digits were presented visually. It is reasonable, in the light of this finding, that interference between auditory and visual chennels should not occur, as we found. Letters of high auditory confusability presented over a visual channel could conceivably have a greater interfering effect, and the possibility is being investigated.

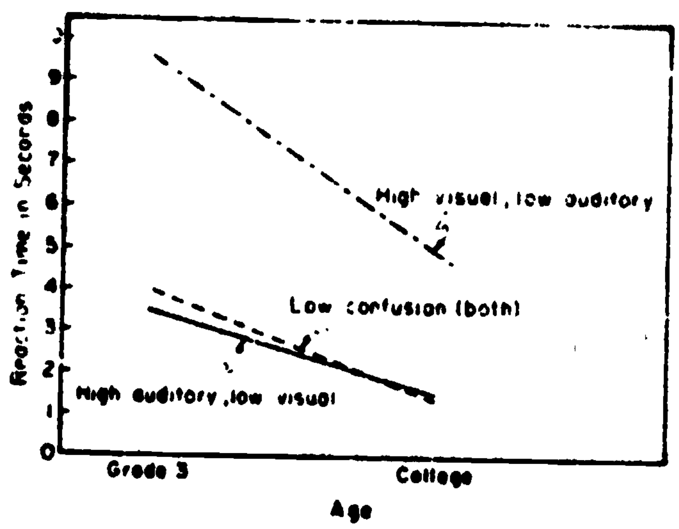


Fig. 1. Mean rate of scan for children and adults on three scanning tasks.

References

Broadbent, D. E., & Gregory, M. Some confirmatory results on age differences in memory for simultaneous stimulation. Brit. J. Psychol., 1965, 56, 77-80.

Conrad, R. Acoustic confusions in immediate memory. Brit. J. Psychol., 1964, 55, 75-84.

Gibson, E. J., & Yonas, A. A developmental study of visual search behavior. Percept. & Psychophys., 1966, 1, 169-171.

Neisser, U. Decision-time without reaction-time: Experiments in visual scanning. Amer. J. Psychol., 1963, 76, 376-385.

Sperling, G. A model for visual memory tasks. Hum. Factors, 1963, 5, 19-31.

Visual and Acoustic Confusability in a Visual Search Task[1]

G. A. Kaplan, A. Yonas and A. Shurcliff[2]

Cornell University

Abstract

Visual and acoustic confusability between a target item and background items was varied in a visual search task. Visual confusability was a highly significant source of difficulty while acoustic confusability had no effect. The results do not seem to be interpretable within a theory which assumes compulsory auditory encoding of visual information.

Several models have recently been postulated which assume a stage of acoustic or auditory transformation of visual input. The models have been derived from studies of such varied phenomena as perception of figure and number strings (Glanzer & Clark, 1962, 1963), latency periods in the reading of words aloud (Levin & Biemiller, 1965; Beimiller & Levin, 1966), amount of recall in a visual-memory task (Sperling, 1963), amount of recall of a written message for a given exposure duration (Mackworth, 1962, 1963), correlation between errors made in immediate recall of visually presented material and those made in the recall of auditorily presented material (Conrad, 1964), and immediate recall of visually presented letter strings (Conrad, Freeman, & Hull, 1965). The above authors have argued that incoming visual information is eventually encoded and stored in a form related to its acoustic representation. However, with the exception of some discussion of the time parameters of this process of encoding, there has been little critical discussion of the generalizability of this process across or within visual tasks.

The present study investigates the relative importance of visual and acoustic factors in a scanning task developed by Neisser (1963), performance on which was shown by Gibson and Yonas (1966a) to be highly affected by the amount of visual confusability between the target and the background. The nature of the task allowed acoustic confusability and visual confusability between the target and the background to be independently manipulated.

1. This research was partially supported by grant OE 6-10-156 from the United States Office of Education.

2. The authors would like to thank Dr. E. J. Gibson for her valuable ad ce and support.

Method and Procedure

The apparatus was a modified version of Neisser's (1963) as described by Gibson and Yonas (1966a). A list was composed of 30 lines, four letters wide, typed on a sign typewriter. A line consisted of four letters randomly sampled (with replacement) from the background set of four letters, with the constraint that no letter appear more than twice per line. Each list was 5/8 in. wide with 1/4 in. lines spaced 1/8 in. apart, and the entire list was 10 in. long. For each experimental group, a single target letter (E or K) could appear once in each list on one of 16 lines (1, 3, 5, ..., 29, 30). The relationship between the target letter and the background letters was varied by manipulating the amount and type of confusability between the two. Background-target combinations of high acoustic confusability/high visual confusability (H_a-H_v), high acoustic/low visual (H_a-L_v), low acoustic/high visual (L_a-H_v), and low acoustic/low visual (L_a-L_v) were generated using confusion matrices for acoustically presented letter names and visually presented letters obtained by Conrad (1964) and Gibson, Osser, Schiff, and Smith (1963), respectively. For each target letter the remaining letters of the alphabet (excluding Z) were ranked both acoustically and visually with respect to the number of times they were confused with the target letter. Conditions H_a-H_v, H_a-L_v, L_a-H_v, and L_a- L_v were then generated by combining letters which were ranked appropriately for a given condition. The target-background combinations together with the rankings for each member of the set are shown in Table 1.

Table 1. Background-Target Confusability

Target - E

		H_a				L_a			
		D	P	G	T	F	H	L	J
H_v	R_a	5.0	1.0	9.0	4.0	22.0	11.0	22.0	24.0
	R_v	3.0	5.5	11.0	11.0	1.0	2.0	5.5	5.5
		V	D	U	C	K	I	Y	X
L_v	R_a	6.0	2.0	3.0	10.0	25.0	15.0	20.0	19.0
	R_v	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0

Target - K

		H_a				L_a			
		N	L	F	H	X	V	Y	R
H_v	R_a	3.0	7.0	10.0	11.0	18.0	15.0	12.0	14.0
	R_v	2.0	7.5	7.5	4.5	1.0	4.5	3.0	7.0
		A	B	O	P	C	G	I	T
L_v	R_a	1.0	2.0	4.0	8.0	25.0	24.0	22.5	17.0
	R_v	20.0	20.0	20.0	20.0	20.0	20.0	20.0	20.0

The visual (R_v) and acoustic (R_a) confusability rankings for each background letter with its target. A rank of 1 indicates highest confusability.

The Ss were 24 college students fulfilling a course requirement. Half were randomly assigned to each target letter and received the four confusability conditions in a random order, the order being counter-balanced across target conditions. Each S searched through 16 lists per condition for a total of 64 lists. The line on which the target was presented and the position in the line where it appeared were balanced across conditions and randomized within conditions.

Ss were instructed to fixate a small spot of light which indicated the top of the list and allowed them to converge properly. When ready they were told to press a small hand-held switch which would expose the list and start a clock. Instructions were to scan from the top down as quickly as possible without missing the target. When they reached the target they pressed the switch which removed the list from view and stopped the clock. Ss were told the target could appear on any line, and at the end of each trial they reported the approximate position of the target. If S missed the target, the list was rerun at a later time. Scanning time was recorded by a Hunter "Klockounter" to the nearest msec.

Results

The data consist of 64 scanning times per S, divided equally across the four conditions. Since what is of interest is the time taken to process each line, the time taken to scan each list was transformed by dividing by the line on which the target appeared on that list. This gives a measure of the rate of scanning adjusted for the target position. In general, a linear relationship would be expected between target position and scanning rate. The data confirm this expectation with the exception of the instances where the target item appeared close to the beginning of the list. This initial nonlinearity was also found by Gibson and Yonas (1966a) and is, presumably, due to contamination of the rate measure by factors peculiar to the beginning of search. Thus, in order to get as pure a measure of the rate as possible, the data for the trials where the target appeared on any of the first four possible lines (lines 1, 3, 5, & 7) were excluded from the analysis. Table 2 presents the means for the two target groups, averaged across the Ss in each group and across the remaining 12 target positions.

Table 2. Mean Scanning Rates for E & K Groups (in sec.)

		E			K	
		Acoustic confusability			Acoustic confusability	
		hi	lo		hi	lo
Visual	hi	.2675	.2674	hi	.3123	.3315
Confusability	lo	.1760	.1461	lo	.2147	.1964

The data for each target group was treated as a single experiment and analyzed by an appropriate analysis of variance. A descriptive summary of the analysis is presented in Table 3. The most striking result is that while the visual factor is highly significant in both groups

Table 3. Percentage of total variance accounted for.

	% of Variance		d.f.	
	E	K	E	K
Subjects (S)	26.82	42.86	11	11
Acoustic (A)	.64	.00	1	1
Visual (V)	32.03	22.48	1	1
Target line (L)	3.95	2.83	11	11
A x V	.62	.56	1	1
A x L	.76	.48	11	11
V x L	.20	.53	11	11
A x V x L	.91	.51	11	11
Ss-interactions	34.06	29.74	517	517
Total	100.00%	100.00%	575	575

($F = 56.37, 49.42$; df = 1/11; $p < .001$, for E and K, respectively), the acoustic factor is negligible and non-significant ($F = 1.25, 0.00$; df = 1/11; n.s., for E and K, respectively). The target line effect is significant in both analyses ($F = 8.32, 8.30$; df = 11/121; $p < .001$, for E and K, respectively). The A x V, A x L, V x L, and A x V x L interactions are not significant in either analysis. However, the interactions of A and/or V with Ss are all significant beyond the $p = .001$ level.

Because the experiments were analyzed separately it is not possible to test directly the possibility of an interaction of the main effects with the two different target groups. This was done directly by using a t test for independent groups to compare the distribution of differences due to the level of acoustic or visual confusability in the E group with the same distribution in the K group. Neither analysis indicated any significant differences. Visual inspection of the data indicated that a test for an interaction between target groups and target position was not warranted.

Discussion

The results of this study support those of Gibson and Yonas (1966b) who found that a potentially distracting auditory context (played through earphones) did not interfere with a scanning task identical to the present one. Likewise, Glucksberg (1965) found no effect due to acoustic similarity of items in a short term visual memory task. Chase (1965) compared a visual recognition task, a visual comparison task, and a memory search task and found no effect of acoustic confusability of items in any of the three tasks. He did, however, find a strong effect due to visual confusability (defined by an overlap measure) in his visual comparison task, a result which is supportive of our own.

It should be pointed out that the acoustic and visual interactions with Ss were all highly significant. We interpret this to indicate that different strategies are used by our Ss. However, there were no indications in the data of any S which would indicate a larger effect due to acoustic confusability than that due to visual confusability. In addition, we find no support for the speculation (Kaplan & Carvellas, 1965) that Ss rehearse the target letter; both the reports of our Ss and the results would seem to preclude such an interpretation.

It seems clear that no process which relies on the acoustic encoding of visual information can account for our results. Visual confusability was the major source of difficulty. The scanning task therefore is at least one instance where explanatory mechanisms that assume compulsory auditory representation are of no predictive value. There will, of course, be situations which require some form of acoustic or auditory transformation of visual information; e.g., perhaps those which require rehearsal or verbal identification. However, the contextual sensitivity of the perceptual mechanisms must be emphasized: Gibson (1966) and Bower (1965) have suggested that perceptual learning and visual selection, respectively, can be thought of as processes in which there is a filtering of information according to its utility. Such a view is consistent with that taken here; given a visual scanning task in which a visual comparison strategy is sufficient for maximal performance, it would be surprising if acoustic confusability had any effect.

References

Biemiller, A. J., & Levin, H. Studies of oral reading II. Pronounceability. Preliminary draft, 1966.

Bower, T. G. R. Visual selection: scanning or filtering. Psychon. Sci., 1965, 3, 561-562.

Chase, W. G. The effect of auditory and visual confusability on visual and memory search tasks. M. A. Thesis, Univ. of Wisconsin, 1965.

Conrad, R. Acoustic confusions in immediate memory. Brit. J. Psychol., 1965, 55, 75-84.

Conrad, R., Freeman, P. R., & Hull, A. J. Acoustic factors versus language factors in short-term memory. Psychon. Sci., 1965, 3, 57-58.

Gibson, Eleanor J. Perceptual development and the reduction of uncertainty. Paper to be read at the International Contress of Psychology, Moscow, 1966.

Gibson, Eleanor J., & Yonas, A. A developmental study of visual search bheavior. Percept. & Psychophys., 1966a, 1, 169-171.

Gibson, Eleanor J., & Yonas, A. A developmental study of the effects of visual and auditory interference on a visual scanning task. Psychon. Sci., 1966b, 5, 169-170.

Gibson, Eleanor J., Osser, H., Schiff, W., & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In A basic research program on reading. Final Report on Cooperative Research Project No. 639 to Office of Education, Department of Health, Education, and Welfare.

Glanzer, M., & Clark, W. H. Accuracy of perceptual recall: an analysis of organization. J. verbal Learn. verbal Behav., 1963, 1, 289-299.

Glanzer, M., & Clark, W. H. The verbal loop hypothesis: binary numbers. J. verbal Learn. verbal Behav., 1963, 2, 301-309.

Glucksberg, S. Decay and interference in short-term memory: comments on experiments by Steffy and Eriksen. Paper read at Psychonomic Society meetings, Chicago, 1965.

Kaplan, I. T., & Carvellas, T. Scanning for multiple targets. Percept. mot. Skills, 1965, 21, 239-243.

Levin, H., & Biemiller, A. J. Studies of Oral Reading III. Contingent versus non-contingent spelling patterns. Preliminary draft, January, 1966.

Mackworth, Jane F. The relation between the visual image and post-perceptual immediate memory. J. verbal Learn. verbal Behav., 1963, 2, 75-85.

Mackworth, Jane F. The visual image and the memory trace. Canad. J. Psychol., 1962, 16, 55-59.

Neisser, U. Decision-time without reaction-time: experiments in visual scanning. Amer. J. Psychol., 1963, 76, 376-385.

Sperling, G. A model for visual memory tasks. Hum. Factors, 1963, 5, 19-31.

Utilization of Spelling Patterns by Deaf and Hearing Subjects

Eleanor J. Gibson, Arthur Shurcliff and Albert Yonas

Cornell University

For more than half a century, we have known that a good reader does not read sequentially, letter by letter, but takes in and processes larger graphic units (Cattell, 1885). Yet, except for the obvious surmise that a familiar word constitutes a unit, there has been little research or even speculation on the relevant grouping principles. How are larger graphic units constituted? By meaning? By frequency? By rules? If rules, what kind of rules? A large number of experiments on the perception of words[1] has shown that word frequency is correlated with speed and accuracy of perception. But more recent experiments have shown that other principles than frequency play a role in the formation of graphic units. Among these are degree of approximation to English (Miller, Bruner and Postman, 1954; Wallach, 1963) and internal structure of letter strings (redundant strings as opposed to random ones, as defined by a grammar or set of rules, Miller, 1958). In general, any property that increases redundancy is a good possibility for facilitation of perception, recognition, or retention. One of the major psycholinguistic generalizations which can be offered, according to Diebold (1965) in a review of recent psycholinguistic research, is that speech recognition increases directly with the increase in redundancy for all sizes of message units. Surely we can expect this principle to apply to written messages as well as spoken.

---

1. See Gibson et al., 1962, for a review of these experiments.

One property which one might think of as contributing redundancy, so as to reduce information and facilitate reading a string of letters as a unit, is invariance of spelling-to-sound correspondence. Hockett (1963), Venezky (1963), and Venezky and Weir (1966) have all worked on the problem of how English spelling patterns are mapped to sound. One often hears the assertion that spelling-to-sound correspondences in English are irregular and unpredictable, but the work of these linguists demonstrates that the mapping rules and constraints are there, if graphic units larger than the single letter are considered. The hypothesis was proposed, therefore (Gibson, Pick, Osser and Hammond, 1962), that units for reading are formed by a relatively invariant mapping to speech sounds. For English spellings this would mean that letter clusters in a given position in a word and in a given environment, when they map with regularity to pronunciation, will operate as units and that grouping is functionally determined by the relationship to speech sounds. A stimulus property that is invariant over a set of items constitutes a constraint or "rule" that is a good bet for creating a unit or "chunk" of otherwise randomly organized parts.

A letter string that has high internal transitional probabilities is not the same thing as one that maps with invariance to speech. "ATI," for instance, is a high frequency trigram, but it is not a unit for reading, since it is pronounced differently, depending on context (cf. relation vs. relative). Certain clusters are always pronounced the same way, wherever they appear (e.g., SH). But some are pronounced differently depending on their location in the word (e.g., GH in GHOST vs. ENOUGH). This condition is rule-like, however, and constitutes a higher-order constraint; GH is always pronounced in one way at the beginning of a word but never this way at the end.

That the reader is not always aware of mapping rules from spelling to speech need not mean that order and regularity are not abstracted and used in the course of learning to read. As Venezky and Weir (1966) have shown, the rules are indeed "high level" ones and a long program of computer-aided research was required to formulate them. The question is, does the skilled reader, knowingly or not, actually use them in perceiving written language?

We tried to answer this question with an experiment in which pseudo-words--that is, letter-strings which were not real words, but which in some cases might have been--were presented tachistoscopically to skilled readers of English. The words were all monosyllables, consisting of an initial consonant cluster, a vowel cluster, and a final consonant cluster. Half of them were constructed so that the initial consonant cluster had a single regular pronunciation in that position, the final one in its position and the vowel cluster a regular pronunciation when preceeded and followed by the selected consonant clusters. These were called pronounceable words, because the clusters had an invariant mapping from spelling to sound in this arrangement. A control set of words was constructed from the same letters, but with the initial and final consonant clusters reversed, rendering them unpredictably pronounceable. These were called the unpronounceable words. For instance, a pronounceable pseudo-word was GLURCK; its unpronounceable counterpart was CKURGL.

This experiment was run and replicated several times (Gibson et al., 1962) and very consistently gave significant results in the predicted direction. In the meantime, several theoretical questions were brought up by members of the research group as to the exact interpretation of the results. What did pronounceability really mean? Was it actually the invariance-to-sound-mapping that was crucial? Partial answers to these questions were sought in the two following experiments. All the words were rated for pronounceability on a 9-point scale (following the method of Underwood and Schulz, 1960). The words constructed to be pronounceable were indeed rated high in pronounceability, their counterparts low. Secondly, 16 subjects were asked to read aloud all the words, and their pronunciations were recorded on tape. The pronunciations were analyzed by two linguists and the variability of pronunciation was determined for each word. Variability was very high for the unpronounceable words, but low for the pronounceable ones. The variability score correlated .83 with the pronounceability rating.

The results were published at this point, and soon afterward several alternative interpretations were suggested. In all, five different interpretations appeared to warrant consideration or test. They are as follows:

1. Rules of spelling-to-sound mapping suggest that mapping-invariance creates larger units for reading and therefore faster processing. This is the original hypothesis described above.

2. Transitional probabilities in written English, without regard to sound, account for the superiority of the so-called pronounceable words. It was suggested (Anisfeld, 1964) that summed bigram or summed trigram

counts would predict the results obtained in the experiments. The counts were made and correlated with number of correct perceptions (Gibson, 1964). The count did not predict success in reading the pseudo-words when pronounceability was partialled out and length held constant.

3. Pronounceable words are more readily perceived because they "match" to an acoustic representation. When a word (or letter-string) is exposed, it is silently rehearsed and matched with a stored auditory representation (Levin and Biemiller, 1965). This hypothesis implies auditory encoding before successful reading of a pseudo-word.

4. Processing of letter-strings in reading involves encoding and matching to an articulatory represetnation or "plan." (cf. Liberman et al., 1964).

5. Complex orthographic rules cover structural patterns of letters permissible in English words. Such rules are not merely transitional probabilities but are a kind of syntax, analogous to grammar. Such rules could be learned, as one learns to read, with or without relating them to speech sounds. An obvious example is the case of consonants or consonant clusters which cannot be used initially but can finally, and vice versa, such as CK and QU. The principle was used in construction of words in the present experiment. Mapping to speech, when invariant, would be an added, redundant constraint.

The Experiment

Because a resolution among these alternatives would have important implications for teaching reading, experiments were sought which might decide among them. A comparison of deaf and hearing subjects, it was thought, should be instructive, especially when various potential predictor variables were weighed against performance with pronounceable vs. unpronounceable words. Accordingly, the original experiment was modified slightly for replication with deaf subjects.

Method

The method followed closely that of the original experiment except that all instructions were presented in writing and a red light was used as a ready signal. The words used were the same as those used by Gibson et al. (1962), except that two pairs were dropped because the difference in their pronounceability ratings was slight. The words varied in length from four to eight letters.

The words were projected tachistoscopically one at a time on a screen for 100 m.sec. Contrast between letters and background was high enough so that some letters could always be read. The pronounceable and unpronounceable items were projected in a random order. After the series had been shown, it was repeated, this time in a reverse order, with the same exposure time. The subjects were seated ten feet from the screen. They wrote what they saw after each presentation on numbered lines, four lines on each page of a scoring sheet, and turned to a new page after one was filled. This procedure was followed to assist them in keeping to the order. The subjects were run in small squads of three or four to permit equivalent viewing angles. The height of the letters projected on the screen was 3 3/4 in., the width about 2 1/2 in. Exact instructions were as follows:

"This is an experiment on reading sets of letters when they are flashed on the screen in front of you for a very short time. The letters do not form real words, but try to read them and write down what you see. We will call them nonsense words. Write all the letters of the nonsense word in the order that you saw them, if you possibly can. If you are not quite sure, write them down anyway, as you think you saw them.

I will show you some practice nonsense words, before we begin the experiment. A red light will flash, so you can be ready. When the light flashes, look at the center of the screen. The nonsense word will be flashed on the screen just one second after the ready signal. After the nonsense word is flashed, write down what you saw at once. Here is a practice word. Watch for the red light, and then for the word."

Subjects

Thirty-four subjects were secured at Gallaudet College for the Deaf in Washington, where both staff and students cooperated in every possible way. An interpreter was present in case the subjects had any questions. We requested subjects who were congenitally deaf, or nearly so, and who had maximal hearing losses. After the staff furnished for *Ss* hearing ratings, speech ratings, age of onset of deafness, and scores on various tests such as reading and verbal aptitude. At the close of the experiment, the subjects themselves answered questionnaires regarding the way they were taught to read.

A new control group of 34 subjects was run with the same procedure as that used with the deaf subjects. They were Cornell students drawn from an introductory psychology class. The native language of all *Ss* was English.

Results

Comparison of deaf and hearing subjects. The deaf subjects, not surprisingly, made more errors than hearing subjects for both pronounceable and unpronounceable words. But the difference between the two sets

Table 1

	Mean Errors	
	Pronounceable Words	Unpronounceable Words
Deaf	21.27	36.36
Hearing	15.86	25.68

of words is just as significant and just as striking for the deaf as for the hearing. Whatever it is that facilitates reading the words in the pronounceable list seems to be operating equally well for them. Labelling the difference between the two lists pronounceability evidently served only to "pull the wool over our eyes," for the deaf students had never heard the words pronounced. Another experiment with similar results, though on a smaller scale, recently came to our attention (Doehring and Rosenstein, 1960). In that experiment, lists of trigrams were shown tachistoscopically to deaf and hearing children. The lists were roughly equivalent in frequency, but one list was all CVC, such as ZIF, while the other was CCC, such as RCH. The authors referred to the first list as pronounceable and the other as unpronounceable. Both groups of children made fewer errors on the CVC trigrams, the deaf being relatively at least as much better on these as the hearing.

Error data, by words, for both our groups of subjects were correlated with data from the earlier experiment, as a check on reliability. The coefficients were very high, .91 and .96 respectively, indicating excellent replicability and similar performance on individual words by the deaf and earlier hearing subjects.

It seemed possible that an examination of the errors might reveal something more, since the scoring used to obtain the criterion for the measures in Table 1 was simply "right" or "wrong". We were fortunate in having available a computer program for the analysis of graphic errors in reading which would allow us to compare part scores of several types, using as data the actual errors-- the wrong spellings recorded by the subjects. These errors were all punched on cards, along with the correct spelling of the word, and the error compared with the word exposed. The spellings were compared for the number of <u>same</u> letters in the two (regardless of position), the length of a correct string reading in a forward direction (starting in any position), the length of a correct string reading in reverse, the number of letters same at the beginning of a word, and the number same at the end.

Means are given for these counts in Table 2, separated for deaf and hearing <u>Ss</u>, and for pronounceable and unpronounceable words. When the five measures are compared for pronounceable and unpronounceable words,

---

Insert Table 2 about here

---

there are some slight differences in the expected direction. The length of a correct letter sequence in the forward direction is longer for the pronounceable pseudo-words. The mean number "same" in sequence at the beginning is somewhat longer for pronounceable words. The number of letters same without regard to order, and the length of span correct in reverse order go, to a small extent, in the other direction, which is reasonable, since length in these cases is symptomatic of error, rather than accurate reading. There is no other significant difference here. The slightly longer mean number correct at the end for pronounceable words for the deaf subjects may simply have to do with the fact that they made more errors. This is certainly the explanation of the slightly larger means in several other cases of the deaf subjects, compared with the hearing, since it was not feasible to correct for total number of errors. A deaf subject might get four letters correct in forward order in a five-letter word, whereas hearing subjects made no errors. At any rate, there is no indication of an interaction here, that we might have missed in our original method of scoring by number of words wholly correct.

Prediction of errors. In an attempt to analyze what accounts for the facilitation of the "pronounceable" words, we examined the effect of a number of potential predictor variables in several multiple regression analyses. The data used in these analyses are presented in a correlation matrix in Table 3. Errors for hearing and deaf subjects were correlated

---

Insert Table 3 about here

---

separately and in combination with word length, pronounceability rating, summed bigram frequency, summed trigram frequency (the two latter measures taken from Underwood and Schulz, 1960, combined count), and two recent measured of summed bigram and trigram frequency compiled by Mayzner and Tresselt (1965), and Mayzner, Tresselt and Wolin (1965). These last two measures take into account letter position in the word and word length.

Table 4 presents the results of these analyses. Errors were, in every analysis, the dependent variable. Each row across the table presents one analysis. Since the analysis was step-wise, the variable tested first is presented in the first column, the next second, and so on, until variables significant at the .05 level or better are exhausted. Under the first variable is listed the percent of variance which it accounts for, and under the following the cumulative variance accounted for when another significant predictor variable is added. When variables were omitted from an analysis, it is so indicated.

Insert Table 4 about here

In the first three rows, all variables were included, with hearing errors as dependent variable in row 1, deaf errors in row 2, and hearing and deaf combined in row 3. The variable tested first (with highest predictability) was length of word. It accounts for 67 to 73 percent of the variance; a reasonable finding, for the longest words were always perceived erroneously and the shortest most successfully. Pronounceability is the second significant predictor variable in all three rows and when added to length, 84 to 87 percent of the variance is accounted for. The only frequency count reaching significance is the Mayzner bigram count, and it adds only one percentage point for the deaf Ss and none for the hearing.

These findings are very convincing in themselves. Pronounceability rating is a significant predictor of error, and the frequency counts add nothing for the hearing subjects given the first two variables. The finding is as strong for the deaf Ss as for the hearing. The fact that the Mayzner count adds to the prediction slightly for the deaf Ss may mean that their attention is drawn more to regularities in the purely visible orthography, but the effect is too small to warrant much speculation.

Pronounceability is obviously a better predictor than any of the frequency counts. A proponent of frequency might ask, however, what would happen if pronounceability were left out. Would the counts then have any predictive value? Row four in Table 3 shows the results of such an analysis. Length, of course, is still significant, and in this case one of the frequency counts, the Mayzner bigram count, adds significantly to the prediction of errors. It correlates better with pronounceability than the Mayzner trigram count, and the latter does not add significantly to the prediction. This count is a more plausible predictor than sheer bigram frequency, because it takes letter position and word length into account. It is not as good a predictor as pronounceability, however, as the previous analyses show. This is not unreasonable, when one considers the sample used in making the count. Tokens, rather than types, were the sample--all the words three to seven letters in length in a text of 20,000 English words. Since the sample was running text, rather than dictionary entries, a trigram like THE occurred very often (2401) times in the first three positions, while QUI, an acceptable and highly constrained pattern for beginning an English word, had a total count of only 23 in the first three positions--little more than CKI which did not occur, but in fact cannot occur. Thus even this frequency count, though it reflects the rules to some degree, does not do so perfectly. This suggests that the pronounceability measure accounts for most of the variance predicted by the one significant frequency count, and something more.

When both pronounceability and length are omitted, the frequency counts in some combination account for 53 percent of the variance. The significance of bigram frequency as a predictor, in this case, is due to its correlation with length. It does not do as well as length, but is clearly taking up some of the variance that that variable would, if present. The Mayzner bigram count again increases the prediction fairly effectively.

When length alone is omitted, pronounceability is the "lead" predictor with 43 percent of the variance. The two sheer frequency counts add to the prediction, again because they are contributing what length would have, had it been there. The Mayzner counts do not appear. Their potential contribution to the prediction is absorbed by the pronounceability rating.

These analyses seem to indicate that the pronounceability rating is actually measuring something more than sheer pronounceability, something which is reflected, but to a lesser degree, in the Mayzner bigram count, and something which is potentially present in orthography alone, since it facilitates the deaf at least as much as the hearing. It is our opinion that it is the rules in the spelling, not mere sequential probability, since the two ordinary frequency counts are not predicting (when they predict anything) the same thing, as their low correlation with pronounceability shows us.

<u>Prediction of deaf errors by linguistic variables</u>. Scores and ratings on a number of variables having to do with aspects of speech, hearing or reading were furnished us for the deaf subjects. They were: rating for degree of residual hearing; rating for comprehensibility of speech; age of onset of deafness; and scores on a reading test, two verbal aptitude tests, a vocabulary test, and a nonverbal aptitude test. These measures were all correlated with errors and with each other. The correlation matrix is presented in Table 5. The matrix was subjected to

Insert Table 5 about here

stepwise multiple regression analysis with errors as the dependent variable. The analysis revealed that only the hearing rating was a significant predictor of errors, and that only at about the 5 percent level of significance. The fact that hearing level, even very low (and they all were quite low, with loss of high frequencies in the speech range) has any predictive value may seem odd in view of the similar correlations between pronounceability and errors for deaf and hearing subjects. It probably reflects the fact that exposure to education in general is facilitated by even a little hearing, a point reflected by the over-all higher scores for the hearing subjects.

<u>Answers to questionnaires</u>. The answers to questions regarding methods of training in early reading proved to be of little value. Most of the subjects reported that some attempts had been made to give them speech training before or during early reading instruction, but there is no way of knowing how successful this was, or to what extent, if any, it was related to reading instruction. Many subjects simply said they did

not remember. Pictures with words, or words lettered on objects were mentioned in some cases.

An interesting point does come out of the answers to the final, open-ended question, however. The question asked was "Describe as well as you can how you were taught to read." Most of the subjects wrote a short paragraph in answer. In examining these paragraphs, we were struck by the fact that there were practically no spelling errors, but by contrast there were numerous grammatical or morphological peculiarities. Wrong tenses, confusion of singular and plural, and elliptical statements were frequent. Here are some examples:

"After I come back from school, my mother taught me how to read the words in the school books. She used both loud voice and clear lip movement to help me to learn how to read. I was taught to form letter when read."

"Learn to read comes from oraling."

"I was learning to read through writing alphabets on notebook... Later my teacher display several cards with pictures on a board."

It is not our aim here to interpret these errors, but it is notable that they are not similar to telegraphic statements or misusages of a young hearing child. Morphological conventions easily picked up by a hearing child seem often to be lacking. Evidently it is very important to <u>hear</u> speech to learn to use these conventions consistently. But this does not seem to be the case for spelling patterns; they have become effective in the deaf group without their hearing the sounds they map to.

Discussion

Let us reconsider now the five hypotheses presented earlier, in the light of these data. Are any of them confirmed or weakened? Consider first the ones which can reasonably be eliminated.

The hypothesis that a word is matched to an acoustic representation before it is read, and that a familiar sound is facilitating cannot be right. It is obviously impossible for the deaf subjects. Even those with the highest ratings for hearing were unable to discriminate speech sounds.

It is equally unlikely that matching to an "articulatory plan" can adequately explain the difference between the pronounceable and unpronounceable words. Most of the deaf subjects spoke very little and furthermore the speech rating (its comprehensibility and therefore its differentiation) did not predict errors.

What about sequential dependencies? Ordinary summed bigram and trigram frequencies were poor predictors of errors, so sequential probability as such, taking no account of beginnings and endings of words, is inadequate. The Mayzner bigram count, which considers structural features of words such as letter position and length, was better, however, and thus indicates the importance of these features.

We come now to the hypothesis that inspired our first experiment--that spellings which map with invariance to sound become "chunks" or larger units because of the one-to-one mapping rule. With great reluctance, we conclude that this hypothesis is seriously weakened. The fact that the deaf Ss were equally or indeed more facilitated in reading pronounceable spellings must mean that the mapping relation to sound is not essential--or rather, that it is not essential that the reader experience the cross-modal invariance.

In another sense, however, the cross-modal invariance is essential. It is essential in the evolution of written language. Our fifth hypothesis was that pronounceability ratings are measuring orthographic regularity (rules governing the internal structure of English words), and that it is this kind of structure in the pronounceable words that facilitates perception. The words are rated pronounceable, because the writing system--and therefore rules for spelling--evolved in relation to sound. Therefore pseudo-words that follow the rules must map to sound with regularity and must be rated pronounceable by hearing subjects.

Writing is a surrogate for speech; but orthographic rules are rules in their own right, and apparently can be learned as such, quite aside from the fact that any word they produce maps predictably to speech sounds. Sound would seem thus to be not necessarily a part of the individual's processing in forming higher units of reading, although historically it formed them in the spelling patterns of the written language. An intelligent deaf reader does master and use the regular spelling patterns of the language in processing graphic material and is facilitated by their presence. The redundancy contributed by invariant mapping to speech sounds may well make it easier for the hearing child to pick up the common spelling patterns and regularities as he learns to read, but clearly it can be done without this.

We need now to know more about the structural constraints within the words that contribute order and reduce the amount of information to be processed. We need also to find out the best way of learning them. Given a hearing child, will he abstract the common patterns in different words more easily if the redundant common (invariant) sound patterns always accompany them? Or will this added information serve at first to distract him, divide his attention and lengthen the time required for abstracting the spelling patterns and rules?

Probably only research with the child as he is actually learning will help us to answer these questions. In a first onslaught on the problem, we are studying the abstraction of very simple spelling patterns by five and six year old children just prior to their entering first grade. A learning set procedure has been adopted, with a task that combines discrimination and classification. A set of cards each containing a word is given the child, who proceeds to sort them into two piles. In one pile go all the words that contain the "concept," a common cluster such as ST at the beginning of the words. In another go all the "negative instances." As the child proceeds from problem to problem, the abstraction of a new pattern may become easier. In pilot data, we found that in a few children the set built up quickly and led to 100 percent success after a few days of practice. In others, success came hard, if at all. Would these children profit by adding the spoken counterparts, so that they can abstract the common auditory feature? Little is known about the effect of cross-modal redundancy in concept formation, especially in children, but the issue seems to us a critical one.

It is possible that beginning with morphological rules already known to the child, such as formation of plurals, and linking them with the appropriate spelling pattern would lead easily into the set to abstract order within words. How much verbal instruction helps is another question. That is does help in a specific instance of a pattern is true, we have found, but it seems also to hinder transfer in some cases. The question is how to build a general set to abstract the rules, and thereby gain the most powerful aid to transfer in reading new words that follow rules, if we are concerned with building skill in reading by larger units and thus reducing the information load.

Summary

As experiment was reported in which deaf and hearing subjects were compared for the ability to read, under tachistoscopic presentation, letter strings (pseudo-words) which did, or did not, follow rules of orthography which rendered them pronounceable or relatively unpronounceable. Deaf as well as hearing readers were more successful in reading the pronounceable ones. This finding must mean that orthographic rules were used by these subjects even though the invariant sound mapping was not available to them. Research is needed to show the best way to teach or promote induction of spelling patterns in order to promote skill in processing written language in units that reduce the information load.

References

Anisfeld, M. A comment on "The role of grapheme-phoneme correspondence in the perception of words." <u>Amer. J. Psychol.</u>, 1964, 77, 320-321.

Cattell, J. McK. Ueber der Zeit der Erkennung und Benennung von Schriftzeichen, Bildern, und Farben. <u>Philos. Stud.</u>, 1885, 2, 653-650.

Diebold, A. R. A survey of psycholinguistic research, 1954-1964. In <u>Psycholinguistics: a survey of theory and research problems</u>. Edited by C. E. Osgood and T. A. Sebeok. Bloomington, Ind.: Indiana University Press, 1965.

Doehring, D. G. and Rosenstein, J. Visual word recognition by deaf and hearing children. <u>J. Speech and Hearing Research</u>, III, 1960, 320-326.

Gibson, E. J. On the perception of words. <u>Amer. J. Psychol.</u>, 1964, 77, 668-669.

Gibson, E. J., Pick, A., Osser, H. and Hammond, M. The role of grapheme-phoneme correspondence in the perception of words. <u>Amer. J. Psychol.</u>, 1962, 75, 554-570.

Hockett, C. F. Analysis of English spelling. In <u>A basic research program on reading</u>, Cooperative Research Project Number 639, Cornell Univ. and U. S. Office of Education.

Levin, H. and Biemiller, A. J. Studies of oral reading. I. Words vs. pseudo-words. Mimeo paper, 1965.

Liberman, A. M., Cooper, F. S., Harris, K. S. and MacNeilage, P. F. A motor theory of speech perception. In <u>Proc. Speech Commun. Sem.</u>, Stockholm, Royal Inst. Tech., 1963.

Mayzner, M. S. and Tresselt, M. E. Tables of single-letter and digram frequency counts for various word-length and letter-position combinations. Psychon. Monogr. Suppl., 1965, 1, 13-22.

Mayzner, M. S., Tresselt, M. E. and Wolin, B. R. Tables of trigram frequency counts for various word-length and letter-position combinations. Psychon. Monogr. Suppl., 1965, 1, 33-78.

Miller, G. A., Bruner, J. and Postman, L. Familiarity of letter sequences and tachistopscopic identification. J. Gen. Psychol., 1954, 50, 129-139.

Miller, G. A. Free recall of redundant strings of letters. J. Exp. Psychol., 1958, 56, 484-491.

Underwood, B. J. and Schulz, R. W. Meaningfulness and Verbal Learning. New York: J. B. Lippincott Co., 1960.

Venezky, R. L. A computer program for deriving spelling-to-sound correlations. In A basic research program on reading, Cooperative Research Project Number 639, Cornell Univ. and U. S. Office of Education.

Venezky, R. L. and Weir, R. H. A Study of Selected Spelling-to-Sound Correspondence Patterns. Cooperative Research Project N. 3090, Stanford Univ., 1966.

Wallach, M. A. Perceptual recognition of approximations to English in relation to spelling achievement. J. ed. Psychol., 1963, 54, 57-62.

Table 2

Analysis of Graphic Errors

	Deaf		Hearing	
	Pron.	Unpron.	Pron.	Unpron.
M No. same in any order	4.58	4.64	4.45	4.70
Length of correct sequence forward	3.77	3.09	3.52	3.08
Length of correct sequence reverse	.52	.73	.60	.74
M No. same in sequence, beginning of word	2.58	2.50	2.64	2.44
M No. same in sequence, ending of word	1.68	1.00	1.14	1.18

Table 3

Correlations between Errors for Pseudo-Words and Various

Predictor Variables (Hearing and Deaf Ss)

	Length	B.F.	T.F.	Pron. Rating	Hearing Errors	Deaf Errors	M.B.F.	M.T.F.	Combined Errors
Length	1.0	.64	.21	.35	.85	.82	-.02	.07	.84
Bigram Frequency		1.0	.73	-.03	.41	.41	.31	.55	.42
Trigram Frequency			1.0	-.14	.02	.06	.28	.75	.05
Prounounceability Rating				1.0	.61	.68	-.63	-.45	.66
Hearing Errors					1.0	.93	-.29	-.16	.98
Deaf Errors						1.0	-.37	-.21	.98
Mayzner Bigram Frequency							1.0	.65	-.33
Mayzner Trigram Frequency								1.0	-.19
Combined Errors									1.0

Table 4

Stepwise Regression Analyses for Predictor Variables

Dependent Variable	Predictor variables signigicant at .05 level or better and percent of variance accounted for (cumulative)			
Hearing Errors	Length 73%	Pronounceability 84%	All Others NS	
Deaf Errors	Length 67%	Pronounceability 84%	Mayzner bigram 85%	All others NS
Hearing and Deaf Errors Combined	Length 72%	Pronounceability 87%	Mayzner bigram 88%	All others NS
Hearing and Deaf Errors Combined	Length 72%	Mayzner bigram 83%	All others NS	
	Pronounceability omitted			
Hearing and Deaf Errors Combined	Bigram F 17%	Mayzner bigram 48%	Trigram F 53%	Others NS
	Pronounceability and length omitted			
Hearing and Deaf Errors Combined	Pronounceability 43%	Bigram F 63%	Trigram F 69%	Others NS
	Length omitted			

Table 5

Correlations between Errors for Pseudo-Words and Various

Predictor Variables (Deaf Subjects Only)

	Errors	Hearing	Speech	Onset	SCAT Verbal	STEP Reading	SRA Verbal	Vocab.	SRA Nonverbal
Errors	1.00	.44	.32	-.24	-.04	-.31	-.37	-.25	.23
Hearing Rating		1.00	.53	-.31	.09	-.16	-.23	.04	.04
Speech Rating			1.00	-.43	-.22	-.54	-.23	-.23	-.01
Age of Onset				1.00	.05	.21	.05	-.10	-.17
SCAT Verbal					1.00	.32	.27	.54	.08
STEP Reading						1.00	.39	.70	.15
SRA Verbal							1.00	.54	-.24
Vocabulary								1.00	.04
SRA Nonverbal									1.00

The Discrimination of Graphic Characters: The Confusability of Hindi Letters by Non-Hindi Readers

Arthur Shurcliff

It is generally agreed among educational psychologists that the ability to discriminate letters is basic to the process of learning to read. What is not so unanimously agreed on is the question of the process or processes underlying this discrimination.

A distinctive feature hypothesis was developed and tested by Gibson, Osser, Schiff, and Smith (1963). They first drew up a list of features for Roman capital letters, the list being subject to the following criteria: "(1) Features must be critical ones, present in some members of the set but not in others so as to present contrast; (2) They must be invariant under perspective and size transformations; (3) They must yield a unique pattern for each grapheme; (4) The list must be reasonably economical. If the list contained 26 features, we should have done no better than a template matching scheme." They then obtained a confusion matrix for the 26 letters, using as subjects four-year-old children. They found a small, but significant correlation between number of common features and confusions. That is, the greater the number of features a pair of letters had in common, the more likely they were to be confused. Further support for the distinctive feature theory of letter discrimination comes from a study using a search task (Gibson & Yonas, 1966) and from confusions made by adult S's (Gibson, Yonas, & Schapiro, in preparation).

An important question to be asked at this point is whether a similar economical list of distinctive features could be found that would uniquely specify each letter and predict confusions for another alphabet. Gibson's list of distinctive features was drawn up expressly for Roman capital letters, and it is conceivable that no such economical list of features would work for another alphabet. The present study was an attempt to answer this question using Hindi letters. A second question, assuming an economical list of features could be found that correlated with confusions, was how closely such a list would resemble the list generated for Roman letters.

Procedure

Subjects were 35 female high school students who had just finished their junior year, and were at Cornell University for a summer advanced placement program.

Materials consisted of slides of Hindi letters, scoring sheets with the same letters, and a projector with a tachistoscopic attachment. Both the slides and the scoring booklets were made by means of a standard Hindi typewriter. The slides were constructed by typing a single, lower case Hindi letter in the center of a Radio-Mat slide. A total of 30 different letters were used, and consequently there were 30 different slides. Each scoring booklet consisted of thirty sheets, with each sheet having all 30 letters printed on it in five rows, six letters to a row. The order of letters was counterbalanced for the different sheets, so that each letter appeared equally often on the top and bottom rows, and at the beginning and end of each row. It was discovered later that one letter was put into the slide tray backwards, and so data for that letter had to be discarded. The projector was a standard Carousel slide projector with a tachistoscopic attachment.

Subjects were run in groups, and were given the following instructions: "I am going to flash a Hindi letter onto the screen for a very brief interval. For the first letter, I would like you to turn to the first page in your scoring booklet. When you see the letter flashed on the screen, I would like you to circle the letter you thought it was. Since each page of the scoring booklet has all 30 letters on it, the letter on the screen will be one of these. After each slide, please turn to a new page, so that you are circling one letter per page. I will say, 'Ready, go' before each slide flashes, to warn you." All thirty letters were then flashed on the screen in succession, following the procedure outlined in the instructions. Each letter was exposed for 20 milliseconds. Then subjects were handed a new book, and the process was repeated, with the slides in reverse order. This order of presentation was counterbalanced by reversing it for half the subjects.

Results

The data were first collated in the form of two confusion matrices, one for the first run and one for the second. The data were then condensed in two ways. First, since both matrices seemed to have roughly similar distributions of confusions, they were combined. Second, since both halves of the resulting matrix appeared similar, it was combined across its diagonal. The result is shown in Figure 1.

The data were then analyzed by a computer program based on a hierarchical clustering scheme developed by S. C. Johnson at Bell Telephone Laboratories, Murray Hill, N. J. Since the report of this program has never been published, a brief word of explanation is in order. A hierarchical clustering scheme is one having a number of levels, or hierarchies. At the bottom level, all the objects are grouped together as a single cluster. At progressively higher levels, each cluster becomes more and more exclusive, until finally at the highest level each letter is a separate cluster. The output from the clustering scheme is presented in Figure 2. The higher the column linking any group of letters, then, the greater the confusability of those two letters.

With the aid of this clustering scheme and the list of distinctive features drawn up by Gibson, Osser, Schiff, and Smith, the following set of

	1 प	2 म	3 य	4 स	5 घ	6 भ	7 न	8 च	9 ज	10 ज्ञ	11 त	12 त्र	13 व	14 ब	15 क	16 ग	17 थ	18 द	19 ह	20 र	21 ट	22 उ	23 ल	24 ठ	25 ष	26 ढ	27 ख	28 श	29 अ
म	14																												
य	6	4																											
स	0	1	3																										
घ	3	0	8	2																									
भ	4	14	4	3	2																								
न	1	1	1	1	0	0																							
च	1	1	1	7	3	1	3																						
ज	1	1	2	2	2	4	3	4																					
ज्ञ	0	0	2	4	0	0	0	2	15																				
त	1	0	5	13	0	1	6	3	7	2																			
त्र	0	0	0	0	2	0	3	3	1	4	0																		
व	2	0	2	1	2	1	0	7	0	0	2	0																	
ब	0	0	1	0	4	0	0	5	3	1	0	1	14																
क	0	1	0	0	1	1	0	4	0	0	0	1	1	3															
ग	0	0	2	0	0	2	4	1	2	0	6	0	1	0	0														
थ	0	0	1	0	3	7	0	0	2	0	0	0	0	1	0	0													
द	2	0	0	2	1	0	0	3	0	2	1	0	1	0	0	0	0												
ह	0	0	0	2	0	0	0	2	6	0	0	0	1	0	2	0	0	14											
र	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0										
ट	0	0	0	1	0	0	0	0	0	1	3	1	0	0	0	0	0	4	0	0									
उ	0	0	0	0	0	0	0	1	5	1	0	0	1	0	1	1	0	1	1	33	3								
ल	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0							
ठ	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	3	0	0						
ष	0	0	1	5	0	0	0	0	0	1	0	0	3	0	1	0	0	5	4	2	0	1	0	0					
ढ	0	0	0	1	1	0	0	0	0	0	0	0	0	4	1	0	0	4	6	0	2	0	0	3	0				
ख	0	0	0	8	1	0	0	2	2	2	0	1	0	0	2	0	0	1	2	0	0	0	0	0	3	0			
श	0	0	0	6	1	1	0	0	3	10	2	0	0	2	2	3	4	5	1	0	0	1	0	0	24	0	8		
अ	0	0	0	4	2	0	0	3	12	1	11	2	4	1	3	0	4	1	2	0	0	6	0	0	3	3	5	3	

Figure 1. Confusion matrix obtained from summing both runs and collapsing along the diagonal.

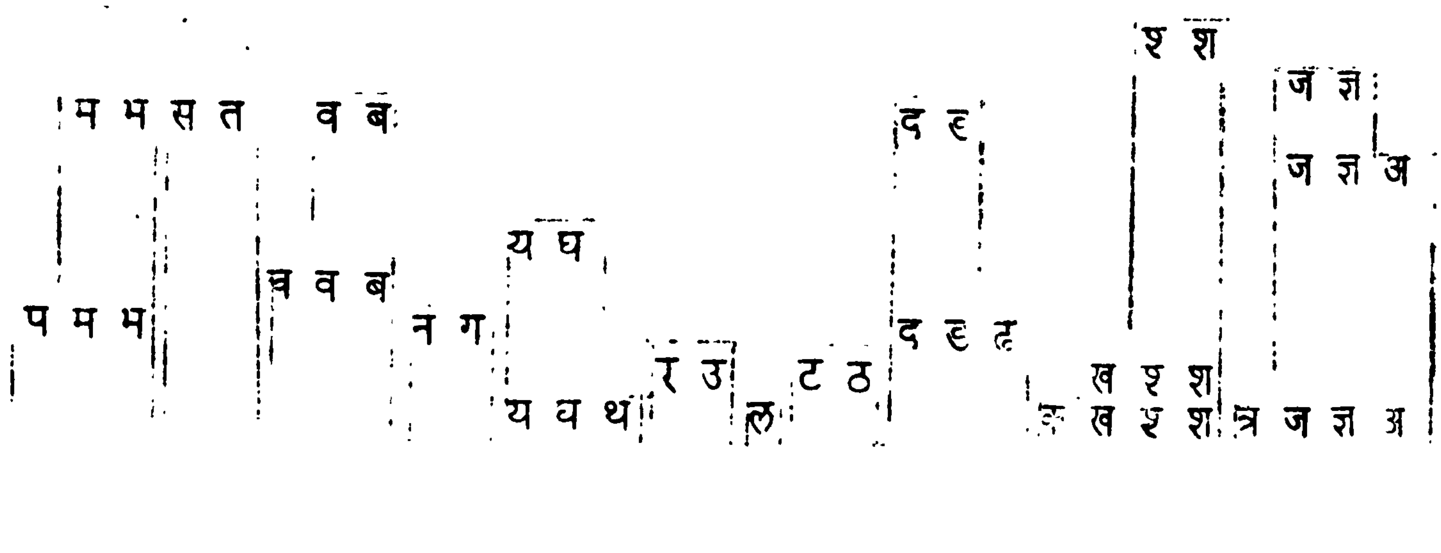


Figure 2. Hierarchical cluster analysis obtained from the confusion matrix.

distinctive features were drawn up.

Feature List

Standard Vertical

Straight line
 Horizontal
 Vertical
 Diagonal

Curve
 Concave left
 Concave right
 Concave up
 Concave down

Closure
 Major closure
 Minor closure

Number of segments
 1
 2
 3
 4
 5

The "standard vertical" refers to the vertical extending down from the horizontal line on top of the letter, present on most of the letters. The top horizontal line was ignored for the purpose of feature analysis, since it was common to all letters and hence was not distinctive. A "major closure" was defined as one formed against the top horizontal.

Next, the distinctive features of each letter were plotted in the following way; the feature list was drawn on the left side of a large sheet of lined paper, and the letters were placed along the top. Each letter was given a plus in the appropriate row below it if it had the feature listed in that row. The superordinate features (Straight line, Curve, Closure) were included in this analysis, because it was intuitively felt that two letters having, for instance, a vertical and a horizontal line respectively should be credited with having the feature "straight line" in common. When the first version of the list was plotted in this manner, it was found that not all the letters were specified uniquely; after minor adjustments in the feature list, resulting in the list shown above, all the letters were specified uniquely.

The next question to be asked was, how well would this list of distinctive features account for the number of confusions between any two letters. The underlying assumption was that if letters are discriminated by searching for distinctive features, then the more features in common between two letters, the greater should be the confusions. Thus, a matrix was drawn up representing the number of features in common between any two letters. By comparing this matrix with the confusion matrix, the number of confusions was plotted against the number of common features in the matrix shown in Table 1. If the distinctive feature list predicted perfectly, one

Table 1. The relation between number of common features and confusions represented as a frequency distribution. Each cell represents the frequency of occurrence of a given number of confusions for a class of item pairs having the specified number of common features.

Number of Common Features	Number of Confusions 0	1	2	3	4	5	6	7	8 and over	% item pairs with one or more confusions
0	15	1	6	2	1		1	1		6.25
1	64	14	10	8	6	2	3	1	3	28.09
2	77	18	9	7	5	1	1		2	39.84
3	42	19	9	5	5	1		2	4	51.12
4	19	9	2	2	1	1	1		1	64.81
5	6	4	1	1		1	1		3	66.67
6	3		1							70.00
7										

would expect to see only the cells running along the diagonal from the upper left to the lower right filled (i.e., the more the number of common features, the more the number of confusions). If the distinctive feature list accounted for none of the variance, one would expect to find cells filled randomly, predominantly in the upper left of the table, reflecting merely the frequency distributions of confusions and common features. A close inspection of the table will show that the result lies somewhere between these two extremes; those letters having five or six common features have a wide range of confusions, while those letters having 0 or 1 common features have predominantly 0, 1, or 2 confusions.

The values of many of the cells were so low that it was impossible to try a trend test. But by considering the percentage of cells having confusions for each row, the problem of sparse cells and uneven distributions was circumvented. For each row (i.e., for each number of common features) the following fraction was obtained; the total of the frequencies for one or more confusions/total of the frequencies. Thus, for example, 6.25% is 1/15+1, or 1/16. This fraction, expressed in the form of a percentage, is shown in the right-hand column of Table 1. As is apparent, this confusion percentage rises dramatically as the number of features in common increases. The probability of this rank order occurring by chance is well below .01.

As a further test of the ability of the feature list to account for confusions, the total number of confusions was calculated for each row.

This quantity was divided by the total number of confusions possible (that is, for a given pair this would be the number of confusions that would result if every subject confused the pair whenever possible). This relation is plotted in Figure 3. The unusually large number of confusions for letters with six features in common is mostly the contribution of letters 25 and 28, which were confused 24 times (the next largest total was 15). Despite the irregular shape of the curve, the confusion ratio shows a clear positive relation to the number of common features.

Discussion

The question asked in the introduction has been answered in the affirmative; an economical list of features was derived for a non-Roman alphabet that was both able to specify each letter uniquely and able to account for a large proportion of the confusions. The second question can also be answered in the affirmative; the list closely resembles the one drawn up by Gibson, Osser, Schiff, and Smith for Roman capitals.

It is interesting to notice some of the differences between the lists. The present list of features for Hindi letters failed to include Gibson's features of Section (Bisects, Intersects) and Symmetry (included in an alternate list), the reason being simply that Hindi letters lack these features. Alternatively, the Hindi list had more features relating to curved lines than did the other list (Concave left, right, up and down). It should be noted that uniform presence of a property, as well as uniform absence, is sufficient grounds for excluding it from the feature list, since such a feature yields no information useful for discrimination. A case in point is the top horizontal bar on all Hindi letters.

One of the weaknesses of the present list of distinctive features for Hindi letters is that it fails to take into account position. For example, the feature list predicted that letter 17 would be highly confusable with letters 24, 25, 26, and 27, since it had four features in common with each of them (see Figure 1 for the numbering scheme). In fact, there were no confusions with these letters. It would seem likely that this is due to the difference in relationship of the common features to other features; as an example, they all share minor closure, but letter 17 has it in a prominent and unusual position. Relations between features are thus important, but we have not yet succeeded in handling them in analyzing an entire alphabet.

The greatest problem in an analysis by distinctive features such as this is that the selection of features is arbitrary, and thus it is always possible that a different list could predict or account for confusions as well or better. In both the present study and the Gibson, Osser, Schiff, and Smith study the lists did considerably better than chance in accounting for confusions, but left a large amount of the variance unaccounted for. In addition to a better selection of feature lists, it seems likely that greater predictive power could be obtained by differential weighting of the features; intuition suggests that certain features are of far more importance in making discriminations than others. Tikofsky and McInish (1968)

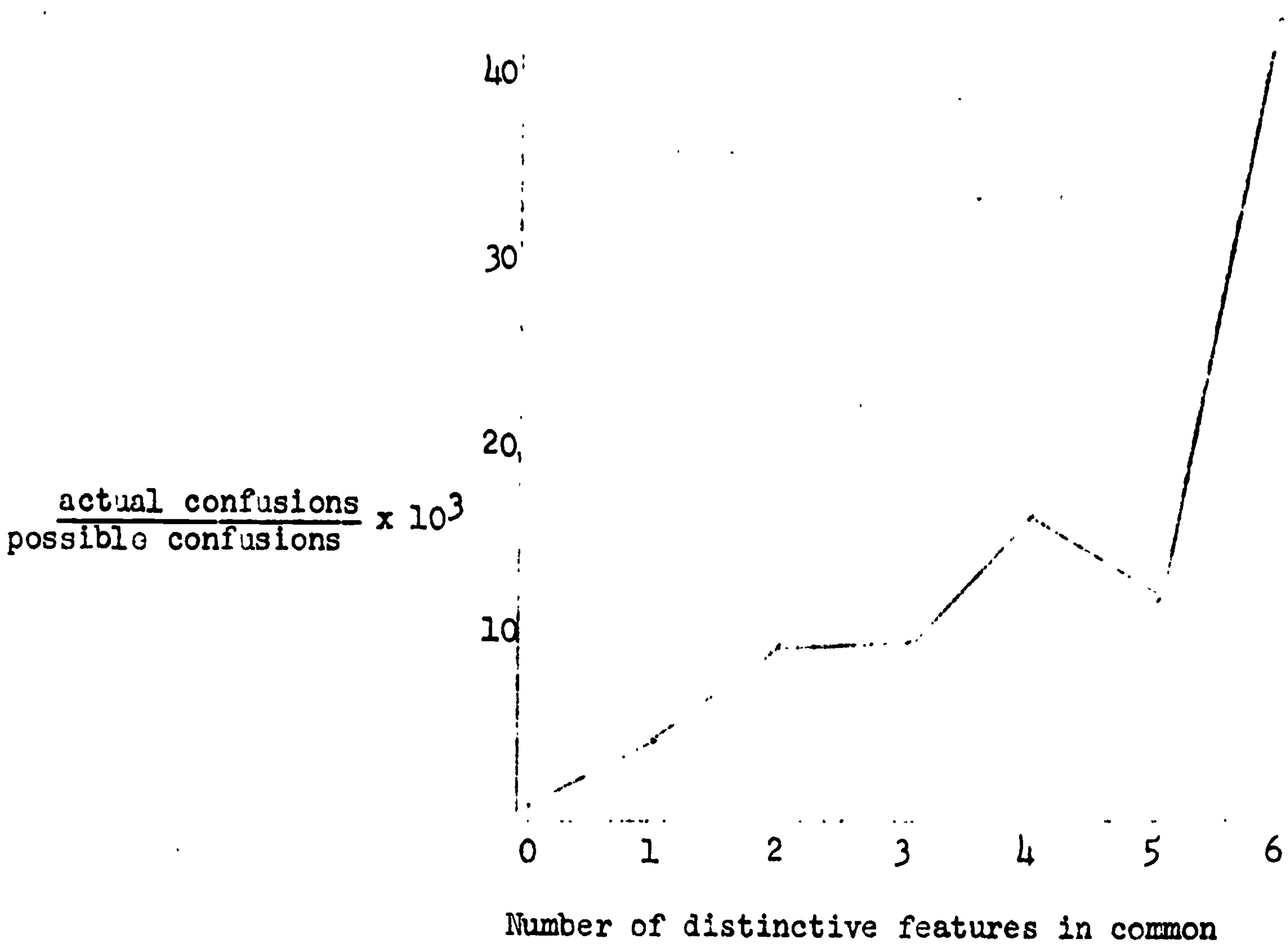


Figure 3. Total confusion ratio as a function of shared features.

found such a weighting scheme useful in predicting auditory discrimination of consonants by seven-year olds.

The present study, if nothing else, gives further evidence that the distinctive feature concept is useful in explaining the process of distinctive letters, be they Roman or Hindi. It would be interesting to see what feature lists are important for the discrimination of Chinese characters, Arabic letters, or Sinhalese letters. This might throw light on the question of whether there is a single set of useful features used by all alphabets, or whether there is a potentially enormous set, with each alphabet selecting only a small subset.

References

Gibson, E. J., Osser, H., Schiff, W., & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In _A basic research program on reading_. Cooperative Research Project No. 639, U. S. Office of Education.

Gibson, E. J., & Yonas, A. A developmental study of visual search behavior. _Percept. and Psychophys._, 1966, _1_, 169-171.

Gibson, E. J., Yonas, A., & Schapiro, F. Latency measures of confusions between letters. In preparation.

Tikofsky, R. S., & McInish, J. R. Consonant discrimination by seven year olds: A pilot study. _Psychon. Sci._, 1968, _10_, 61-62.

A DEVELOPMENTAL STUDY OF FEATURE-PROCESSING STRATEGIES IN LETTER DISCRIMINATION

Albert Yonas and Eleanor J. Gibson

Cornell University

The question we were concerned with was whether people can change their perceptual processing strategies to include tests for the presence of only those stimulus features necessary for the task at hand. Eleanor Gibson has proposed a theory of perceptual learning which hypothesizes that differentiation requires the search for, and processing of, distinctive features of stimulus displays. These are analogous to Jakobson and Halle's distinctive features of phonemes. Furthermore, perceptual processing, given practice in a discrimination task, would be expected to progress toward strategies which use the most economical feature list. We have tried to demonstrate that such perceptual learning does occur when the task presented to the subject makes such a change adaptive.

The method was to set up an experimental situation where it would be possible, given practice, for the subject to differentiate the displays presented on the basis of a single distinctive feature, as contrasted with a control condition where an equally economical search would not be possible. If performance in the experimental and control conditions is initially the same but learning curves show asymptotic performance to be lower in the experimental condition, we infer that the perceptual process has taken advantage of the potential strategy and narrowed the search to the single feature, thus reducing the information processed before the decision is made. We asked whether such learning occurs in the course of practice and whether the ability to use such optimal strategies improves with age. We therefore compared the performances of second, and sixth grade children with that of college sophomores over 135 trials of practice.

METHOD

We chose a disjunctive reaction time procedure (similar to one used by Sterinburg) in which the subject is told to look for one or more letters (which we will call the positive set). When the single letter presented on a screen before the subject is a member of the positive set, he is required to respond "yes" by moving a lever, as quickly as possible, to one side. If it is not a member of the positive set, he responds "no" by pushing the lever in the opposite direction. The stimuli consisted of nine simplified Roman capital letters--A, O, F, N, V, E, C, H, B. When the subject pressed a button, a tachistoscopic shutter on the slide projector was activated to produce the display, which remained on the screen for 20 milliseconds. At the same time, a Hunter millisecond clock was started and it stopped when the subject pushed the lever.

There were three conditions in the experiment. In the first condition, the positive set contained only one letter (E); the negative set contained the other eight (A, O, F, N, V, C, H, B). In the second condition, the positive set contained three letters, A, O, and F, while the negative set contained the other six (N, V, E, C, H, B). In the third condition, the positive set also contained three letters (A, N, V), the negative set the other six (O, F, E, C, H, B). The division of the letters into the two sets in this last condition is such that processing a single feature, in this case, diagonality, would be sufficient for a decision. That is, a single feature's presence or absence differentiates the negative from the positive set. This is not the case in the AOF condition.

Every subject took part in all three conditions of the experiment. At the beginning of a condition, he was told the positive set and, as an aid to memory, the lever background was labelled with the target letters. Order of conditions and the direction of a positive lever response were counterbalanced over subjects.

The subject was given 135 trials for each condition, each letter appearing approximately 15 times in a random order. Response bias thus favored the negative response, but this was equal for the three conditions. Reaction time and errors were recorded.

For each condition, a subject's reaction times were divided into five blocks of 27 trials, and a mean obtained for each block so that initial performance could be compared with subsequent performance. We expected that response time at the beginning of practice would be roughly equivalent for the two conditions containing three members in the positive set, but that with practice they should diverge. The ANV condition, with a single differentiating feature should decrease in latency so as to approach the single target condition. The AOF condition, without any single differentiating feature, should benefit less from practice.

Fig. 1 shows curves for the three conditions, the five blocks of trials on the abscissa and the mean latency on the ordinate. All three age groups have been combined. As you can see, the two conditions with three letters in the positive set are equal in latency in the first trial block. Both curves descend with practice, but the ANV conditions shows a greater decrease in latency, approaching the curve for the single target condition, which is lower throughout.

We expected that the single target condition would be lower, since differentiation of one letter from a set, rather than three, should reduce the number of features that must be processed.

Fig. 2 shows reaction times for the three conditions plotted by blocks of 27 trials, with the three age groups graphed separately. The relationship between condition and practice is similar for each age group, but latency decreases enormously with age.

A mixed analysis of variance was run with subjects-within-grade, grade, condition, and practice as factors. The means of the 27 trials for each subject within each block were used as the data for the analysis. The main effect of conditions was significant at less than the .001 level; the interaction between practice and conditions was also significant at the same level. The difference between the two three-target conditions on the fifth block of trials was significant at less than the .01 level by a Tukey test.

Grade and condition did not yield a very significant interpretable interaction. Fig. 3 shows age curves for the three conditions on the last block of 27 trials. The youngest children, despite the fact that their responses are much slower, show the same greater decrease in latency for the ANV condition as do adults. If the younger children are actually not as adept at switching to an economical processing strategy as adults, perhaps this is made up for by the advantage of their not having to overcome long-practiced habits of exhaustive processing of a complete feature list.

The predicted difference between the two three-target conditions, although statistically significant, is not very large (29 milliseconds). This seemed reasonable to us since the amount of practice given was not enough to bring the subject anywhere near asymptote, and it was further counteracted by years of overlearned processing habits for letter discriminiation. We decided, therefore, to run a few subjects on all three conditions until they reached asymptote. Three subjects were started and given 405 trials per day, approximately 40 minutes. One proved to be irregular in attendance and the second showed erratic curves relating principally to vacation times and illness. However, the third stayed with the practice for 34 days.

Figure 4.

The curves show an asymptotic trend and, most interestingly, the condition with a single differentiating feature, ANV, appears to have dropped to the level of the single target condition. Unfortunately, this is only the data of a single subject and the day-to-day variability is great. It seems clear that this sort of perceptual learning is a slow process. With letters, it has been going on for a long time and a shift to a new task, where some distinctive features can be disregarded, cannot be expected to bring instant changes, however adaptive for that task.

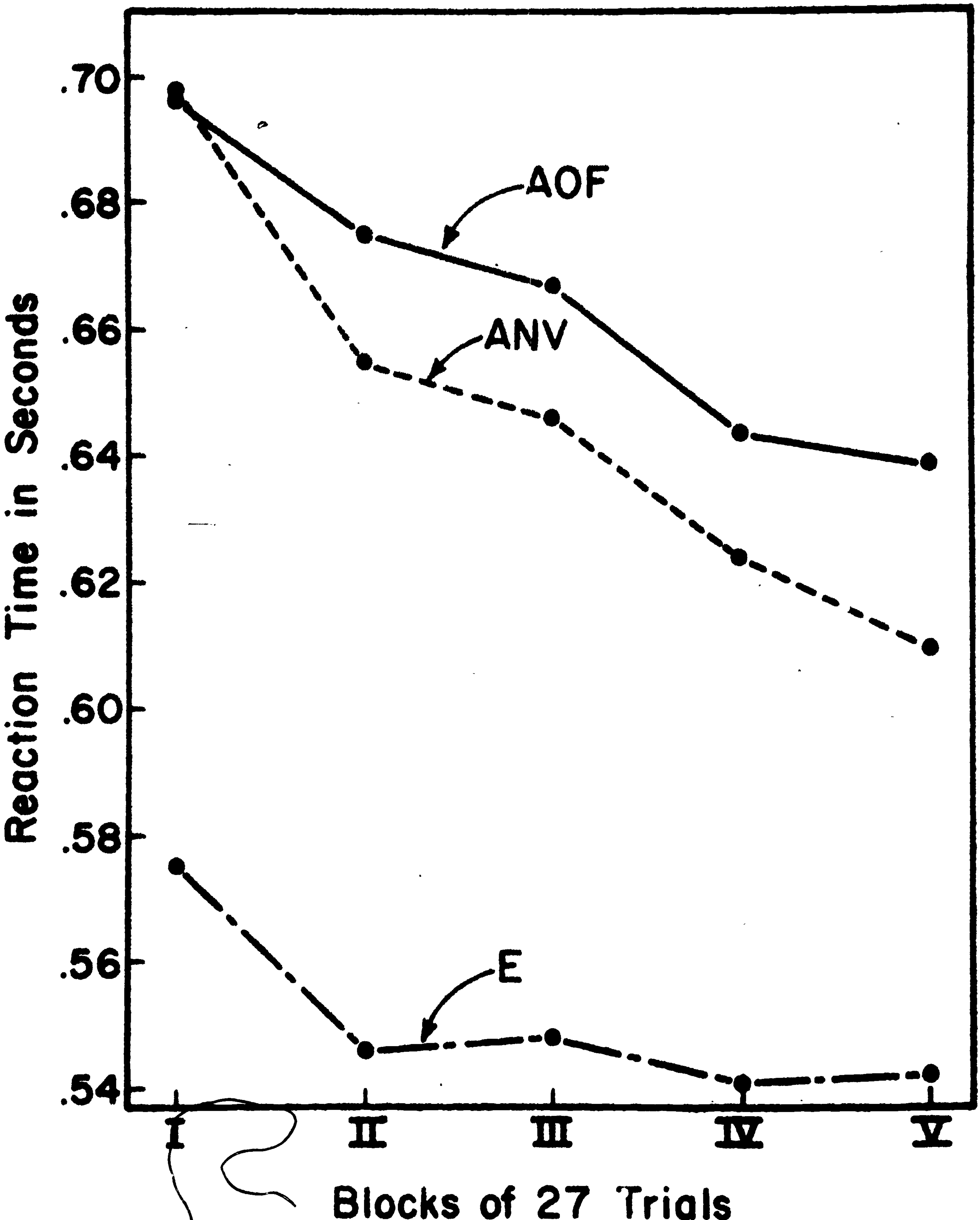
It would be wise, of course, to avoid the difficulty of requiring that learning in an experiment go against long-standing habits. This is not easy; such features as curves and diagonal lines that are critical for alpha-numeric characters will form part of the subject's potential vocabulary even for artificial characters. Nevertheless, we are preparing to repeat this experiment using made-up forms rather than letters, and this may give us a more easily manipulated learning situation.

In conclusion, I feel we have demonstrated that perceptual learning does occur. And, although motor skill does improve in the present experiment, it cannot account for the differential rate of improvement in our main experimental conditions.

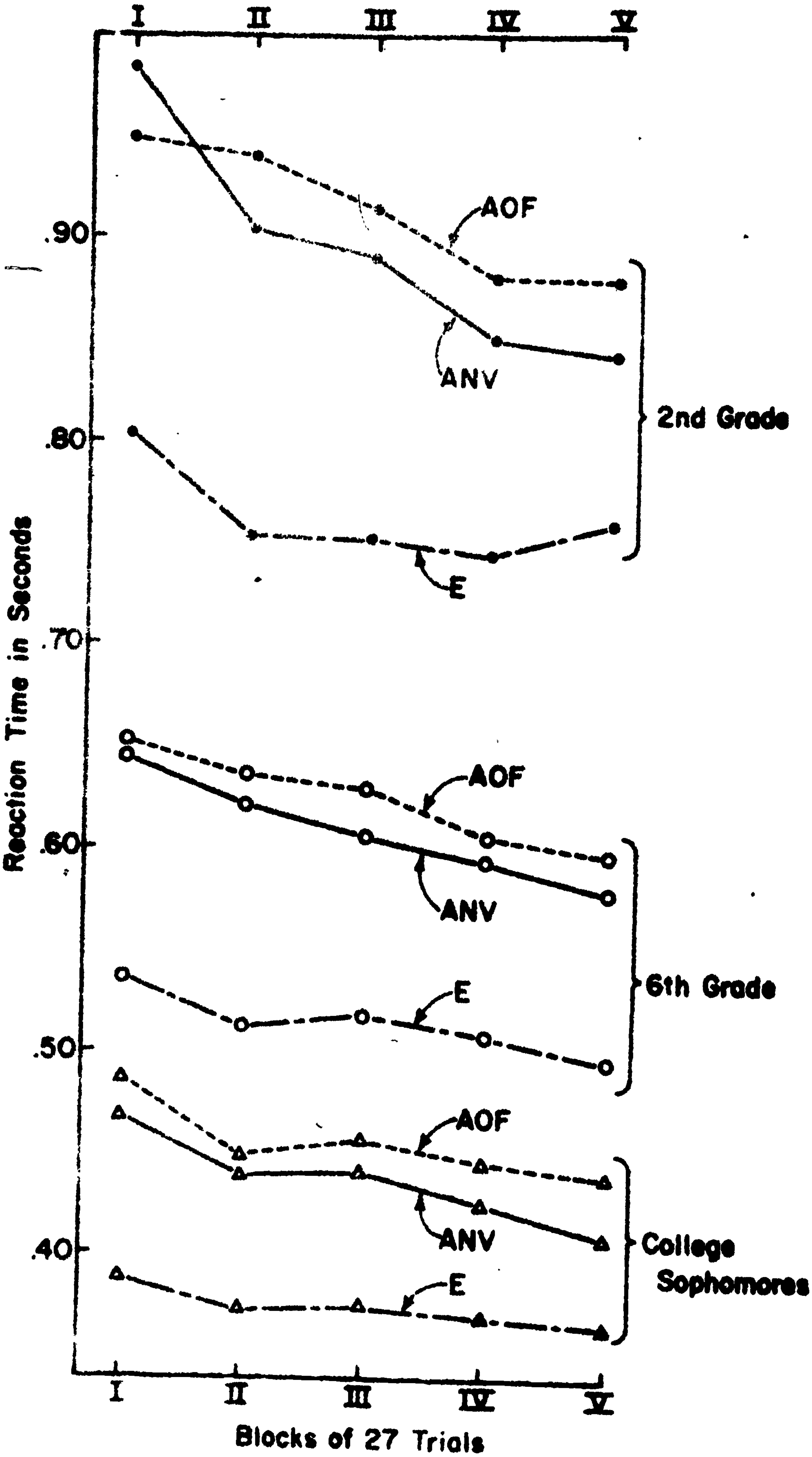
On the other hand, if we can rule out improvement in motor skill, is it possible to locate the learning in some more intellectualized process than perception, some sort of cognitive insight or deliberate instruction of the "intelligence?" I do not think so, because we found that most subjects could not tell us any single differentiating feature in the relevant condition. Also, the improvement was a gradual thing, far from looking like "insight."

We believe, therefore, that perceptual learning <u>occurs</u>--and it is an adaptive, self-regulating change in the direction of reducing the information to be processed. At this point, the question of where and how these changes in processing strategy occur is unanswered.

Figure 1



.70
.68
.66
.64
.62
.60
.58
.56
.54
Reaction Time in Seconds
AOF
ANV
E
I
II
III
IV
V
Blocks of 27 Trials



I
II
III
IV
V
.90
.80
.70
.60
.50
.40
Reaction Time in Seconds
AOF
ANV
E
2nd Grade
AOF
ANV
E
6th Grade
AOF
ANV
E
College
Sophomores
Blocks of 27 Trials

Figure 5

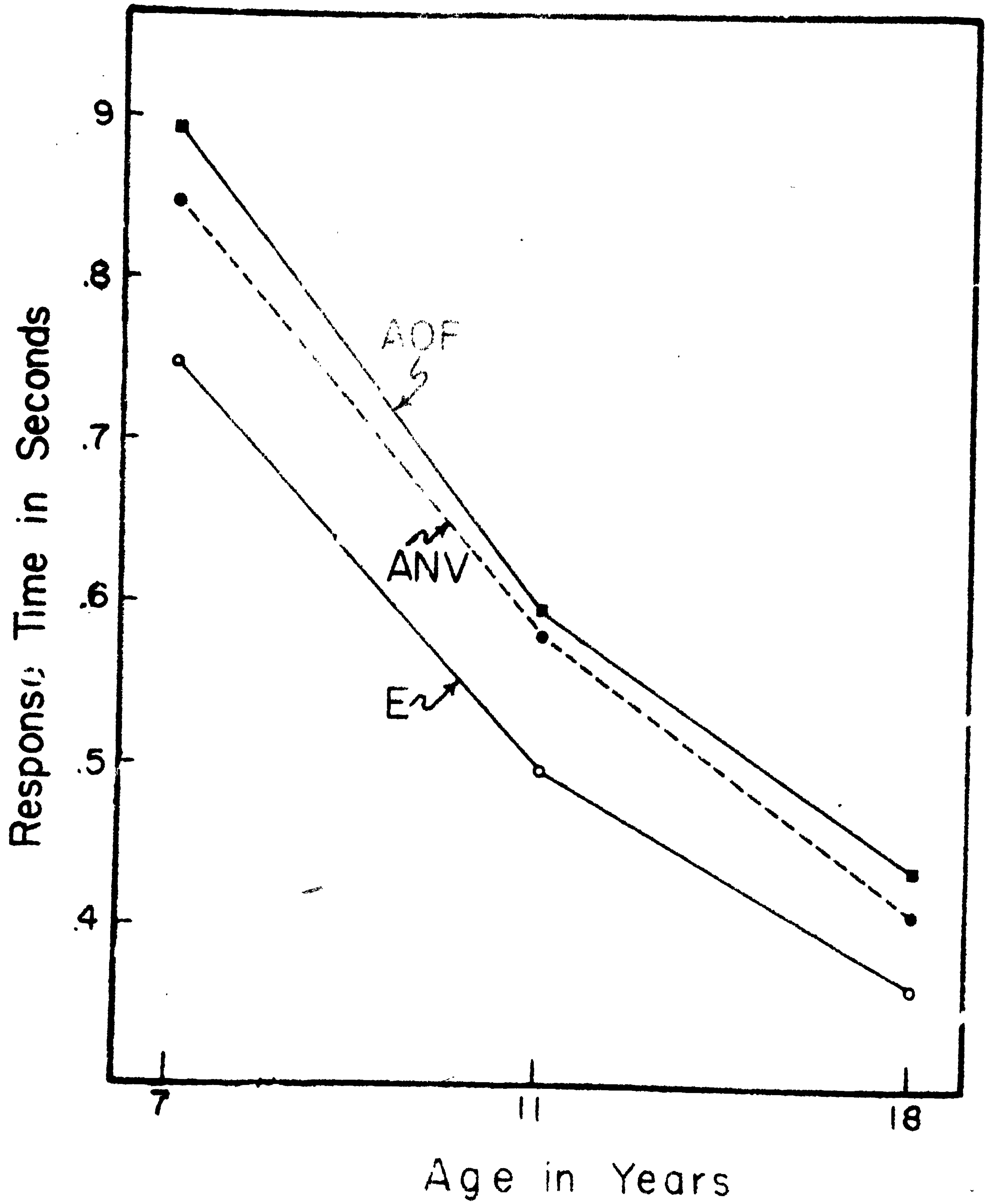
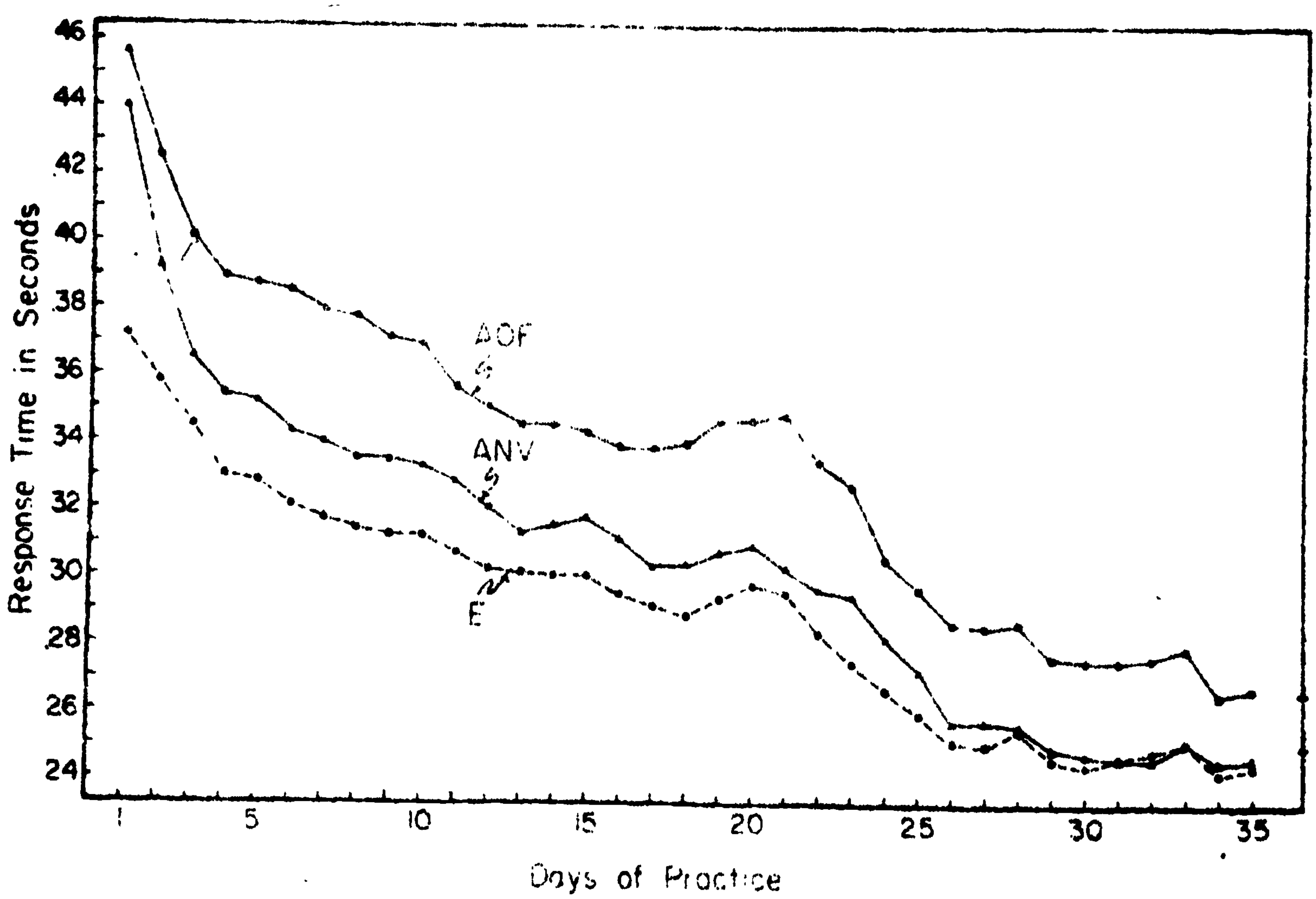


Figure 4


46
44
42
40
38
36
34
32
30
28
26
24
Response Time in Seconds
AOF
ANV
E
1
5
10
15
20
25
30
35
Days of Practice

TEST OF A LEARNING SET PROCEDURE FOR THE

ABSTRACTION OF SPELLING PATTERNS

Eleanor J. Gibson, James Farber, and Sharon Shepela

Cornell University

Since Cattell's research in 1885, psychologists have known that a skilled reader processes graphic material in "chunks" larger than the single letter. With the advent of information theory, the notions of "approximations to English" and sequential probabilities within words were added to "meaning" and "familiarity" as useful concepts for understanding facilitation of processing graphic information. With the advent of psycholinguistics, the concept of structure in the system began to be considered; are there rules for how letters can be composed to form a word, beyond the letter to letter probabilities? And are there, in addition, rules for mapping sequences of letters to sound which might aid in forming units for reading?

Fries exploited the concept of spelling patterns in English orthography in his book on Linguistics and Reading. Hockett and Venezky both worked on rules for spelling to sound mapping. A number of experiments on non-word letter-strings have provided evidence that various forms of redundancy do seem to operate in reading under tachistoscopic conditions or in recall (cf., for instance, George Miller's "Recall of Redundant Strings of Letters").

Tachistoscopic experiments in the Cornell laboratory investigated a hypothesis derived from the work of Hockett and Venezky that spelling-to-sound correspondences facilitate unit formation, by comparing pseudo-words which were compatible with these correspondences (called "pronounceable") with ones which were not. Many replications have confirmed the finding that "pronounceable" combinations are read better than unpronounceable ones. Facilitation begins quite early, at least by the end of third grade. In an effort to analyze further exactly what factors are operating, we recently ran such an experiment with congenitally deaf college students and found, somewhat to our surprise, that although they read fewer words than comparable hearing subjects, the words which followed what we thought of as the rules still had an advantage over the others.

This finding seemed to indicate that spelling patterns have a structure in their own right, which apparently can be learned quite apart from the fact that every word they produce maps predictably into the spoken language. Sound thus seems not to be necessary for the processing of higher order units in reading. The normal skilled reader, however, probably uses not only the structure inherent in spelling patterns but also his knowledge of the rules of the spoken language, and the higher order relations between these systems. He may well process written material on several levels. It seems to us that some knowledge of how the graphic information itself is utilized is preliminary to our understanding of these more complex processes.

It is necessary, therefore, to be more precise in specifying the kinds of order or constraint which can be picked up by a reader and which may provide structure for unit-processing. As an example of orthographic rules, there are many constraints on letter-clusters specifying which may begin a word, or which may end it (FR is a high frequency digraph which may begin a word but not end it; CT is a frequent one which may end but not begin). Sound correspondences are systematically related to these, but do not provide the sole constraints. They are redundant.

Because the spelling patterns are rule-like, it seemed to us that a child who does more than memorize words and their sounds as paired-associates must, as he learns, abstract them from multiple contexts. How he does this is a matter of pressing concern. The research upon which we are engaged is a first attempt to study the abstraction of spelling patterns in very early stages of learning to read.

A method which seemed particularly suited to the kinds of questions we wanted to investigate was the discrimination learning-set paradigm. We wanted to know whether, when a subject learns to discriminate members of a set, he makes use of the structural constraints within the stimulus set. Given a set of structured stimuli, (1) can a subject learn to discriminate among the stimuli on the basis of this structure, and (2) does a subject learn to look for such structure in new problems involving similar type of material?

In a learning set experiment, these questions correspond to (1) whether there is learning within problems that transfers to novel problems involving the same structure, and (2) whether, across problems, there is learning of how to solve discriminations involving patterns of different structure.

In addition to providing an opportunity for abstraction of simple patterns, we wanted to facilitate their discovery so that the abstraction would occur sooner as occasions recurred and, by transfer of the habit of looking for simple patterns, make discovery of more complex patterns easier. Utilizing constraints and redundancies, and transferring the practice is a sophisticated form of behavior. Could we get children as young as five and six doing this? If we hoped to have a procedure for studying the role of added redundancies, and for finding the kinds of pattern that are easy or hard. We hoped also that we could eventually study the role of negative instances in abstraction of spelling patterns. The set to abstract them seemed essential for this task, and so we began with the learning set paradigm. We tackled orthographic patterns alone, without sound redundancies or constraints, aiming to investigate the potential contribution of the latter as a second stage of the research.

Procedure

Subjects were kindergarten and first grade students.[1] For stimuli, we devised a series of discrimination problems, each one consisting of four pairs of four-letter words, half of which (the positive set) contained a common two-letter "pattern." The possible patterns corresponded to initial and final consonant clusters, vowel clusters, or disjunctive "final e" units. A negative instance differed from the corresponding positive in one of the letters in the pattern only. The negative set itself contained no pattern, since research in concept identification (so far) suggests that a constraint within the negative instances makes the concept harder to attain. Four of these pairs, repeated four times according to a Gellerman series, comprised a set. A typical set was:

positive	negative
LACK	LAKC
MUCK	MUCH
DECK	DERK
SOCK	SOAK

Over five days, twenty-four such sets were presented to each S. A summary, by common pattern and day, is given below. Day 1 began with three simple three-letter problems for practice.

Day 1	Day 2	Day 3	Day 4	Day 5
SP_ _	_OO_	TH_ _	_AI_	PL_ _
_ _LL	WH_ _	_ _CK	CH_ _	_ _NG
_EE_	_EA_	_OA_	PR_ _	SH_ _
_U_E	_ _SH	_ _SS	_ _SK	_ER_
	_A_E	_I_E	_O_E	_E_E

1/We wish to thank Mr. F. Ballard, Supervising Principal of the Dryden, N.Y. Elementary Schools, and the Mesdames Barry, Fairand and Teeter of the Freeville Elementary School for their cooperation with this research.

Each type of pattern was represented at least once each day, with the disjunctive always in the last position. The disjunctive appears to be a higher-order abstraction, and it was hypothesized that it would be more difficult to solve.

The task was a two-choice discrimination problem, with immediate feedback. The positive and negative words were printed individually on separate cards, and presented in pairs to the S, who was asked to pick the "correct" one. He was told whether his response was correct, the pair was removed, and the next pair in that set presented. Four such groups of 4 pairs of stimuli were repeated unless the S reached the criterion of 6 correct responses in a row, in which case the E went on to the next set.

Errors were recorded, as well as observations on apparent strategies and the Ss comments when they were relevant to the experiment.

Because this task was quite difficult for the kindergarten children, their motivation and attention declined disastrously after a few days. We revised the task then to include some apparatus which we hoped would make the task more interesting—essentially a light box which the child activated himself, and a buzzer for reinforcement. This had some of the desired effect, but didn't alleviate all the problems.

It appeared that the memory load between trials was too great for these children. While not knowing their alphabet, they had to remember not only the distinctive differences between the positive and negative instances, but also to which set they were to respond with "yes."

A sorting task, with correction, seemed to impose less of a memory load, and had the important advantage of allowing more activity by the child. One group of kindergarten children and all of the first graders were therefore tested on sorting the whole set of cards into two categories, corresponding to the positive and negative items. Each card was presented as often in this task as in the choice discrimination, the important differences being that 1) the items appeared sequentially, thereby focussing attention on the individual word, and 2) the S could compare his present choice with his last choice in both the negative and positive sets. This task was decidedly easier than the choice discrimination, both in time to complete a day's sequence of problems, and in the total percentage of problems solved. We plan on continuing with this procedure in the main experiment.

Results

The data reported are from a total of 12 kindergarten children who were able to complete the five days of testing, and from five first graders. There were a number of kindergarten children who, primarily for motivational reasons, could not finish.

The task was definitely easier for the first graders, 3/5, or 60% of whom showed evidence of forming a learning set, as contrasted with 1/12 or 8.3% of the kindergarten Ss. Criterion for learning set was a progressive increase in the percentage of problems solved per day with a final level of at least 80%.

Analysis by type of constraint for kindergarten children showed no difference in the percentage of problems solved for the initial, middle, and final clusters. The disjunctive pattern was substantially more difficult for all Ss, however, while the double letters (e.g. final LL) in both the middle and final positions were substantially easier. The same relationship between pattern positions held for the first grade Ss.

It is interesting to take a close look at a particular S in the experiment--one who performed exactly as we had hoped they all would. Lynn is a kindergarten girl (age 5) who was tested on the sorting task. She developed a learning set for solving these problems by looking for the two letters which were invariant in a particular position for each problem. Her progressive percentage of problems solved per day was 25, 40, 60, 100, 100. By the third day she verbalized her letter matching strategy, and first mentioned that the reversal had "the right letters in the wrong spots," which indicates that she was utilizing the letters of the pattern in proper sequence. On day five, she solved all the problems but one in the first sorting: the exception was due to application of her correct strategy to the wrong letters (a fault of the stimulus material which we shall correct). It is interesting to note here that while Lynn was capable of excellent discrimination of letters, she could name only one, L, which began her name.

Because there were very great individual differences, we looked at individual cases to see whether there were different learning styles. Most of the children fall into one of the three groups described below.

1. Several of the children (especially in the kindergarten group) reached criterion on only a few problems. On the others, they typically performed, over all trials, a little above chance but far from criterion. It looked as if, insofar as they were learning anything, they were learning individual pairs. When they reached criterion, one cannot say they "solved" the problem: they may simply have learned, individually, which four items were called "right." They may on the other hand have abstracted the distinctive feature, in this one case, which made it easier to learn which four were right, but the feature was specific to the problem and transferred to nothing else. To decide between these interpretations, we intend to add a transfer task with new instances of the same specific pattern.

2. Another group of four children reached criterion on five or six problems, and typically stayed around chance performance on the others. When the problems these children solved are examined, they tend to fall in a single category, such as all the problems with double letters as a common pattern, or all those with an initial cluster in common. These children had abstracted something not merely specific to the problem, but instead a <u>kind</u> of pattern that could and did recur, and they transferred this to the appropriate problems. That they developed a hypothesis about it is suggested by the fact that they did no better than chance (and even sometimes worse) on the other problems, indicating that they were looking for something that didn't work in these problems, rather than memorizing individually which items were correct.

3. Four children (three in first grade and one in kindergarten) showed clear evidence of development of a generalized learning set. Their rates of successful solution progressed steadily over the five days reaching 100% and in all cases a "rule" was clearly verbalized. Two other children showed evidence of a developing set, reaching 100% on one day and giving some verbal evidence of it. Further practice would probably have consolidated it.

The progression in levels of abstraction in these styles of learning word discriminations or (when it happens) spelling patterns, suggests that we may, as one of our next steps, be able to work out a program for introducing problems so as to "shape" a child toward progressively higher order abstractions.

Some Effects of Redundant Stimulus Information on Learning to Identify Letters

Eleanor J. Gibson and Sharon Shepela

Abstract

The effect of introducing a redundant stimulus property on learning to identify letters was studied in kindergarten children. Color was made redundant with the vowel sound of a letter's name for one group of Ss. Color was present but uncorrelated with the sound for a second group. All letters were presented in black for a third group. Introduction of color did not affect the rate of learning, but correlated color and sound did produce interference as measured by confusion errors, suggesting that the redundancy was effective. There was transfer of identification to black letters, with little loss. Significantly more colors than chance would predict were remembered by the kindergarten children, but this was not the case for fourth grade children in a comparable task. Sixth grade children again remembered more colors than chance, suggesting that attention does not develop toward a strict trade-off of selected for incidental information but toward some optimal level of selectivity and exploratory behavior.

The notion that the perceptual differentiation and identification of letters depends on pick-up of distinctive features that yield a unique pattern for each letter has been developed by E. J. Gibson (in press). Perceptual detection of the distinctive features of a set of objects is in fact a process basic to many educational tasks. Learning the features that are critical for discriminating rocks in a geology class or plants in a botany class might be examples. The educator considering his training problem is bound to ask "how do we facilitate the pick-up of distinctive features?" The obvious answer would seem to be to enhance them so that they stand out from noisy or non-informative context. But what is the best way to enhance them? One answer might be to introduce along with the critical properties redundant correlated cues that are attention-getting and easy to pick up.[1] But this strategy brings its own problems. For one, will the student attend only to these cues and be actually distracted from noticing the critical features themselves? Or will he eventually detect that they are correlated but irrelevant so that he can transfer his newly learned differentiation in their absence? Second, can a young child make use of redundant information in the sense that he detects the correlation? If he does not, the procedure would almost certainly be worthless.

It has been suggested that children need more redundancy of stimulus information than adults to achieve the same effectiveness of perceptual

1. Every-day examples of this procedure are numerous, such as drawing little faces on notes of the musical scale for teaching young children.

differentiation (Wohlwill, 1960; Bruner et al., 1966). But the evidence for this is not clear. It might be that redundant information only seems to be useful to children because some children make use of one aspect of the correlated properties for discrimination and some children, another. Group means might look better if this were the case, even though an individual child was not really affected and was using only one property. Even more disquieting is some evidence that children are less able than adults to detect redundancy when it is useful; that is, to perceive the correlation or constraint between two properties and reduce the information to be processed (Munsinger, 1967; Munsinger and Gummerman, 1967).

We decided to make an exploratory (and what seemed a very simple) attack on this problem with a letter identification task. Five-year-old children who had not "learned their letters" were to receive training in identifying letters by name. Redundancy was introduced into the task by printing letters whose name had a common sound in the same color. The correlated stimulus information was thus only partially redundant (a condition of more interest than perfectly correlated, one-to-one redundancy because in the latter case economical discriminatory processing would not have to filter out the critical features). The questions were: (1) would this partial redundancy facilitate learning in the beginning; (2) would it interfere when progress reached a point where the correlated color cue no longer helped to complete the identification of each letter uniquely; (3) would the redundant cue be filtered out at some stage of learning; or (4) would it make no difference at all compared to appropriate control groups. It is possible that groups of letters having a common sound (for example, A, J, K) will be identified earlier if the common sound is emphasized, for this will classify the material in an early stage of learning and by giving it more structure, reduce the information to be processed. It is also possible, however, that at some stage where discrimination demands a finer analysis by features, interference from the partially correlated cue could result.

A second question of interest to us was the potential usefulness of crossmodal (e.g. visual-aural) redundancies in learning an identification or a concept. The question is important, because in reading, visual-aural correspondences in partial correlation are typically present. The problem is whether the auditory dimension plays a facilitating role in helping the learner to make the necessary discriminations. If so, it could be especially useful when detection of higher-order constraints or rules (as in spelling patterns) is involved. The rare pertinent investigations (Lordahl, 1961; Haygood, 1965; Laughlin et al., 1968) have been concerned only with adult Ss and highly artificial concept attainment.

Experiment I

Metho

Experimental design. The experiment included three conditions, each one employing a different group of children. In all three groups, 9 letters were presented for the child to identify. In Condition I (Experimental group), the letters were divided into three sets, each set having a common vowel sound in the letters' name (A K J, E Z D, L N S). The letters were printed in color, a different color for the letters of each set. In Condition II (Color control), the letters were printed in the same three colors, but assignment of a color to a letter was uncorrelated with the sounds of the letters. In Condition III (Control) all nine letters were printed in black. When training ceased, the children in Conditions I and II were tested with the set of black letters for transfer in the absence of color. The children in the control group were given a comparable recall trial with black letters. Following this test, the children in Conditions I and II were shown the black letters again and asked what color the letter had been. There was no previous suggestion that they would be asked this question.

Procedure. Pilot experimentation with children in a summer Headstart group taught us that learning to name nine letters is surprisingly hard for many children and in fact impossible of achievement in one or two "sittings." It is also impractical to demand that all children reach a criterion of 100% identification in a limited training time. Consequently, we gave each child three individual training sessions on successive days, each one lasting about one-half hour. The letters were Roman capitals 1½ in. high prepared on individual cards 3½ x 3½ in. The procedure was as informal as possible while controlling the amount of information given the child. The experimenter showed the child the nine cards, presenting them one at a time on a stand, and told him the name of the letter. Then the pack was shuffled and the cards presented again, the child attempting to name the letter. The experimenter told him if he was right, and gave the correct name of the letters if he was wrong or gave no name. All the child's responses were recorded. Non-specific encouragement was given between runs. Five runs through the pack constituted a day's training. The three days of training were followed by the two tests described above.

An early try-out of rewarding the child was an M and M candy for each correct response proved unsatisfactory. Motivation to continue through the training was achieved much more satisfactorily by letting the child choose a gift from an array of toys, putting his name on it, and giving it to him when the experiment was completed.

Subjects. The subjects were Kindergarten children in the Ithaca and Dryden, N. Y., school systems. Children who already knew as many as three of the letters were eliminated. There were 12 children in each condition.

Results

The results of major interest were the number of letters learned in the three respective experimental conditions, the number of letters correctly recalled when color was omitted or added, the distribution of errors in relation to the color variable, and the number of letters correctly assigned to a color. We shall call the latter incidental learning, since the child's attention was never drawn to the color until this final test.

The number of letters correctly identified varied considerably from one child to another, but the variability was similar between groups. The greatest number of letters identified correctly in the final recall trial was 7, the least 0. The best comparison of the rate of learning to identify the letters in the three conditions is seen in the learning curves in Figure 1, where the mean number of correct responses is plotted over the three days of training. Following the last training session, the number of letters identified correctly on the transfer test is plotted. The learning curves for the three groups are very similar.[1] To judge from the increment of correct responses given as training progresses, introduction of the color seems to have had no effect. On the average, the children had learned about 4.5 of the 9 letters in all three conditions.

On the transfer test, the number of letters recalled by the two groups who practiced with colored letters dropped slightly, but not significantly compared to the number recalled by the children who practiced with black letters and were tested on black ones. Omission of the color thus seemed to have had little effect on retention of those letters learned. If a letter could be distinguished and named when it was presented in color, it generally still could when it was presented in black. But the fact that the criterion of learning achieved was so low over all makes a conclusion on this point rather tentative. Facilitating and interfering effects might have cancelled each other or have been concealed in the averaging process.

The confusion errors offer a more sensitive means of detecting potential interference effects. One can determine the effect of sharing a color on errors by determining the number of confusions within a same-color set of three letters, and comparing with them the number of confusions of those letters in the results of the group shown only black letters. The effect of a common color combined with a common sound can be found by comparing confusion errors within a set for Condition I where color and sound were correlated with the effect on confusions of common color alone (color control group) and common sound alone (control group).

In order to make these comparisons, confusion matrices were constructed for all three groups, entering all confusion errors made over the fifteen trials. The three matrices are presented in Table 1, A, B, and C. The entries down the diagonal represent correct responses; the others represent confusions between a given pair of letters. The total number of entries is not equal, since some children gave no response oftener than others, and a few errors were intrusions of some letter other than the nine being taught (the first letter of a child's name, for instance).

To compare the effect of sharing a color on confusion errors, we can compare the number of confusion errors within the three sets of letters sharing a color in the "color control" condition, with the <u>same</u> letters in the control condition, by pulling these confusions out of the matrix and totalling them for each condition. These totals were then converted to percentages of the total number of confusions in the matrix. The percentage of confusions within same color sets of letters in Condition II was 23.16%; and for the same sets of black letters in Condition III, 23.37%.

1. The curve for the Color Control Group is consistently very slightly lower than the other two groups. This difference is accounted for by two very slow children who happened to be in that group.

Table 1A

Confusion Matrix (Condition I, Experimental Group)

The confusions of interest are those between
A K J, E D Z, and N L S

	Stimulus Letter								
Response	A	D	E	J	K	L	N	S	Z
A	68	8	3	4	0	7	12	7	3
D	8	56	24	5	2	16	5	6	12
E	12	23	46	8	11	14	17	13	12
J	1	0	1	73	3	7	5	0	3
K	16	1	3	11	86	7	4	6	2
L	4	5	6	10	8	27	7	5	4
N	7	3	3	3	2	7	40	0	2
S	2	3	1	0	1	8	7	61	2
Z	4	10	15	1	7	7	10	7	103

Table 1B

Confusion Matrix (Condition II, Color Control Group)

The confusions of interest are those between
E K S, J D N, and L Z A

Stimulus Letter

Response	A	D	E	J	K	L	N	S	Z
A	95	24	31	12	18	13	15	12	7
D	4	48	15	4	8	9	9	6	4
E	4	7	22	11	3	8	8	4	3
J	3	6	2	61	4	10	5	9	4
K	12	13	11	5	59	12	14	12	10
L	9	4	8	13	5	65	9	9	5
N	3	4	3	2	12	2	36	0	5
S	3	3	4	1	3	3	3	54	4
Z	3	6	5	3	1	3	2	6	79

Table 1C

Confusion Matrix (Condition III, Control Group)

Stimulus Letter

Response	A	D	E	J	K	L	N	S	Z
A	80	0	6	2	4	4	5	12	3
D	6	66	16	3	0	10	12	4	4
E	6	1	29	1	0	3	4	4	0
J	6	3	3	78	0	7	2	3	1
K	6	2	12	5	113	11	9	6	2
L	3	3	4	14	1	37	4	5	1
N	5	1	4	1	2	5	26	2	2
S	0	5	3	0	2	3	2	49	2
Z	2	3	3	2	0	6	6	3	90

It thus seems that common color alone did not increase confusions. Does sharing a common sound alone increase them? We can answer this question by pulling out of the matrix of the Control group (Condition III) the confusions between letters having common sounds and comparing them with the confusions between the sets of letters compared with Condition II, where there were no common colors. The percentages are respectively 24.39% and 23.37%. It thus appears that neither did common sound alone significantly increase confusions.

Now we can ask whether redundancy of another property with the common one will enhance its effect. The interesting comparison here is between confusions of letters with common sounds in the control group and the same letters in the experimental group where sound and color were correlated. There were 35.73% color-sound confusions in the experimental group and 24.39% confusions of the same letters, sharing a sound alone, in the control group. This difference is significant ($P = .03$). It thus seems reasonable to infer that enhancement of a property by adding a redundant one, even from another sensory channel, may have an effect. It is not the color alone which is effective (as the appropriate comparison of the two control groups shows) but the combination of two properties.

How does this reinforcing effect of two correlated properties take place? Probably not consciously, for no child ever noticed that the sound and color were correlated in Condition I and yet there was an effect of the two combined that was not an addition of the two alone. In this case the effect showed up as interference by increasing confusion errors, but not in slowing the rate of learning over all trials. It is conceivable that it might even have resulted in facilitation eventually if we had continued training to a criterion of 100%, since there was evidence of transfer to identifying the black letters. It might be suspected that color is eventually filtered out as irrelevant and no longer noticed. That this did not occur, at least at the stage of learning reached in our training, is manifest by the results on recall of colors.

When the children in Conditions I and II were shown the black letters in the final test and asked what color they had been, the children in Condition I remembered, on the average, 5.08 correctly; the children in Condition II, 5.5. This is a better recall than that for the letter names in the just previous test (a mean of 3.67 for Condition I and 4.08 for Condition II). Since there were only three colors but nine letter names, one might expect more colors to be correct by chance if names and colors were equally attended to; but why should they be? The children were never asked for the color but were persistently asked for the names.

Several alternative hypotheses suggest themselves. The first is that these children, who could not previously identify letters, did not know what the distinctive features of letters are and were trying to use color as one of them; they simply had not yet learned the distinguishing properties and thus could not dismiss color as irrelevant. If this were correct, they should eventually have done so if we had continued training until they all reached perfect, consistent identification. Would they, in that case, begin to filter out the color dimension and not recall the colors?

Another hypothesis is that younger children do more incidental learning than older children because they have not matured in the ability to inhibit or attenuate input from an irrelevant context of task-relevant features. Several experiments have produced evidence that although task-directed (intentional) learning increases with age, incidental learning does not or even decreases (Maccoby and Hagen, 1965; Siegel and Stevenson, 1966; Hagen, 1967; Hagen and Sabo, 1967). These experiments differed in several ways from the present one. The irrelevant material was 100% correlated with the relevant; criteria of learning and method of testing differed; and instructions differed. Whether or not the child is told in the instructions what the irrelevant properties are, it seems obvious that for irrelevant context to be suppressed, it must be recognized as irrelevant. Instructions and the kind of material employed will make a profound difference in this respect. In our experiment, the child had to know from previous experience that color differences are not integral to the distinctive features of letters, or he had to discover what the distinctive features of letters are in the course of the experiment.

The concept of a filter suggests that there is a limit to the amount of information that can be processed. Redundant information could, in that case, overload the system. But then there should be a trade-off between the amount of relevant or irrelevant information that gets processed, with a negative correlation between the two. We ran a correlation, therefore, between the number of letters correctly recalled and the number of colors correctly recalled. The correlation (+.317) was positive and not significant. There is no evidence, therefore, of a trade-off. Maccoby and Hagen (1965) likewise found no such trend. Hagen (1967) used percent of total recall as a measure and found an age-related trend--a positive correlation in earlier years, which reversed with age. Hagen and Sabo (1967), however, did not confirm the trend.

Experiment II

It seemed of interest to know whether older children who are aware of the distinctive features of letters would exhibit the same tendency to recall colors as well as or better than letter names under these conditions of partial redundancy and no mention of the color variable. We therefore modified the experiment so that a comparable task could be given to older children.

Method

Nine artificial graphemes were copied from an experiment by Gibson, Yonas and Schapiro (in preparation). They were drawn on cards in the same proportion as the letters used before. Only Condition I was replicated. Three of the graphemes were colored red, three yellow, and three blue, and within a set were to be labeled A, K, J; L, N, S; E, D, Z, as in the experiment with kindergarten children and real letters. Two different random assignments of names to the artificial graphemes were used, each for half of the Ss. Again, the Ss did paired associate learning. The order of the cards displaying the graphemes was randomized from trial to trial. The Ss were run to a criterion of 7 out of 9 correct. This was more than the average number correct for the kindergarten children (about 4) but that criterion was so easy for the older Ss as to be attainable on a single trial, which seemed too little to give a chance for any incidental information to be retained. After criterion was reached, the graphemes were presented one at a time in black and the S asked what color each had been.

The Ss were 12 fourth grade girls and 12 sixth grade girls enlisted from Girl Scout troops.

Results

The trials to criterion, number of correct responses on the transfer trial and number of colors correctly recalled are shown in Table 2. The criterion was easily attained in one training session for all fourth grade and sixth grade Ss. The entries for the kindergarten Ss are taken from the previous experiment, where the graphemes were real letters. There is no reason to think that real letters are harder to discriminate than the artificial graphemes. Indeed, discrimination latency data (Gibson et al.) suggest that they might be easier.

Table 2

Recall of Incidental vs. Critical Responses by Three Age Groups

Grade	Mean Trials to Criterion	Mean Correct responses, transfer test	Mean correct color responses
K	15	3.88	5.29
4	3.93	6.25	3.0
6	3.50	7.16	5.41

The kindergarten Ss recalled correctly significantly more colors than chance would predict ($p < .02$). The fourth grade children not only did not recall more than chance expectation, but some of them had not even noticed that the graphemes were different colors, or else stated erroneously what the three colors were. But there is not a consistent trend in the direction of decreased incidental recall, for the sixth grade Ss did remember more colors correctly than chance would predict. The number (5.41) accounts, however, for less of the total recall (43%) than does incidental recall of the kindergarten children (5.29 colors is 57% of their total recall). These data are somewhat comparable to a trend found by Siegel and Stevenson (1966). In their experiment, there was a decrease in incidental learning between 11 and 14, and then an increase again for adults. The age ranges are different, however, and it seems likely that the ease of the task is very important. There is certainly no magic age where maturation of "inhibitory power" or the like reverses trends.

Discussion

What do these differences in recall mean? The simplest interpretation would seem to be that the kindergarten children are learning the colors during the training sessions because they do not yet know that they are irrelevant. Fourth grade children, however, have learned that color is not an invariant distinctive feature for differentiating letters. When they are asked to learn letter names for some artificial letter-like graphemes, they turn their attention to the features that serve to distinguish letters; straight line, curve, intersection, opening, etc. The sixth graders undoubtedly do the same thing, but they remember more altogether. The task does not overload them with information and we have already seen that there does not seem to be a strict trade-off between selected and incidental information. When there is not an information overload, this would seem to be an adaptive characteristic of behavior. If attention developed toward rigid selectivity, it would return to the inflexibility and captive nature of attention so often pointed out in the young infant (Stechler and Latz, 1966; Ames and Silfen, 1965). Selectivity with some low but optimal level of exploratory activity would seem to be the mature ideal. This view does not necessarily imply that ability to ignore the irrelevant does not develop, but it places more emphasis on learning to select what is relevant.

To return to the questions asked in the introduction, we have only tentative answers. Does a partial redundancy such as might enhance some critical feature facilitate learning? It did not in our experiment, but it might have if learning had been carried to completion. On the other hand, there is evidence of interference from the error data. This interference does not preclude the possibility of overall facilitation, for it may be that an increase in overt confusions hastens the ultimate differentiation. The introduction of redundancy as an enhancer can have both positive and negative effects and only a very careful analysis of the task could yield sensible predictions. Even then, they surely need to be tested before adoption in any educational program.

Does a partially redundant "enhancer" or cue get filtered out eventually? We would have had to push training to completion to answer this question satisfactorily, but there is some reason to think that it does. When the children were transferred to the test with black letters, there was little loss in correct identification. Insofar as they had learned to differentiate and identify a few letters, the color dimension of the stimulus item could be dispensed with. Whether this would happen if it were perfectly redundant, as in certain experiments with adults (e.g. Weiss and Margolius, 1954; Saltz, 1963; Saltz and Wickey, 1967) is another question.

Our final question, do children pick up a visual-aural redundancy at all, whatever its effect, receives some affirmation. The combination of common sound and common color within a letter set increased confusions of pairs within the set, whereas common sound or color alone did not. The contingency seems to have had an effect, although it was never remarked on. As regards the general question of how the pickup of distinctive features can be enhanced, this finding suggests that further exploration of the effects of crossmodal redundancies is worthwhile. It should be compared with intramodal redundancy, as well as with other potential means of enhancing invariant distinctive features.

References

Ames, E. W. & Silfen, C. K. Methodological issues in the study of age differences in infants' attention to stimuli varying in movement and complexity. Paper presented at the meeting of the Society for Research in Child Development, Minneapolis, Minn., 1965.

Bruner, J. S., Olver, R. R., Greenfield, P. M., et al. Studies in cognitive growth. New York: Wiley, 1966.

Gibson, E. J. Principles of perceptual learning and development. New York: Appleton-Century-Crofts (in press).

Gibson, E. J., Yonas, A., & Schapiro, F. Confusion matrices for letters and for artificial graphemes with a latency measure. In preparation.

Hagen, J. W. The effect of distraction on selective attention. Child Development, 1967, 38, 685-694.

Hagen, J. W. & Sabo, R. A. A developmental study of selective attention. Merrill-Palmer Quarterly, 1967, 13, 159-172.

Haygood, D. H. Audio-visual concept formation. J. educ. Psychol., 1965, 56, 126-132.

Laughlin, P. R., Kalowski, C. A., Metzler, M. E., Ostap, K. M., & Venclovas, S. Concept identification as a function of sensory modality, information, and number of persons. J. exp. Psychol., 1968, 77, 335-340.

Lordahl, D. S. Concept identification using simultaneous visual and auditory signals. J. exp. Psychol., 1961, 62, 282-290.

Maccoby, E. & Hagen, J. W. Effects of distraction upon central versus incidental recall: Developmental trends. J. exp. child Psychol., 1965, 2, 280-289.

Munsinger, H. Developing perception and memory for stimulus redundancy. J. exp. child Psychol., 1967, 5, 39-49.

Munsinger, H. & Gummerman, K. The identification of form in patterns of visual noise. J. exp. Psychol., 1967, 72, 75-81.

Siegel, A. W. & Stevenson, H. W. Incidental learning: A developmental study. Child Development, 1966, 37, 811-817.

Stechler, G. & Latz, E. Some observations on attention and arousal in the human infant. J. Amer. Acad. Child Psychiat., 1966, 5, 517-525.

Saltz, E. Compound stimuli in verbal learning: Cognitive and sensory differentiation versus stimulus selection. J. exp. Psychol., 1963, 66, 1-5.

Saltz, E. & Wickey, J. Further evidence for differentiation effects of context stimuli: A reply to Birnbaum. Psychol. Reports, 1967, 20, 835-838.

Weiss, W. & Margolius, G. The effect of context stimuli in learning and retention. J. exp. Psychol., 1954, 48, 318-322.

Wohlwill, J. F. Developmental studies of perception. Psychol. Bull., 1960, 57, 249-288.

Confusion Matrices for Graphic Patterns Obtained with a Latency Measure

Eleanor J. Gibson, Frank Schapiro, & Albert Yonas

One of the more difficult but essential tasks for a child in learning to read is the identification of the letters of the alphabet. This achievement involves not only the attaching of a unique name to each letter; there must be a one-to-one matching of the graphic pattern to the name, and that accomplishment requires visual differentiation of each of the graphic patterns as uniquely different from the others. How are these patterns differentiated? It has been suggested (Gibson et al., 1963; Gibson, in press) that this is accomplished not by a process of matching the total form of a letter to a stored representation that is independent of other letters, but rather by a more economical process of discriminating a smaller set of relational distinctive features that yields a unique pattern for each letter. The features are distinguishing with respect to the set, not components fitted together to build the letter.

What might these features be, and how can we investigate such an hypothesis? It is apparent that a set of distinctive features smaller than the total set of letters requires that some letters share certain features with others. Uniqueness for a given letter can be achieved only by a pattern of features that is distinct from the others. It follows from this that two letters which share a number of features and differ from one another minimally by only one or two, should be harder to discriminate than two letters which differ from one another by many features. We can test this prediction by determining experimentally the actual confusability of the letters with one another.

This program was followed by Gibson et al. (1963) and a confusion matrix was obtained for the 26 Roman capitals, using errors in a matching task as the measure. Subjects were 87 pre-school children. Since every letter had to be matched against every other with equal opportunities for error, a very large number of comparisons were required. Any one child could go through only a small block of the design before becoming fatigued. Since we did not want errors due to inattention or misunderstanding of the task, there was a very small yield of errors, despite a large investment of experimental time. The error matrix contained many "holes" where no errors occurred. Those that did occur, however, showed a low but significant relationship to an intuitively generated feature list. While the correlations between degree of feature-difference and confusion errors were not large, it seemed worthwhile to make a new attempt to collect more satisfactory data on confusability of letters.

It was decided to use a latency measure, rather than errors, since there would then be a yield of information on every trial and no empty spaces in the final matrix. The old literature on disjunctive reaction time (Cattell, 1902; Lemmon, 1927; Woodworth, 1938) suggested that response latency is increased when highly similar items must be discriminated. Latency should, therefore, yield a good measure of confusability. It was possible to check this expectation by correlating latencies with such errors as occurred.

The judgment chosen was "same" or "different." A pair of items was presented simultaneously and the subject responded by pressing a button with one hand if they were the same and by pressing another button with the other hand if they were different. We were interested in this judgment because the comparison of mean latencies for "same" responses with mean latencies for "different" responses is pertinent to the question of what processes go on in discrimination, as well as is the distribution of latencies for "different" pairs (Egeth, 1966).

Two sets of nine letters each were chosen for test. Each set was chosen to include a range from very different to very similar pairs, as predicted by feature differences. This would give us two 9 x 9 confusion matrices to check on our predictions. We did not attempt to obtain the 26 x 26 matrix of all the letters, since running only once through the simplest design for this would require 1300 judgments. We also devised a set of nine artificial graphemes, constructed so that the features we thought might distinguish letters would also distinguish them. This would permit another check on the predictability of features, one with unfamiliar letter-like patterns. We were also interested in comparing mean "same" and "different" latencies for these unfamiliar patterns, which might lack codability or "Gestalt-like" properties characterizing real letters.

It was possible that our rather low correlations between feature differences and errors in the earlier experiment had to do with the age of our Ss (a mean around four years) who were as yet rather inexperienced in discriminating letters. We therefore ran Ss of two age groups; adults and seven-year-old children. Whether the same features are selected for discrimination at all developmental levels is an interesting question, and it is possible that adults, after long experience, achieve some higher-order more economical means of processing visual letter-patterns. Since there are so many unknown possibilities for processing differences dependent on age and material, it was decided to analyze the data by a cluster method in order to compare it with our a priori feature analysis.[1]

Method

Material. Both the letters and artificial graphic stimuli were prepared photographically on slides. The first set of letters run was C, E, F, G, M, N, P, R, W. The second set was A, D, H, K, O, Q, S, T, X. The type was simplified Roman capitals. The same "master" copies were photographed for all the slides. The artificial graphemes are sketched below in Fig. 1. A master copy was prepared for each one, black on white, with the same line-thickness as the letters. The artificial graphemes differed from one another in such features as curve, straight, diagonality, inter-

1. A multi-dimensional analysis is also in process.

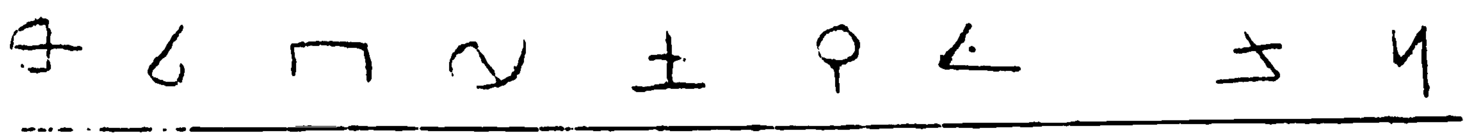


Figure 1. Artificial graphemes.

section, and others deemed characteristic of letters. We attempted to include items that were very similar and others that differed by many features, on an a priori basis.

Apparatus. The S sat in a chair with his eyes approximately $5\frac{1}{2}$ feet from a display apparatus. His foot was on a starting pedal and the index finger of each hand on a response button. The slides were presented to him simultaneously, via a projector. They were projected in two small windows of a Foringer display apparatus. The windows were 2 x 2 in. and were separated by 1 1/8 in., one to the right of the other at the same height. They could be observed without shifting fixation. The slides were presented automatically, paced at a 3 to 4 sec. interval by a programming device. If the pace became fatiguing for the S, he could remove his foot from a pedal and stop the displays. He could start them again when he was ready by pressing the pedal. A pair of slides presented for the same-different judgment were exposed until the S made his decision, when he pressed one of the two buttons for a "same" or a "different" judgment.

When S pressed the button, a shutter closed, terminating the display, and the S's response time from the beginning of the display was measured and recorded automatically by an IBM keypunch. A green light went on when S responded correctly. A red light came on if he made an error.

Procedure. The adult Ss made 360 judgments of same or different, following a brief practice period to become accustomed to the set-up. Half of the presentations were "same" pairs and half were "different" pairs, to control for response bias. Each of the 9 same pairs was presented 20 times. Each of 36 different pairs was presented five times. Order of presentation was randomized on replications within and between Ss. Only one arrangement for a given different pair was used (e.g., either AB or BA but not both), but a given letter, throughout, appeared equally often on left and right. Procedure was similar with the artificial characters.

It was essentially the same with the children, but the number of judgments was reduced to two replications for each "different" pair and eight for each "same" pair. The children were allowed to choose a toy when they came in and were presented with it at the close of the experiment. The experiment was conducted in the laboratory with adult Ss but in a mobile laboratory at the school with the children.

Instructions were as follows:

"Every three or four seconds two letters (or forms) will appear on the screen before you. You are to decide if these two are the same or different. If they are identical press the button marked same with your index finger. (Keep your finger on that button so that you can respond rapidly.) If they are different, press the button marked different with the index finger of your other hand.

Focus between the two windows.

So long as your foot is pressing the foot pedal, the experiment will continue. If you need to stop, lift your foot; this stops the projector until you are ready to continue.

You are to respond as quickly as you can without making mistakes.

We will have 3 practice trials. Try to learn which hand is same and which is different. The green light means you were correct; the red light indicates that you were in error."

An S was asked which was his dominant hand. The dominant hand was assigned to the same key or the different key with alternate Ss.

Twenty-four adult Ss were run in the first experiment with the first set of letters. This experiment was then replicated with another 24 adult Ss, since we wanted a check on reliability of our measure. Sixty 7-year-old children were run on this set of letters. Twenty-four adult Ss were run on the artificial graphemes, and 24 more on the second set of letters. Twenty-four children were run on the artificial graphemes, but the time cost was so great that this phase was discontinued.

Subjects. The adult Ss were college students obtained from the subject pool of the introductory Psychology course. The children were obtained from a summer day-camp at the school.

Results

Latency Data for Letter Pairs

Adults. The results for the two groups of 24 adult Ss run on the first set of letters were very similar; the correlation between the two was +.82. The data were threfore combined, and are presented in Table 1.

The latencies for responding to a pair of different letters by pressing the appropriate key varied from 458 msec. (GW) to 571 msec. (PR), a range of over 100 msec. The difference in latency of responding to these two pairs, the extremes of the range, was very significant ($P < .0001$). In fact, differences as low as 30 msec. are significant at .05 or better. This method of testing confusability thus gives a useful spread of responses and, judging from the confirmatory results of the replication, is reliable. The correlation is particularly satisfactory, since the replication was run by a different experimenter in a different place.

There is also a range of latencies for judging "same" to identical

Table 1

Mean Latencies (in msec.) for a Response of "Same" or "Different" to Pairs of Nine Roman Capital Letters[1]
(48 Adult Ss)

	C	E	F	G	M	N	P	R	W
C	467	500	495	552	481	465	496	483	467
E		475	560	479	502	495	504	490	485
F			488	464	488	491	495	495	482
G				496	463	470	501	481	458
M					497	545	477	481	538
N						500	463	486	510
P							466	571	461
R								487	489
W									488

pairs (down the diagonal), from 466 msec. (PP) to 500 msec. (NN). The range is by no means as great as for different pairs, but the extremes of the range are nevertheless significantly different ($P < .01$). Deciding that NN or MM is a "same" pair requires a longer time than does PP or CC. Something that is peculiar to the letter is influential in even such a simple perception as sameness. Letters containing diagonality seem to be associated with longer latencies, but the observation may not hold up and the reason for it, in any case, is not clear. It may be that speed of a "same" judgment depends on the composition of the series to some extent, such as whether other letters highly confusable with it are part of the set being tested.

The results for the second set of letters run on 24 adult Ss are presented in Table 2. The range of latencies for "different" judgments extends from 472 msec. (QT) to 593 msec. (OQ), a slightly longer range than the first set. The mean for all the judgments is also a little higher for this set, whether due to the sample of letters or to the sample of Ss is not clear. In any case, there is a wide spread of latencies and the difference between extremes is highly significant ($P < .0001$). Differences over 30 msec. are significant at .05 or better.

There is again a variation in "same" judgments, from 476 msec. (AA) to 555 msec. (QQ). This is a considerably longer range of "same" latencies than for the first set. The difference between extremes is very significant ($P < .0001$). Letters containing diagonals do not generally take

1. In calculating means, a cutoff was employed to discard extremely short or extremely long responses. An S occasionally made false responses (impossibly short ones) or ones so long that he could not have been attending. The cut-off was set at 200 msec. for too short and at 1 sec. for too long. A computer program automatically discarded these items and also the latencies for errors.

Table 2

Latencies (in msec.) for a Response of "Same" or "Different" to Nine Pairs of Letters (24 Adult Ss)

	A	D	H	K	O	Q	S	T	X
A	476	509	534	521	505	489	535	484	520
D		494	497	510	548	521	507	511	491
H			514	580	505	484	499	540	542
K				526	504	493	496	492	588
O					491	593	512	490	490
Q						555	506	472	490
S							486	514	501
T								487	524
X									523

longer. The Q, which takes longest, has a very confusable letter, O, in the set, but the OO pair is judged "same" comparatively rapidly. There may be some structural feature of the "same" pair as a whole, like symmetry, that influences speed of processing. We shall return to this speculation in a later section.

<u>Children</u>. Latency data for the 60 seven-year-olds are presented in Table 3. Latencies are much longer than those of adult Ss, but again there

Table 3

Latencies (in msec.) for a Response of "Same" or "Different" to Nine Pairs of Letters (60 Seven-year-old Ss)

	C	E	F	G	M	N	P	R	W
C	1047	1097	1112	1252	1130	1093	1128	1107	1110
E		1033	1229	1107	1112	1124	1054	1114	1144
F			1072	1108	1139	1132	1092	1148	1106
G				1082	1127	1148	1099	1091	1042
M					1043	1279	1063	1057	1203
N						1075	1084	1112	1155
P							1030	1138	1070
R								1050	1036
W									1038

is a range of latencies. The shortest latency was 1036 msec. (RW) and the longest 1279 msec. (MN), a range of 243 msec. The extreme differences are satisfactorily significant ($P < .001$), although the children are quite variable.

The range of latencies for "same" responses, on the other hand, is relatively small--1030 msec. (PP) to 1082 msec. (GG). Yet the fastest same pair was the one that was fastest for adults and the difference between the extremes is significant at $<.05$.

The correlation between the children's latencies and those of the adults for the same set of letters was +.53. This is not as high as the correlation between the replications for adults. The lower correlation could reflect a lesser reliability in the judgments of the children, or it might mean that the children are not processing the feature differences in the same way as the adults. A comparison of cluster analyses of the two matrices should answer this question.

Confusion Errors between Letters

Adults. Error data ~~for the 48 Ss run~~ on the first set of letters are presented in Table 4. While there were not a great many errors there is a fair range of confusions between letter pairs, from none (GM) to 28 (PR). There are errors for "same" judgments, as well. These are not more frequent than "different" errors, as might appear, since there were four times as many opportunities for them.

Table 4

Number of Errors of Mistaking a Different Pair
as Same or a Same Pair as Different
(48 Adult Ss)

	C	E	F	G	M	N	P	R	W
C	21	6	4	19	3	4	8	4	3
E		14	14	5	10	6	5	8	4
F			25	4	5	4	8	6	8
G				41	0	2	7	4	3
M					38	17	3	5	14
N						41	5	9	11
P							13	28	6
R								26	3
W									33

The interesting feature of the error data is its relation to the latency variable. As Figure 2 shows, the two variables are positively correlated, despite the fact that errors are not numerous. The cluster analysis, to be presented later, makes this abundantly clear. The more two different letters are mistaken for the same, the longer it takes to discriminate them correctly as different. We therefore feel confident in using the latency measure as the dependent variable for testing predictions about features and dimensions of stimulus similarity, since there

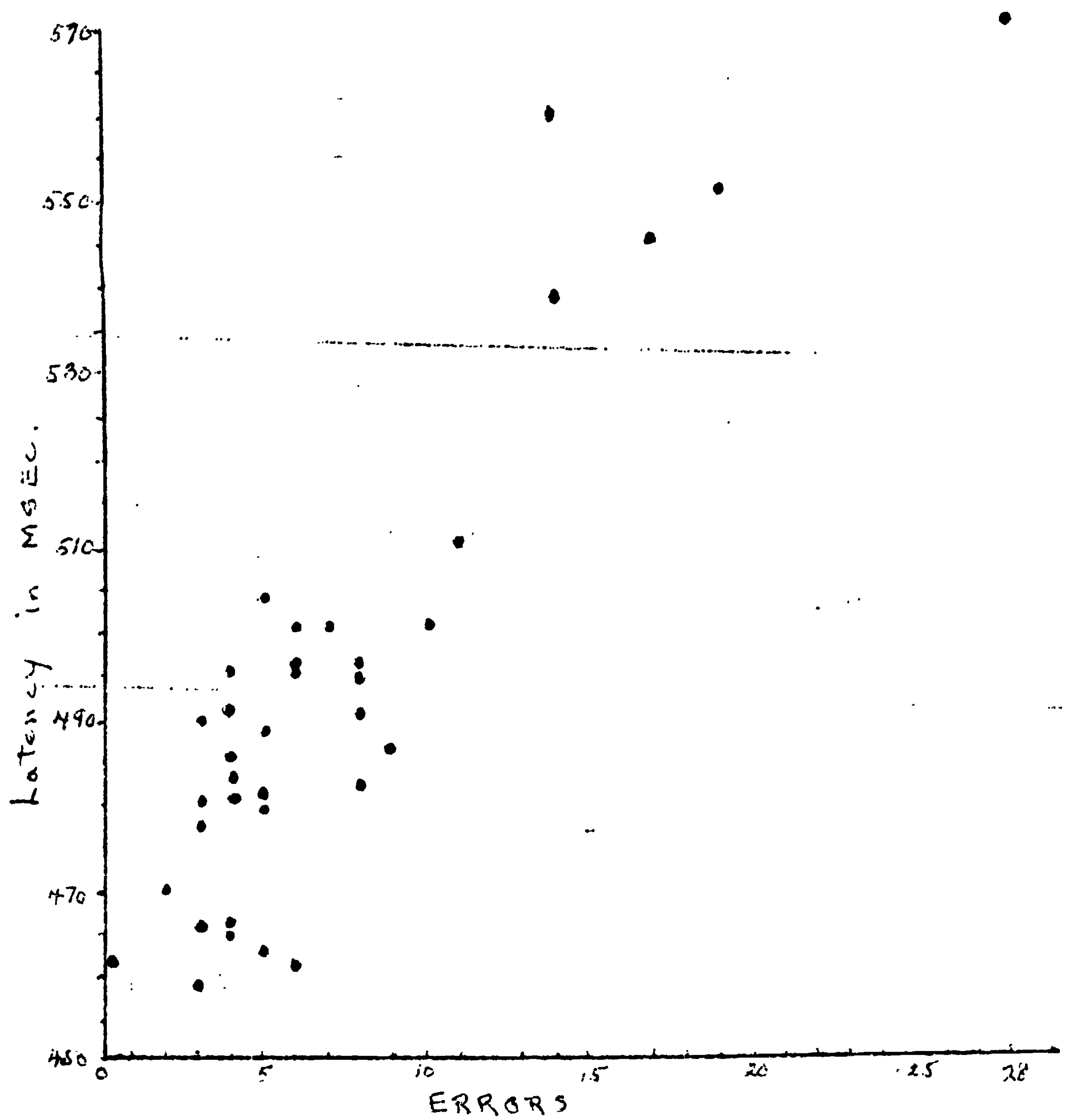


Figure 2

Latency plotted against errors for pairs of different letters (first set of letters, 48 Ss) is a wealth of latency data, shown to be very reliable.

The error data for the other set of letters are presented in Table 5. These subjects made very few errors, perhaps related to the fact that their overall mean latencies were higher than those for the first set of

Table 5

Number of Errors of Mistaking a Different Pair as Same or a Same Pair as Different (24 Adult Ss)

	A	D	H	K	O	Q	S	T	X
A	2	2	-	2	2	-	-	1	-
D		4	1	2	1	-	-	-	1
H			7	3	-	-	-	2	1
K				11	-	1	-	2	3
O					4	3	-	2	2
Q						13	3	-	1
S							4	-	2
T								3	2
X									29

letters. They may have set a higher criterion of confidence for a decision. Nevertheless, the three longest latency pairs, OQ, KX, and HK also have the most errors.

Children. Error data for children are presented in Table 6. Errors are few, despite the large number of Ss, because each S made so few judgments. The children's long latencies may also indicate a high criterion for errors. Nevertheless, the highest confusions (EF, CG, MN, and PR) coincide with the highest latencies for discriminating a difference.

Table 6

Error Data for Pairs of Different Letters (60 Seven-year-old Ss)

	C	E	F	G	M	N	P	R	W
C	12	2	1	8	3	2	4	6	5
E		14	10	1	2	3	3	3	5
F			12	3	1	1	3	3	4
G				15	3	6	5	1	2
M					8	8	4	3	5
N						17	4	6	4
P							15	9	4
R								13	7
W									11

Artificial Graphemes

Adults. The latencies for judging pairs of artificial graphemes "same" or "different" are presented in Table 7. Judgments of "different" vary from 459 msec. (vs.) to 527 msec. (vs.), a range

Table 7

Latencies (in msec.) for a Response of "Same" or "Different" to Pairs of Nine Artificial Graphemes (24 Adult Ss)

	479	472	488	473	493	497	477	507	496
		456	477	470	487	515	467	469	454
			459	476	462	467	527	481	475
				455	474	459	463	482	492
					477	474	488	516	503
						481	496	468	498
							479	472	498
								487	514
									484

of 68 msec. This is not as long as the range of "different" latencies for either set of letters. Is this because the set of artificial graphemes is not as well differentiated as the very familiar letters? If that were the case one would expect the overall mean latencies for the artificial graphemes to be considerably longer than those for letters, but they are not. They are slightly shorter, in fact, than the overall means for the second set of letters. We conclude that although the artificial graphemes are not familiar in the sense of figures identified or codable as wholes, nevertheless their distinguishing features (curve-straight, diagonality, etc.) are the same as those of letters and are thus readily detected. The shorter range may very likely be due to the fact that we simply didn't construct a pair as similar as, for instance, OQ. The difference between the shortest and the longest latencies for the artificial graphemes is nevertheless significant ($< .001$).

The judgments of "same" vary from 455 msec. ( ) to 487 ( ), a range of 32 msec. The extreme difference is significant at $< .001$. Again, detection of replication is easier for some pairs than for others. Why? We know very little about detection of regularities in general and our ignorance here is manifest. There are the same number of lines in the two figures. But are there more features to be processed in one than the other? Or is a judgment of "same" made on the basis of a higher-order feature characterizing the pair as a whole?

Error data for the artificial graphemes are presented in Table 8.

They are too few to permit any strong conclusions, but the two longest latency pairs also have the most confusions ([illegible] , [illegible]) and ([illegible] , [illegible]).

Table 8

Errors in Discriminating Pairs of Artificial Graphemes
(24 Adult Ss)

	31	5	1	7	4	3	4	3	4
		5	3	5	1	3	1	2	0
			15	3	1	5	14	2	0
				12	3	4	3	5	4
					19	3	5	10	2
						24	2	1	4
							29	1	1
								19	3
									21

Children. Latency data for artificial graphemes for 24 children are presented in Table 9. The number of judgments for each pair was small

Table 9

Latencies (in msec.) for Discriminating Pairs
of Artificial Graphemes as Different
(N = 24 Children)

	1216	1295	1267	1305	1332	1244	1304	1253	1313
		1183	1275	1265	1105	1177	1086	1299	1223
			1161	1285	1213	1383	1271	1269	1355
				1177	1245	1277	1178	1137	1240
					1139	1255	1252	1257	1239
						1178	1385	1228	1231
							1175	1144	1220
								1234	1317
									1208

(two per child) and the variability between Ss very great, so few of the latency differences between pairs are reliable, and they do not correlate very well with those of the adults. The errors with these artificial graphemes (see Table 10) are distributed nearly evenly over the matrix, with a range of only 5, so we reluctantly conclude that there is little to be learned from them.

Table 10

Error Data for Pairs of Artificial Graphemes
(24 Seven-year-old Ss)

	6	3	2	2	2	1	0	5	2
		0	1	2	3	3	2	1	0
			3	2	1	3	0	0	2
				3	2	1	2	1	3
					9	2	0	2	1
						5	1	1	3
							5	1	1
								6	2
									4

Latency of "Same" and "Different" Judgments

If the latency of judging "different" increases as the similarity of two items increases, should it not take longest of all to judge that two identical items are the same? If one conceives of a sequential feature-testing process that compares two items to find a match, the process might be expected to take longer if an exhaustive search of every single feature must be carried out, as a judgment of "same" would presumably require. A decision of "different" could be reached as soon as any difference was found and the search could stop. The range of latencies for judgments of "different" supports such a notion; the more two items differ, the sooner a difference should be found, and that is exactly how the latencies for "different" pairs look.

But when one looks at the mean latencies for "same" as compared with "different," this reasoning breaks down. The overall means for "same" judgments are shorter than the means for "different," as Table 11 shows. This trend is present in the adult's judgments of both sets of letters, in

Table 11

Mean Latencies of "Same" and "Different" Judgments in msec.

	Different	Same
Adults, first set letters	493	485
Adults, second set letters	514	505
Children, letters	1121	1052
Adults, artificial graphemes	507	496
Children, artificial graphemes	1270	1186

the children's judgments of letters, and in the judgments of artificial graphemes for both children and adults. The trend was particularly striking (and surprising) in the children. Of the 60 children run on letters, 53 had shorter mean latencies for "same" judgments. Of the 24 run on artificial graphemes, 21 had shorter mean latencies for "sames."

How is this to be interpreted? Does it mean that a feature-testing model of discrimination must be wrong? Would a template-matching model do any better? It might for successive discrimination, for, as Sorenson (1968) has suggested, a subject could be "set" with an appropriate template for a "same" judgment. If the second item matches the template that is "ready," the decision is made. If not, a further search of some sort must go on. But what sort? If it were a comparison with a set of templates, why should there be such a systematic range of latencies for "differents?" Indeed, why should there be a significant difference in latency between some of the "same" judgments?

One simple explanation that cannot be easily dismissed is that the "same" pairs are repeated more often. This procedure was necessary to control for response bias in our experiment, since there were 36 "different" pairs and only 9 "sames." It seems reasonable that repetition should reduce latency. The finding that "sames" are shorter, on the average, has been reported by others before us (Nickerson, 1965; Egeth, 1966), but so has the opposite finding (Bindra, Williams, and Wise, 1965; Chananie and Tikofsky, 1968). Bindra, Donderi, and Nishisato (1968) compared "same" with "different" latencies as a function of several variables (stimulus modality, simultaneous vs. successive presentation, inter-trial interval, etc.). All pairs of stimuli were presented equally often, which appears to dispose of repetition as the sole explanation of discrepancies. They concluded that the discrepant findings were due to codability (or lack of it) of the stimulus items. Letters, for instance, are readily codable (identifiable by name) and yielded shorter latencies for "sames." Pairs of tones differing in pitch, or pairs of lines of different length are not easy to identify absolutely, and these items showed longer latencies for "sames." This observation, interesting as it is, seems unsatisfactory as an explanation, because it gives no inkling of <u>why</u> "sames" should be shorter or longer depending on their codability. It also does not explain the range within "same" distributions. Certainly one letter is just as codable as another, and yet decisions are made significantly faster on some pairs than on others (and not even on the most frequent ones). Finally, the artificial graphemes in our experiment were far less codable than the letters, but the same relationship held as with letters.

These rather baffling discrepancies suggest that the process of deciding that two things are the same is not like that of deciding that they are different; that a simple model that says check out all the features, either sequentially or in parallel, is just not appropriate for both. It seems more likely that under certain circumstances, a decision of "same" is a <u>direct perception of replication</u>; of a structural property of the pair as a whole that requires no further look at subordinate features. Replication would be more readily detectable under some circumstances than others--perhaps when simplicity and symmetry are present--but the conditions for it are certainly not well understood. In any case, perception of replication can be a fast short-cut to deciding that two things are the same, as contrasted with a feature-by-feature check. This interpretation is born out by an experiment of Sekuler and Abrams (1968). Recognition of identity of a pair in their experiment was very much faster than finding similarity, even though the similarity decision involved finding a "same" feature (any feature) in two pairs. It is thus not the judgment of "same" as such that is faster, but the opportunity for immediate apprehension of replication--"Gestalt processing" as Sekular and Abrams put it.

Cluster Analysis of "Different" Pairs

The question of greatest interest to us is what is linked with what in pairs of different items so as to make them confusable to different degrees. Can we pull out from our data some indication of what the basis for differentiation or confusion is? A hierarchical cluster analysis (Johnson) seemed to offer a promising method. This is, loosely speaking, a method of progressively clustering the set of letters. If we find systematic differentiation in our letter sets and artificial graphemes, perhaps we can identify the features that account for clustering. If the same ones turn up in replications of the same letter set and with different letter sets, we are in luck.

Tables 12a through 12e show the results of cluster analysis of latency and error matrices for both letters and for both sets of letters. Consider Table 12a, based on latency data of 48 adult Ss for the first set of letters. Results of two methods, "connectedness" and "diameter" are presented. These are very similar, so we shall discuss only one, that by diameter. The analysis pulls out in the first row the most compact and isolable cluster, PR; in the next row, EF appears. Other pairs appear progressively, until longer and looser clusters are left, winding up with only two separated ones, CGEFPR on the one hand, and MNW on the other. One can think of the analysis the other way around, as a progression from the total undifferentiated set toward more and more specific clusters. A tree structure shows nicely the contrasts that emerge when the clusters appear. At the first branch, all the letters on the right contain diagonals (MNW) while those on the left have straight lines and/or curves. At the next branch, MNW splits into MN versus W (all diagonals); the big cluster CGEFPR splits into round letters, CG, vs. letters with verticality, EFPR. At the next branch the cluster EFPR differentiates into a purely vertical-horizontal cluster, EF, vs. one with curves and verticals, PR.

The error data contain some ties, which makes the analysis less satisfactory in terms of meeting one of the underlying assumptions of the method. It is interesting to note, however, that an identical hierarchical structure emerges, strong evidence of the correlation between latency and errors. The children's latency data (Table 12c, 60 Ss) are very straightforward. The first branch is a curve-straight differentiation, all the letters with curves bunched together. Then the cluster with curves separates into the "round" cluster and the curve and straight cluster (PR). The right branch separates into a diagonality cluster and a vertical-horizontal cluster, etc., exactly as with both sets of adult data. The cluster analysis of children's errors does not yield quite as orderly a

Table 12a

Cluster Analysis and Tree Structure

Set 1 letters, Latency (48 adult Ss)

Connectedness Method	Diameter Method
PR	PR
PR EF	EF PR
CG PR EF	CG EF PR
CG PR EF MN	CG EF PR MN
CG PR EF MNW	CG EF PR MNW
CG PREF MNW	CG EFPR MNW
CG PREFMNW	CGEFPR MNW

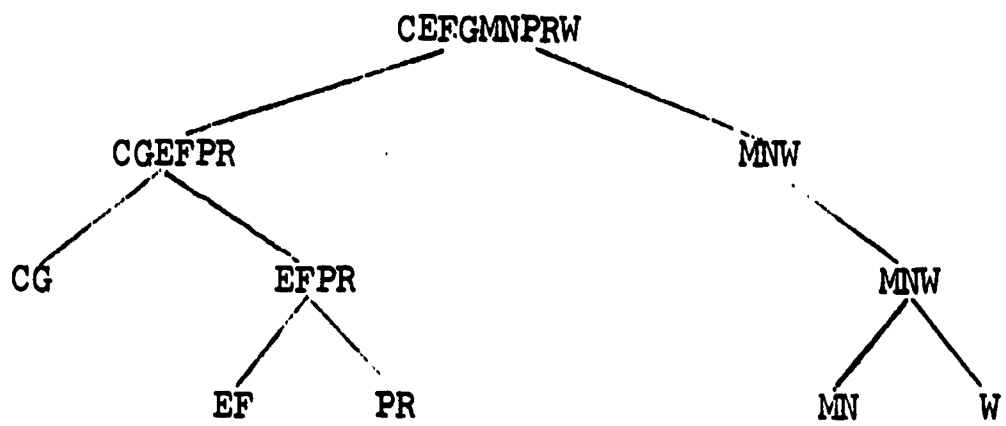


Table 12b

Cluster Analysis and Tree Structure

Set 1 letters, Errors (48 adult Ss)

Connectedness Method	Diameter Method
PR	PR
CG PR	CG PR
CG PR	CG PR MN
CG PR MN	CG EF PR MN
CG PR EF MNW	CG EF PR MNW
CG PR EFMNW	CG EFPR MNW
CG PREFMNW	CGEFPR MNW

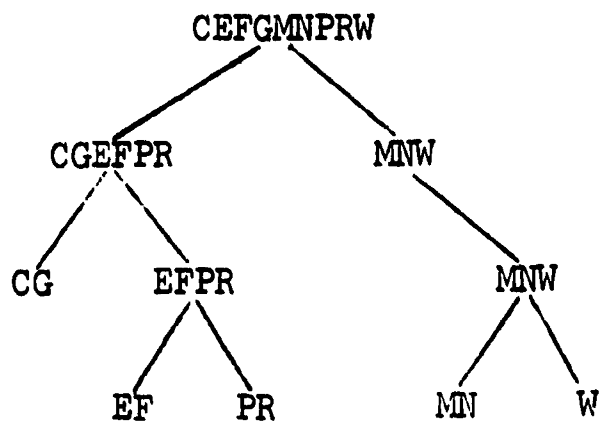


Table 12c

Cluster Analysis and Tree Structure

Set 1 letters, latency (60 7-year-old Ss)

Connectedness Method

MN

MN CG EF

MNW CG EF

CGMNXW EFR

Diameter Method

MN

CG EF MN

CG EF MNW

CG PR EF MNW

CG PR EFMNW

CGPR EFMNW

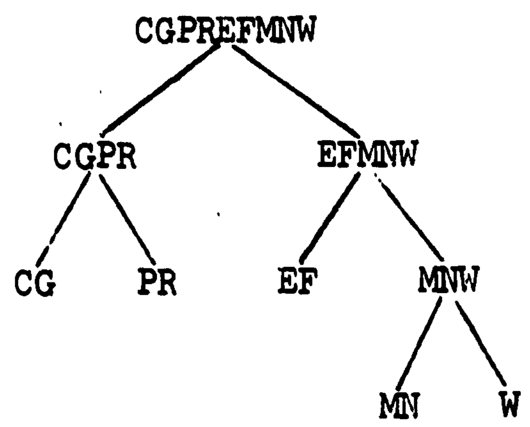


Table 12d

Cluster Analysis and Tree Structure

Errors for Letters (60 children)

Connectedness Method

EF			
EF	PR		
EF	CG	MN	PR
EF	CG	MN	PRW
EF	CGMNPRW		

Diameter Method

EF			
EF	PR		
EF	CG	MN	PR
EF	CG	MN	PRW
EF	CG	MNPRW	

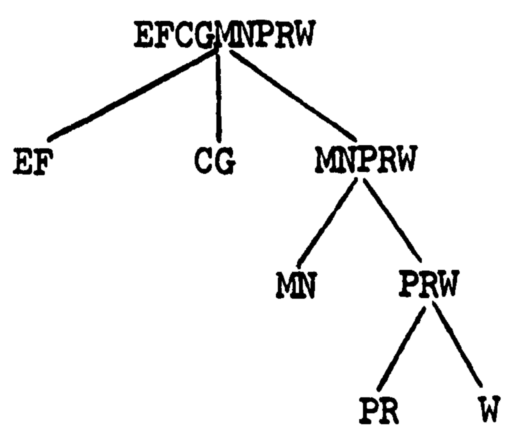
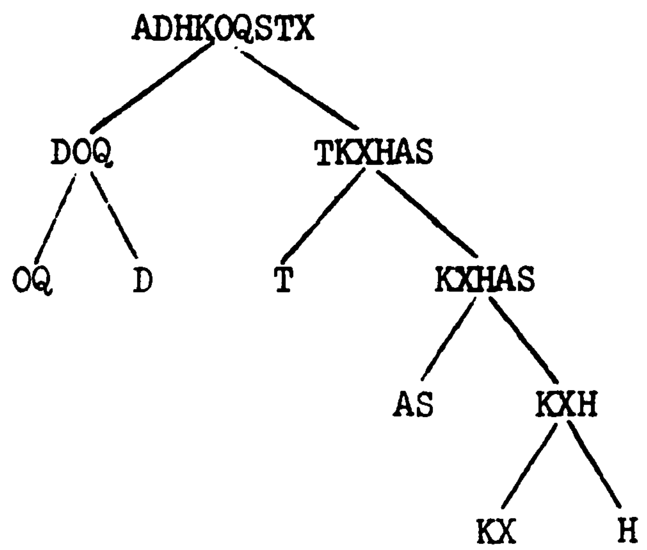


Table 12e

Cluster Analysis and Tree Structure

Set 2 letters, Latency (24 adult Ss)

Connectedness Method	Diameter Method
OQ	OQ
OQ KX	OQ KX
OQ KXH	OQ KXH
DOQ KXH	OQ KXH AS
DOQ TKXH	DOQ KXH AS
DOQ TKXH AS	DOQ KXHAS
DOQ TKXHAS	DOQ TKXHAS



structure, probably because there were just too few errors, but all the small clusters are the same as for latency, suggesting the same basic feature contrasts.

The other set of letters yields similar contrasts, as Table 12e shows. Latency data for 24 adults are presented (errors were too few for an analysis). Again the first branch contrasts curved-straight, branching again on the left to "round" vs. curve and straight (OQ vs. D). On the right the last branch again yields a diagonality vs. vertical-horizontal split. The structure here is not as orderly as for the other set (the AS cluster looks strange), probably because there were only 24 Ss and a few confusions are accidental. There is a hint here of something that does not show up in the other letter set in the T vs. KXHAS split. One might call it intersection or "information in the middle." Since the data in each case are based on a sample of only nine letters, a new feature could easily turn up in the second set.

Since we do not have the complete matrix of all 26 letters, we cannot expect these analyses to generate all the features that would be necessary to provide a unique pattern for all 26 letters. But one can easily see in the two sets of shapes certain features that are high in the tree-structure: curve vs. straight; diagonality; vertical vs. horizontal; intersection; and relatively closed (round) vs. open. These are the first features chosen intuitively in our earlier attempt at a feature analysis. The replications within these experiments indicate to us that perception of difference in a pair of shapes does depend on detecting a set of distinguishing features; and that the set of features is not random or idiosyncratic for either a shape or a perceiver but is rather an orderly hierarchy of fairly abstract properties.

Conclusions

The latency for discriminating a pair of graphic characters gives a wide range of times, depending on the pair to be discriminated. Differences between pairs are significant, for both adults and seven-year-old children. The latencies, furthermore, reflect the tendency to confuse the members of the pair. Two sets of nine letters and a set of artificial graphemes bore out these trends. The mean latency for deciding that a pair was the same was slightly but reliably shorter than deciding that it was different. This trend held for artificial graphemes as well as for the familiar letters, and for children as well as adults. Not all "same" pairs had equal latencies, some being significantly shorter than others. The reason for this, we think, is that a judgment of same may be in some cases a direct perception of replication, without further analysis of features.

There is every reason to think, however, that a judgment of "different" for these graphic characters involves an analysis of distinguishing features. A pair that differs by many features is seldom confused and the decision is faster than for a pair sharing many features. Furthermore, hierarchical cluster analyses of the matrices yielded tree structures that showed an orderly progressive differentiation of features. Both latency data and error data for children and adults on the first set of letters yielded very similar structures. The latency data of the adults for the second set of letters, while not yielding identical structures, suggest that a very similar hierarchy of features are detected in the discrimination process for both sets of letters. We conclude that perceiving a difference between two letters is not a matter of matching to a Gestalt-like template, or decoding to a name, but involves detection of distinctive features.

References

Bindra, D., Donderi, D. C. & Nishisato, S. Decision latencies of "same" and "different" judgments. Percept. & Psychophys., 1968, 3, 121-130.

Bindra, D., Williams, J. & Wise, S. S. Judgments of sameness and difference: Experiments on reaction time. Science, 1965, 150, 1625-6.

Cattell, J. McK. The time of perception as a measure of differences in intensity. Phil. Stud., 1902, 19, 63-68.

Chananie, J. D. & Tikofsky, R. S. Reaction time and distinctive features in speech discrimination. Report No. 49, in a Program on Development of Language Functions, Univ. of Michigan, 1968.

Egeth, H. E. Parallel versus serial processes in multi-dimensional stimulus discrimination. Percept. & Psychophys., 1966, 1, 243-252.

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

Gibson, E. J., Osser, H., Schiff, W. & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In Final Report on A Basic Research Program on Reading, Cooperative Research Project No. 639, Cornell Univ. and U. S. Office of Education, 1963.

Johnson, S. C. Hierarchical clustering schemes. Bell Telephone Labs., Murray Hill, N. J.

Lemmon, V. W. The relation of reaction time to measures of intelligence, memory, and learning. Arch. Psychol., 1927, 15, No. 94.

Nickerson, R. S. Response times for "same"-"different" judgments. Percept. mot. Skills, 1965, 20, 15-18.

Sekuler, R. W. & Abrams, M. Visual sameness: A choice time analysis of pattern recognition processes. J. exp. Psychol., 1968, 77, 232-238.

Sorenson, R. T. Guidance of attention in character recognition: The effect of looking for specific letters. Ph.D. dissertation, Cornell Univ., 1968.

Woodworth, R. S. Experimental psychology. New York: Holt, 1938.

Studies of Oral Reading[1]

I. Words vs. Pseudo Words

Harry Levin and Andrew J. Biemiller[2]

Cornell University

This study is the first of a series on oral reading. The experiments will share a common empirical model and a common analysis of the sub-skills which comprise the complex process which is reading. The independent variables will be variations of the stimulus materials: words, pseudo words, phrases, sentences, and the like. The dependent response variable will be latency: the period of time between the presentation of the graphic stimulus, and the onset of the subject's verbal response, reading the word or words aloud. In the present study, we also examined the effects of the experimental variations on reading errors, and the relationships between errors in reading and response latency.

For analytic purposes, we think that reading involves two sub-processes which, at least during the period when the skill is unformed, take place in sequence. The first is decoding in which the reader converts written material into associated language. The second process is code use during which the reader converts the decoded writing into information, guides to actions, etc. Our purpose in the present study has been to establish response latency as a valid index to the process of decoding. We chose, therefore, stimulus materials which, on *a priori* grounds, represented two

---

1. This study was supported by funds from the Cooperative Research Program, U.S. Office of Education.

2. We wish to thank Miss Susan Bostwick for her help with this study.

extreme degrees of difficulty. In addition, the ages of the subjects were chosen to represent various degrees of skill in reading. We reasoned that if latency was responsive to the variations in stimuli and in subjects, we could be confident in its use as a dependent variable in subsequent studies.

When we originally conceived the present study, it seemed reasonable to expect that words which a child had encountered frequently, either in print or aurally, would be decoded and responded to more rapidly than words which a child had never before encountered. We operationalized this notion by constructing two sets of words, eight "real" words and eight "pseudo" words (e.g., BLERM) and presenting them to second, third, and fourth graders to read aloud.

The validity of response latency as an index of decoding difficulty would be sustained by the following results:

1. Shorter response latencies to real words than to pseudo words.
2. Shorter response latencies associated with higher grade levels in school.
3. Shorter response latencies during the second trial compared to the first.

METHOD

Stimulus materials. The words used appear in Appendix A. The word list was generated as follows:

a. Four initial spelling patterns were selected: BL, CH, GR, ST.

b. Four final spelling patterns were selected: CK, ND, RM, SS.

c. Each initial pattern was combined with each final pattern.

d. By manipulating the vowel letters a snd e, two real and two pseudo words were generated for each initial and each final spelling pattern. All words had five letters.

e. Where possible, two words with e and two with a were provided for each spelling pattern. This was not always possible. The vowel o was used in storm as the only possible word meeting criterion (d) above. The word bland was treated as a pseudo-word on the assumption that it would be unfamiliar to most of our sample. We now think this was an error. Three additional words were used as "warm-up" words.

Subjects. We tested 54 children from the West Hill Elementary School[3] in Ithaca, N.Y.. This school serves a population with a wide variation in economic levels. Eighteen children each were drawn from the second, third, and fourth grades. Half of each group of eighteen were boys and half girls. These subgroups were chosen according to "reading ability" by drawing three boys and three girls from the "best" reading group in a class, three, from the "worst who can read" and three, from a group "in between." (We have not reported data for this ability grouping because we do not feel that the assessment is reliable).

Procedure. Each child was informed that "we are getting recordings of how children talk. I'm going to show you a lot of words on the screen and I want you to tell me what they are. We made up some of the words, so you needn't feel badly if you don't know them."

A lapel microphone was attached and the words projected on a screen three feet from the subject. The projected words were about six inches

3. We are grateful to the staff of the school for their cooperation.

long by two inches high. After the child responded to a word, there was a two second interval before presenting the next word. If the child made no response to a word in fifteen seconds, he was asked, "Would you like to go on to the next one?" (He invariably did.) Similarly, if the child suggested going on, the next word was displayed.

RESULTS

The subject's responses were tape recorded. The tapes were then played through a rectifier which was connected to a pen-writing Brush recorder. This system activates the pen when sound is present and the pen comes to rest during silence. A characteristic sound made by the projector served as a marker indicating the presentation of the stimulus. Latency was measured from this point to the onset of the last word the child gives as a response to that stimulus. All false starts, vocal segregates, etc., are treated as part of the response latency. Omissions were arbitrarily scored as a 12.5 second latency. (We observed that the maximum latency followed by a response was twelve seconds.)

The data were transformed according to the following formula: $x = \log (2.5y - 1)$, where x = transformed score; y = response latency in seconds. This transformation is discussed in Woodworth and Schlosberg (1954, p 39).

The transcription of all tapes was used to make a qualitative analysis of the reading errors. The results of this analysis appear in Appendix B to this paper.

The reliabilities of individual children's response latencies are given in Table 1. Each S read the list twice. The list orders were the same for all children. Subjects were quite consistent from trial to trial, although, as will be seen shortly, the mean latencies decreased over trials.

Another measure of consistency is the relationships between the latencies to real and pseudo words, within trials. As can be seen in Table 1, children are consistent in reaction times to these two types of words, in spite of the fact that their latency for real words is shorter than their latency for pseudo words.

Table 1

Correlations Between Individual's Scores On First Trial vs Second Trial

	Grades 2nd	3rd	4th	Combined
First Trial x Second Trial (all words)	.87	.62	.81	.86
Real Words x Pseudo Words:				
First Trial	.74	.68	.68	.83
Second Trial	.99	.70	.62	.82

The main results of the study are summarized in Table 2. An analysis of variance was calculated on these data, with the classifications, real vs. pseudo-words, grade levels, and trials. The results of this analysis are:

1. Children show longer latencies in reading pseudo words than real words ($F = 70.96$, $df = 1, 51$; $p < .01$).
2. Latencies are longer on the first than on the second trial ($F = 28.83$, $df = 1, 51$; $p < .01$).
3. Younger children evidence longer latencies than older children ($F = 10.56$, $df = 2, 51$; $p < .01$). This finding is qualified by an interaction between grade levels and word type ($F = 8.59$, $df = 2, 51$; $p < .01$). Perusal of the means in Table 2 indicates that second graders show a small difference between real and

pseudo-words, whereas for the third and fourth grades the mean latency for pseudo-words is substantially longer than for real words.

Table 2

Average Response Latencies (in seconds)

	Grades			
	2nd	3rd	4th	Combined
All words	4.51	2.66	2.43	3.20
Real Words	4.10	1.85	1.55	2.50
Pseudo Words	4.86	3.48	3.31	3.90
First Trial	4.74	2.96	2.64	3.45
Second Trail	4.27	2.36	2.22	2.95

Errors in Reading. From the tapes, we judged whether or not the word was read correctly. Of the total of 1728 responses, 1035, or 60% were read correctly. Further, 42% of the pseudo-words and 78% of the real words were correct. In this section, we shall examine the determinants of frequency of errors and the latencies in reading words correctly and incorrectly.

The mean number of errors per subject are given in Table 3. An analysis of variance according to type of word, grade levels and trials yielded two significant main effects and no significant interactions. There are more errors in reading pseudo-words ($F = 104.5$, $df = 1, 51$; $p < .01$) and in the second compared to the third and fourth grades ($F = 7.3$, $df = 2, 51$; $p < .01$).

In addition to the frequency of errors, we investigated the latencies in reading words correctly or incorrectly. The mean latency for correctly read words is 2.08 seconds and for incorrectly read ones, 4.85 seconds. The difference between these means is statistically significant ($t = 64.29$, 181 df. $p<.005$).

Table 3

Average Number of Errors

	Grades			
	2nd	3rd	4th	Combined
All Words	4.6	2.5	2.2	3.1
Real Words	3.4	1.0	0.7	1.7
Pseudo Words	5.7	3.9	3.7	4.4
First Trial	4.6	2.5	2.4	3.2
Second Trial	4.5	2.5	2.1	3.0

The first unadorned analysis indicated that reactions were more rapid to real words than to nonsense ones and that younger children responded more slowly, in general. The subsequent analysis of whether the reading was correct force us to make serious qualifications to the original findings. Pseudo-words are more frequently read incorrectly and younger children make more errors. The latencies are longer to errors than to correct read words. Are the first findings, then, due simply to the differences in frequency of errors? Two subsequent analyses clarify this issue. The latencies for each child were divided into correct and incorrect readings and within this control, the mean latencies to real and pseudo-words were inspected. The results are given in Table 4. Interestingly, the real-pseudo difference holds up <u>only for correctly read words</u>. If the child makes an error, his response is roughly equally slow for both types of words.

Again using the correctness of reading as a control, we inspected age differences in latencies. The initial finding generally holds: older children read words more quickly than younger ones.

Table 4

Average Response Latencies (in seconds) for Words Read Correctly and Words Red Incorrectly.

	Grades 2nd	3rd	4th	Combined
All Words:				
Correct	2.66	2.14	1.67	2.08
Incorrect	5.88	3.41	4.62	4.83
Real Words:				
Correct	2.54	1.61	1.24	1.69
Incorrect	5.97	3.82	4.40	5.32
Pseudo Words:				
Correct	2.91	3.25	2.42	2.82
Incorrect	5.83	3.91	4.66	4.68
First Trial:				
Correct	2.80	2.34	1.72	2.21
Incorrect	6.09	4.30	4.87	5.27
Second Trial:				
Correct	2.49	1.94	1.63	1.95
Incorrect	5.67	3.47	4.34	4.71

DISCUSSION

Reading is a private process. The principle barriers to research on the process of reading are the lacks of clear external indices to the process. Eye-movements are one such index, but the measurement of eye movements are extremely complex and fraught with difficulties of interpretation. Frequently, reading is studied by tests of speed of reading or comprehension of what has been read. For our purposes, these measures confound the decoding and information processing subskills. We decided, therefore, to take oral reading as an index which will be common to a series of experiments. The process will be inferred from the ways in which the common index varies with systematic variations in the stimulus materials. Taking such an external manifestation of reading leaves us vulnerable to the contention that reading aloud and silently involves different skills, basically. We doubt that this is true, although a firm answer must itself wait on research which compares the two modes of reading. McLatchy's 1949 study of second graders shows a high association between scores on oral and silent reading tests. (Edfeldt, 1960; Flavell, 1965). We might point out, also, that developmentally, oral precedes silent reading and it is a common observation that when the materials being read are difficult, there is a tendency to mouth or to say the words.

This study was designed to test the validity of response latency as a behavioral index to reading. As such we chose stimulus materials and an age range which should maximize differences among groups. If the index were not sensitive to these extreme variations it would be useless although we do not yet know its potential value in detecting more subtle variations. In general, the results indicate the merit of latency for future research.

Although our strategy was empirical, the results, even at this early stage of research, tempt us to theorize about the process of decoding written words to their language equivalents. One tactic of the reader, and a highly unlikely one, is that the reader starts at the left of the word and sounds out the letters or groups of letters serially. If it happened, this would be a pure instance of decoding from spelling to sound. There is ample evidence that reading does not work this way. In our data, such a decoding process would not yield the differences in latency or errors between real and pseudo-words. Also, we observed that when our subjects read the words aloud, they usually read smoothly with the sounds blended together, even when the response was an error.

Our theory must account for these findings: (1) correct real words are read more quickly than correctly read nonsense materials, and (2) incorrectly read real and nonsense words take equally long. As a first approximation to a theory we hypothesize that speakers of a language store in memory auditory representations of the sounds of their language, English in our case. We say, for example, that a snatch of language we hear and which we do not understand "sounds like English." When the word is exposed to the child he rehearses it. He matches the consequences of this rehearsal to his auditory memory and emits it with varying latencies and correctness, depending on a number of factors.

If he decodes to a familiar group of sounds, there is a close match between his response and his memory and the word is emitted. Correctly read real words are emitted rapidly (an average of 1.69 seconds). The pseudo-words in this experiment were designed to abide by English spelling patterns so that their correct rehearsal would yield English-like sounds (the latencies to these words average 2.14 seconds.) It is tempting to think that the real words are read rapidly because they are familiar to the child or because in their decoding the child makes a judgment about their meaning. While these steps may take place, we prefer the more general formulation of a dimension of familiar sounds in which familiar, previously heard words anchor the dimension at one end.

In the light of this reasoning, consider the relationships between errors and real and pseudo words. Decoding errors, in both cases, move the result toward the unfamiliar end of the sound dimension. The children's equal and long latencies for both categories probably reflect their perplexity with the outcome of the decoding. It may be that the subjects rehearse the sounds, checking to see whether they can bring the sounds into line with their auditory memories. Whether or not the words "look" familiar has little effect, since decoding errors lead to roughly equal latencies for both types of words.

The responses to the word bland are instructive. From our pretest experiences we put the word into the pseudo category because none of the children knew what it meant. Nevertheless, in this study, the mean latency in reading bland was the shortest of all the pseudo words and briefer than some of the meaningful words. The word is made up of some common English sound elements--land, and--so that decoding yielded a familiar sound pattern, but not a familiar word.

If our reasoning is correct, errors in reading which eventuated in real words should have briefer latencies than errors which were finally read as nonsense forms. Such was actually the case. The mean subject latency for errors read as real words was 2.9 seconds, while for errors read as pseudo words it was 4.0 seconds. This effect is clear both on real words and pseudo words. Sign tests of this difference are significant at the .001 level for errors on real words and the .005 level for errors on pseudo words.* These findings imply that when the rehearsal and matching process yields words, the process is terminated more rapidly than when the consequence is unfamiliar to the child.

The process leading to the word read aloud, as we see it now, goes something like this. The child decodes the word into an auditory equivalent (forms an "auditory image"). He checks this image against his auditory memory of words he knows or sound patterns that he is familiar with. The

* The sign tests were run only on subjects with both types of errors. The mean subject latencies for real word type errors only was 2.2 seconds (n=12) and for pseudo word type errors only was 2.6 seconds. (n-9).

closer the match, the more quickly he says the word. Unfamiliar sound patterns may increase latency by leading to further decoding, rehearsal, matching, or confusion. What are the implications for the age differences in decoding that we have found? We doubt that there is much difference in the familiarity with words or the English sound patterns between seven and nine year olds. Their ability to decode the writing into corresponding sounds, though, is probably vastly different. We find, therefore, more errors at the younger age levels and especially long latencies for errors (unfamiliar sound patterns) by the second graders.

We emphasize that our formulation is tentative and subject to change as we accumulate more data. Several directions are visible, however. The next study systematically varies familiarity of sound patterns by presenting words differing in pronouncibility (Underwood and Schultz, 1960). Another idea is to confuse the auditory matching phase by feeding in sound during the exposure-response interval. If possible, we should like to use a list, in another study, in which the spelling to sound correspondences are simple but the resulting sound patterns are unfamiliar. Finally, we would expect more signs of rehearsal such as lip movements, EMG recordings from the larynx, and practice vocalizations (Flavell, 1965) during instances of unfamiliar words.

This formulation of the process of decoding and reading aloud may be summarized by an analogy to playing the piano. Some scales are more difficult to play than others. The mastery of the scales come from mastering the correspondences between written notes and finger movements. Likewise errors are recognizable by their degree of dissonance from a practised and anticipated musical sound pattern.

SUMMARY

This study was designed to test the usefulness of latency in reading words aloud as a response index to the process of reading. Children in the second, third, and fourth grades were shown a randomized list of sixteen words -- eight real and eight pseudo-words. The time they took to give a verbal response to the word after its exposure was measured. The results were as follows:

1. Children are highly consistent in their behavior across trials and between the two types of words within trials.
2. Younger children took longer to read the words than older children.
3. Latencies decrease over trials.
4. It takes longer to read pseudo-words than real words.
5. More errors in reading are made to pseudo than to real words.
6. Younger children make more errors than do older ones.
7. Latencies are longer to words read incorrectly than to ones read correctly.
8. When frequency of errors are controlled, there were no differences in the latencies of real and nonsense words read incorrectly, but for correct responses, real words were read more quickly.

These findings indicate the usefulness of response latency as a measure of reading. The results were interpreted tentatively according to a formulation which analyzes oral reading into the processes of decoding and matching to auditory memory.

APPENDIX A

Latency Means, Latency Standard Deviations, and Percentage of Errors-Both Trials Combined

	Real Words								Pseudo Words							
	Black	Stand	Grand	Bless	Grass	Storm	Check	Charm	Bland	Steck	Chass	Grack	Grerm	Chend	Blerm	Stess
All mean latency	1.51	2.09	2.32	2.75	2.76	2.78	3.05	4.11	2.78	3.72	4.13	4.14	4.25	4.64	4.88	5.39
s.d.	1.39	2.19	2.20	2.73	2.72	3.50	4.10	3.98	3.01	3.13	3.60	3.56	3.42	3.75	4.13	3.98
Percent Errors	9	9	24	31	20	15	35	27	37	63	63	52	93	44	53	62
4th grade																
mean latency	.92	1.00	1.67	1.80	1.91	.92	2.08	3.05	1.26	3.60	4.44	3.32	3.73	3.46	3.72	5.36
s.d.	.25	.35	2.61	1.42	1.61	.35	3.61	3.84	.56	3.83	4.42	4.52	4.65	3.40	3.71	4.71
Percent errors	0	6	16	14	5	3	11	11	34	36	41	52	66	36	47	41
3rd grade																
mean latency	1.12	1.74	1.46	2.46	1.88	2.64	2.36	2.55	2.70	3.17	3.17	3.33	4.14	4.88	4.60	5.09
s.d.	.39	1.41	1.12	2.60	1.11	2.86	3.52	2.67	3.28	2.00	2.49	2.88	3.28	3.86	3.69	3.85
Percent errors	3	6	11	25	11	8	30	11	39	69	66	52	94	36	47	61
2nd grade																
mean latency	2.48	3.52	3.84	4.00	4.50	4.77	4.70	6.74	4.39	4.40	4.79	5.77	4.88	5.58	6.31	5.78
s.d.	2.06	2.92	3.55	3.45	3.69	4.65	4.74	4.05	3.59	3.15	3.65	4.00	3.56	3.86	3.74	3.50
Percent errors	25	16	44	10	44	34	64	58	39	86	83	52	100	61	66	77

APPENDIX B

Analyses of Errors.*

We have classified errors two ways. The first classification is concerned with errors read as real words versus errors read as pseudo words. The second classification is concerned with the part of the word where an error is made.

Table 6 suggests that neither grade nor real vs. pseudo type words affect the proportion of errors given as real or pseudo (with the possible exception of the third grade's response to real words).

Table 6

Classification of Errors

	Real Words Grade				Pseudo Words Grade			
	2nd	3rd	4th	Combined	2nd	3rd	4th	Combined
Number of Errors	126	38	26	190	204	168	133	505
Percent read as real word:	50	37	48	47	47	50	50	49
Percent read as pseudo word:	38	55	38	42	42	37	36	40
Percent omitted:	12	8	15	11	11	13	14	11

Tables 7 and 8 suggest that the determinants of whether word errors are real, pseudo, or omitted have a lot to do with the word in question. In Table 7, the frequency of occurrence in the Thorndike-Lorge Juvenile list is inversely associated with the number of errors and number of

* We wish to thank Miss Susan Bostwick for her help in preparing this section.

Table 7

Classification of Errors - By Real Words

	Frequency occurrences Per million words*	Number Errors	Number real	Number pseudo	Number omitted
Bless	44.5	34	12	19	3
Charm	58.5	29	3	17	9
Grand	80.6	26	9	12	5
Check	92.8	38	16	11	11
Storm	122.0	5	7	5	
Grass	155.5	22	12	6	4
Black	220+	10	6	4	0
Stand	220+	10	6	3	1
Total		186	69	79	38

*Adapted from Thorndike and Lorge, 1944.

Table 8

Classification of Errors - By Pseudo Words

	# errors	# real	# pseudo	# omitted
Steck	69	50	5	14
Chass	69	46	10	13
Grerm	101	38	56	7
Bland	40	35	5	0
Grack	57	32	17	8
Cherd	48	19	20	9
Stess	65	16	33	16
Blerm	58	11	32	15
Total	507	247	178	82

pseudo-type errors. It is less clearly associated with omissions and real-type errors.

In Table 8, a case could be made for a relationship between the number of real-type errors and the number of letter changes needed to change the pseudo word to a real word. Inspection of the transcript sheds some doubt on this hypothesis.

Our second classification of errors involves breaking words down into three parts: initial (first letters); medial (next three letters); and final (last letter). A word can be categorized as correct (c) or incorrect (i) in any of these three parts. Thus, if the word BLACK were read "BLECK", it would be classified cic. These errors are summarized in Table 9.

Table 9

Errors by Part of Word

	Real Words					Pseudo Words				
		percent					percent			
	#	2nd	3rd	4th	Com	#	2nd	3rd	4th	Com
icc, ici, & iii	21	17.5	5.3	7.7	11.0	67	20.1	17.9	10.5	13.2
cic	34	18.3	13.2	23.1	17.9	125	18.6	26.2	32.8	24.8
cic (vowel only)	22	8.7	15.8	19.3	11.6	85	12.7	23.2	14.9	16.8
cci	5				2.6	21				4.2
iic	14	5.6	15.8	3.9	7.4	45	11.3	5.9	9.0	8.9
cii	72	38.1	42.1	30.4	37.9	110	30.4	13.7	18.7	21.8
omissions	22	11.9	7.9	15.4	11.6	53	11.0	13.1	14.2	10.5
total	190									

A qualitative analysis of the errors themselves may be seen in Table 9. The highest difference between the real and pseudo groups is in the proportion of errors made with the initial consonant. The Pseudo group of words has twice as many of this type of error proportionally than the real group. With this exception and that of the cii group which accounts for 39% of the real group errors and only 22% of the pseudo group errors, there is little difference between the types of errors made on both groups of words.

When the error types are examined by grade level (See Table 3) the following major differences may be noted: 1. That medial errors account for almost 50% of the errors for both the real and pseudo groups and when the medial error is combined with the terminal error the two account for almost two-thirds of all the errors for both groups. 2. With the exception of the 2nd Grade which remains constant, the proportion of cii errors in the real group is almost twice that of the Pseudo group. 3. The 3rd Graders show in almost every case the greatest amount of fluctuation in specific errors types employed. That is, they appear to have two distinct approaches, one for 'real' words and one for 'pseudo' words.

References

1. Edfeldt, A. W., Silent speech and silent reading. Chicago: Univer. Chicago Press, 1960.

2. Flavell, J. H. The function of private speech in children's thinking.. Paper read at Society for Research in Child Development Meeting, Minneapolis, March, 1965.

3. McLatchy, J. H. An oral-reading test as an appraisal of progress. Educ. Res. Bull. (Ohio State Univ.), 1949, 28, 230-239.

4. Thorndike, E. L., Lorge, I., The Teacher's Word Book of 30,000 Words. Bureau of Publications, Teacher's College, Columbia Univer., New York, N.Y., 1944.

5. Woodworth, R. W., & Schbsberg, H., Experimental Psychology, Henry Holt and Company, New York, 1954.

Studies of Oral Reading[1]

II. Pronounceability

Andrew J. Biemiller and Harry Levin

Cornell University

In an earlier study (Levin and Biemiller, 1965) we hypothesized that the reaction time period in reading a word aloud was taken up by the processes of translating the written symbols to sounds, matching these sounds to an auditory schema and saying the sounds. The schema, we suggested, was the memory for English-like sound sequences. There was some evidence, especially from an analysis of reading errors, that subjects rehearsed the decoded sounds in an attempt to bring them into line with the schema.

Verbal reaction time, to adopt Fraisse's (1964) term, will increase as a function of the difficulty of decoding the print to sound and of the acceptability of the sounds so generated. In practice, these two processes are inter-twined although it should be possible to disentangle them experimentally. For example, a series of letters predictable in their correspondences to sounds may decode to an odd sound sequence or a rare sequence of letters may yield familiar sounds.

The present experiment is concerned with verbal reaction times to word forms that differ in their pronounceability. We assume that the more unpronounceable words yield less familiar sound patterns. At the same time, the letter sequences of unpronounceable words violate the

[1] This research was supported by a contract with the U.S. Office of Education.

spelling pattern for English and so create recoding difficulties. In other words, the variation of pronounceability confounds both the ease of decoding and the predictability of sounds. Either process singly or in combination with the other should increase verbal reaction times.

Unpronounceable word forms require higher thresholds for perception (Gibson, Pick, Osser and Hammond, 1962). We have argued that perceptual measures also involve decoding to sound, although the evidence is only inferential. Fraisse (1964) argues categorically that the measurement of thresholds for words and the verbal reaction times to words are part of the same process. He reports a correlation of .81 between the two measures. When the familiarity of the words are held constant, the correlation is still a substantial .55. In fact, a correlation as high as .81 means that the two measures are practically interchangeable. More important, the degree of relationship implies that in the determination of the threshold at which the word is recognized, some central processing is taking place.

Specifically, we postulate that to a literate person, an array of letters (such as a word) is an automatic instigation to saying that word to one's self. As a case in point, in an experiment in which consonant sequences were presented visually and the recall was written, errors were predicted by acoustic confusability (Conrad, Freeman, and Hull, 1965) which means that Ss said the letters to themselves. Likewise, even in short term tachistoscopic exposures of letter groups, the viewer says them to himself. The pronounceable groups are decoded more easily, stored and retrieved more efficiently. The process is the same as verbal reaction time and hence the similarity of findings.

Our speculation that the decoded sound is matched against an auditory image gains support from two studies by Smith (1965 a, and b). Recognition of words as sequences of digits is facilitated when the correct response is heard prior to the visual display. We interpret these findings to mean that the auditory input creates a schema against which the decoded visual stimulus is matched. Smith (1965b, p. 158) puts it "Present results support the interpretation that the facilitation of recognition found is due to a perceptual interaction effect, namely, an effect of hearing on seeing."

Method

Subjects. 36 children, nine boys and nine girls from a third and a fourth grade were randomly selected.[1,2] Each of the classes had children with a range of abilities.

Stimulus Materials. Twenty-four pseudo-words were taken from a study by Gibson, et al. (1962). These words had been generated from spelling patterns described by Hockett (1960). Each of the words had been rated for pronounceability by Underwood and Shultz's (1960) method. The words and their pronounceability ratings are given in Table 1.

Procedure. Each child was informed that he would be shown words on a screen and asked to read them. He was told that the words had been made up so he did not have to feel bad if he did not know them. The words were projected on a screen three feet from the S. The projected words were about six inches long by two inches high. After the child gave his response, there was a two second interval before the next word was presented. If

[1]We wish to thank the staff of the Belle Sherman School, Ithaca, N.Y., for their help.

[2]Non-native speakers of English were eliminated before selection.

Table 1. Stimulus Words with Pronounceability Ratings, Adult Mean Errors, and Mean Latencies in Seconds by Grade.

	Pronounceability[1] Rating	Number[2] Written Errors	Latency 3rd & 4th Grades
Number of Ss	165	25	36
Stimulus Words			
DINK	1.1	7	1.3
CODS	1.3	18	1.2
CLATS	1.5	34	1.6
VUNS	1.5	15	1.9
GLOX	1.6	8	1.8
GRISP	1.6	12	1.8
FUNTS	1.6	6	2.2
SLAND	1.7	5	1.6
SULB	1.9	30	2.5
TILMS	2.2	28	2.9
BLORDS	2.3	71	1.7
BESKS	2.3	46	1.8
FRAMB	2.4	13	1.8
PREENT	2.9	69	3.0
BLASPS	2.9	58	2.9
GLURCK	2.9	74	3.0
BRELP	3.0	34	2.8
QUEESK	3.0	80	2.5
KLERFT	4.2	88	3.0
TIRPTH	4.5	83	3.2
PRILTHS	4.6	97	3.3
TRILFTHS	5.3	112	3.9
SMAWMP	6.2	122	3.2
DRIGHK	6.3	89	4.2

1, 2) From Gibson, E.J., et. al. (1962).

the child made no response to a word in 15 seconds, he was asked if he would like to go on to the next one. He invariably did. Similarly, if the child suggested going on, the next word was displayed.

The Ss' responses were tape recorded and, for analysis, the tapes were played through a rectifier connected to a pen-writing recorder. The pen was deflected by sound on the tape so that we measured the period of time from the exposure of the word to the onset of the final pronunciation given by the S. That is, repetitions, false starts, stutters, etc., were included in the reaction time period.

Results

The 24 words were divided at the median of pronounceability and an analysis of variance calculated according to the classifications, pronounceability, sex of Ss, and grade level.[1] The means are given in Table 2.

Table 2. Mean Latencies in Seconds by Rated Pronounceability, Grade and Sex.

	(N)	More Pronoun. 12 Words	Less Pronoun. 12 Words	All Words
All Grades and Sexes	36	1.58	2.37	1.98
Third Grade, all	18	1.93	3.26	2.60
boys	9	1.75	2.70	2.22
girls	9	2.12	3.83	2.97
Fourth Grade, all	18	1.70	2.85	2.28
boys	9	1.41	2.05	1.73
girls	9	1.99	3.65	2.85

[1] The observations were transformed for this analysis according to the formula, $Y = \log (X-0.4)$, Y is the transformed score and X = latency in seconds. This transformation is discussed in Woodworth and Schlosberg (1954, p. 39).

The single significant source of variation was pronounceability ($F = 35.2$; $df = 1, 32$; $p < .001$). No other main effects or interactions are significant. It is clear that children take longer to read the less pronounceable words.

The main hypothesis of this study concerns the correlations between the latencies to the various words and their pronounceability. Latencies were calculated for each word, across Ss. The results appear in Table 3.

Table 3. Correlations (rho) between Mean Response Latencies and Rated Pronounceability per Word by Grade and Word Length.

		Grade		
	(N)	3rd	4th	Combined
All words	24	.89	.92	.87
Four and five letter words	13	.65	.74	.61
Six, seven and eight letter words	11	.89	.93	.88

The correlations are all significant and clearly indicate that the harder a word is to pronounce, the longer is the interval from the exposure of the word to the verbal response. However, the less pronounceable words tend also to be longer (see Table 1). To control for length, separate correlations were calculated for short and long words. As can be seen in Table 3, the association between pronounceability and latency holds even when word length is controlled.

The comparisons between the latency data and the error scores for these words reported by Gibson, et. al. are interesting, especially in light of Fraisse's findings. First, though, the differences in the experiments should be clear. Gibson and her co-workers exposed two groups of pseudo words: 25 pronounceable (P) and 25 unpronounceable (U) items. We are using her P list which itself includes a range of pronounceability as can be seen from their ratings in Table 1. Each list was tachistoscopically presented five times at exposure speeds 30, 50, 100, 150, 250 m. sec. The S wrote what he saw. The number of correct reproductions was summed for each word across Ss and exposures.[1]

In our data the mean latency for each word was tabulated across Ss. The correlations (rho) between the two scores from the two experiments are given in Table 4. The overall correlation is .86. However, note the

Table 4. Correlations (rho) Between Mean Response Latencies and Mean Errors (Gibson, et. al., 1962) per Word.

		Grade		
	(N)	3rd	4th	Combined
All words	24	.79	.80	.86
Four and Five letter words	13	.36	.08	.27
Six, seven and eight letter words	11	.75	.84	.79

differences in the degree of relationships which are a function of the length of the words. For the shorter, four and five letter words the correlation between errors and reaction times is negligible, whereas, for longer words the correlation is substantial, rho = .79. It is tempting to think that length of word influences the relationship between the measures of accuracy of recognition at the threshold and reaction time. However, it is more likely that the differences in the two sets of cor-

[1] For simplicity in exposition we have calculated the number of errors rather than the number correct from Gibson, et. al's data. (Compare their Table 1 with our Table 1).

relations is due to the restricted range of errors scores to the shorter words. The range of errors to the four and five letter words is 41; to the longer words, 88.

Fraisse (1964), it will be recalled, reported a correlation of .81 between verbal reaction time and recognition threshold. This correlation is very close to the .86 which we find. His stimuli were 25 real words representing a range of frequency in French. The words contained 5 or 6 letters. His correlation is between the median recognition threshold and reaction time to reading the word. Both measures were taken on the same Ss.

It appears that there are a complex of responses which are highly interrelated. The correct recognition of words at threshold, the level of the threshold and the time it takes to read a word when the stimulus is available ad lib. all correlate highly. These findings hold regardless of whether the Ss are adults or children, whether the stimulus words are real or nonsense. The correlation of any of the measures varies with the pronounceability of the word.

Pronounceability is a measure of the ease or difficulty with which the groups of letters can be sounded. We have argued in this and an earlier study that all of these measures have a common process at their base: the decoding to sound and in the case where the procedure requires a verbal response, the private rehearsal of the decoded sound sequence against an experience-generated schema of "acceptable" sounds.

We agree with Gibson that the results are not a reflection of response biases, in the usual sense. The various measures we have discussed are not primarily reflections of experiences with letter or sound sequences (frequency), but of experiences with the relationships between letters and sounds and further a superordinate development which we might call a language sound schema. Such a schema should predict whether a speaker will accept a novel sound sequence as belonging to a language; that is, as being congruent with the schema.

Summary

Thirty-six third and fourth grade children read a list of 24 pseudo words which varied in pronounceability. It took the children longer to read the less easily pronounceable words. In addition, the errors in reproducing these words after rapid tachistoscopic exposures correlate highly with the reaction times to read the words.

References

Conrad, R., Freeman, P.R., and Hull, J. Acoustic factors vs. language factors in short-term memory. Psychon. Sci., 1965, 3, 57-58.

Fraisse, P. Relations entre le seuil de reconnaissance perceptive et le temps de reaction verbale. Rev. psychol. Franc., 1964, IX, 77-85.

Gibson, Eleanor J., Pick, Anne, Osser, H., and Hammond, Marcia. The role of grapheme-phoneme correspondence in the perception of words. Amer. J. Psychol., 1962, 75, 554-570.

Hockett, C.F. Analysis of English spelling. Mimeo. paper, June, 1960.

Levin, H. and Biemiller, A.J. Studies of oral reading, I. Words vs. pseudowords. Mimeo paper, 1965.

Smith, W.M. Visual recognition: facilitation of seeing by saying. Psychon. Sci., 1965a, 2, 57-58.

Smith, W.M. Visual recognition: facilitation of seeing by hearing. Psychon. Sci., 1965b, 2, 157-158.

Underwood, B.J., and Schultz, R.W. Meaningfulness and verbal learning. Philadelphia: Lippincott, 1960.

Woodworth, R.W., and Schlosberg, H. Experimental Psychology. New York: Holt, 1954.

Preliminary Draft

January, 1966

Studies of Oral Reading[1]

III. Contingent versus Non-contingent Spelling Patterns

Harry Levin and Andrew J. Biemiller

Cornell University

In previous studies we have examined two factors which affect that part of the reading process in which graphic materials are decoded to their oral counterparts: real compared to pseudo-words (Levin and Biemiller, 1965) and pronounceability (Biemiller and Levin, 1965). In the present study we begin to examine the effects which certain correspondences between spelling and sound have on reading aloud. The correspondence system in English is complicated. The pronunciations of any vowel and most consonants depend on their environments in the word. The correspondences of clusters of letters compared to single letters are fairly predictable (Hockett, 1963; Venezky, 1965).

A competent reader of English has been taught or has induced most of these complex relationships. He 'knows', for example, that the pronunciation of the letter c in the initial position of a word depends upon the following letter: ca, ce, ci, co, cu, ch. This implies an added step in the processing of the contingent instances compared to words in which the first letter is invariably said the same way; e.g., d, m, l.

We asked ourselves how the necessity for additional information--the subsequent letter--would influence the verbal reaction time for reading words of these types. The prediction is not obvious. The initial letter itself is indeterminate, but the cluster, ca or ce, for example,

[1]This research was supported by a contract with the U.S. Office of Education We thank Miss Susan Bostwick for her help with the analysis of errors.

is perfectly determined. Gibson (1962) suggests that readers learn spelling patterns as units based on their relationships to invariant corresponding sounds. This theory leads to the prediction that words with contingent relationships, as we have defined them, will be read as quickly as words whose initial letters are not dependent for decoding on their environments.

However, one may look at the formation of such higher order units developmentally. It is clear that mature readers do not process words letter by letter. Beginning readers process in the order, beginnings, ends, middles of words (Marchbanks and Levin, 1965). Therefore, we expect the two-stage processing of initial letters will result in longer reaction times than the non-contingent instances, for younger readers. Besides, it is reasonable that two letter units take longer to form than single letter units. (Units here mean the level of maximal predictability from spelling to sound.) In the course of learning to read, one level of correspondences may be well established while the other is still infirm.

In this study, we also examine the effects of the various spellings on the frequency and types of reading errors.

Method

This study was very similar in method to our first two studies. Subjects read aloud words presented on a screen. Their responses were tape-recorded. The interval between presentation of the word on the screen and the subject's oral response constituted the verbal reaction time. Errors were analyzed from a transcript of the tape. A more detailed description of this procedure, and of our measuring of verbal reaction bine will be found in Levin and Biemiller (1965).

Experimental Design and Stimulus Materials.

Three lists of words were drawn up. The first two lists, contingent-common and contingent-uncommon, used words beginning with an initial consonant whose pronunciation was dependent upon the following letter. Contingent-uncommon words used the less common pronunciation of the initial letter, (e.g., celt). Contingent-common words had the more frequent pronunciation, (e.g., colt). The third list, non-contingent, used words having all the same letters as contingent-uncommon words except the first. The first letter would be an initial consonant whose correspondence is invariant, (e.g., belt).

Three initial consonants were used in the contingent lists: c, g, and k. The letter c appeared in its /s/ form in the contingent-uncommon list and its /k/ form in the contingent-common list g appeared in its /j/ and silent, (gn), forms in the contingent-uncommon list and its /g/ and /g/+ glide forms in the contingent-common list: k appeared in its silent (kn) form in the contingent-uncommon list and its /k/ form in the contingent-common list. Three words were used for each form. Where possible, contingent-common and non-contingent words were selected on the basis of being less frequent in the Thorndike-Lorge Juvenile List (1944) than the contingent-uncommon words. The entire list of words and Thorndike-Lorge frequencies appear in Table 1.

(Insert Table 1)

Subjects.

54 children from the West Hill Elementary School in Ithaca, New York were tested.[2] Eighteen fourth grade, eighteen third grade, and eighteen second grade children were randomly selected from one classroom in each grade. Due to extreme imbalances in the distribution of the sexes in the classrooms, no effort was made to balance the grades for sex.

Results

1. Verbal Reaction Time. The main hypothesis of this study concerns verbal reaction times to words beginning with "contingent" spelling patterns versus words not beginning with contingent spelling patterns. Results appear in Table 2.

(Insert Table 2)

An analysis of variance performed on these data indicated that the main effects (type of initial pattern, grade, and letter group) were all significant at the .01 level as was the interaction between initial pattern type and letter group. (A summary of the analysis appears in Appendix A)

Inspection of the means in Table 2 shows that our hypothesis is only partially confirmed. The mean reaction time for the contingent-common words does not differ significantly from the mean reaction time for non-contingent words. Contingent-uncommon differs from the other two groups. These points will be taken up later.

2. We wish to thank the principal and teachers for their considerate help.

A second dependent variable, errors, was expected to show the same pattern of effects as verbal reaction time. Error percentages are shown in Table 3.

(Insert Table 3)

An analysis of variance revealed that the effects for percentage of errors are identical with those obtained for verbal reaction times. Again, responses to contingent-common words did not conform to our expectations. (A summary of the analysis appears in Appendix A.)

The significant interactions obtained between initial pattern types and letter groups are caused by changes in the relative effects of contingent-common and non-contingent patterns on the different letter groups. These changes are probably a function of the particular words used to represent these types of initial spelling patterns.

Latencies for Correctly Read Words. In our first study (Levin and Biemiller, 1965) we found that words read incorrectly required longer verbal reaction times. This suggests that the significant latency findings presented in Table 2 are artifacts of the greater number or errors made to words with less frequent contingent initial spelling patterns. In Table 4 are presented the frequency of correctly read words and the latencies to these words.

(Insert Table 4)

Although the overall means for each grade conform to the pattern of high latencies for contingent-uncommon words and low reaction times for contingent-common and non-contingent words, inspection of letter group and letter group by grade values indicate that several letter group by grade sets do not conform to the patterns observed in Tables 2 and 3.

On the other hand, the fourth grade Ss conform most clearly to the pattern observed in Table 2. Fourth graders make fewest errors. Hence we may hypothesize that the contingency effect comes out most clearly with competent readers.

The data in Table 4 suggested a further analysis. Different numbers of subjects are providing our estimates of mean latency per word for the correctly read words. These varying frequencies could seriously bias the results. For example, we find that fourth graders made about 40 fewer correct responses to contingent-uncommon words than to the other two types of words. The same subjects who failed to make correct responses to the contingent-uncommon words may have been generally poorer readers and hence taken longer in processing those words they could read. This would result in increased mean latencies to the contingent-common and non-contingent lists. We decided, therefore, to examine differences in the reaction times of subjects who gave correct responses to both contingent words and associated non-contingent words. Three comparisons were made, contingent-uncommon to contingent-common, contingent-uncommon to non-contingent, and contingent-common to non-contingent.

In comparisons involving contingent-common words, groups of words must be compared because single pairs of words are not matched. Thus, comparisons are made between mean latencies for words beginning with c and a vowel, g and a vowel, etc. Unfortunately, there are not enough second and third graders with groups of correct responses to make comparisons worth while. Results for the fourth graders are shown in Table 5.

(Insert Table 5)

Again we find the same pattern of high latencies to contingent-uncommon words and relatively low latencies to contingent-common and non-contingent words. Analysis of this small sample by t-tests indicates that the difference between mean latencies to contingent-uncommon words and the other two types are just short of significance. ($t = 1.85$ and 1.64 with 23 degrees of freedom)

These various analyses may be summarized as follows: latencies are highest to words that have contingent-uncommon initial spellings; common contingent spellings do not differ from control words whose initial spellings have a one-to-one correspondence to sounds.

The Nature of Errors. The finding is well established that reading errors occur least frequently at the beginnings of words, somewhat more often at the ends of words and most frequently in the middle of words (Jensen, 1962; Marchbanks and Levin, 1965). In Table 6 it can be seen that the errors made to the contingent-common words and to the control words follow this pattern. For our purposes we have analyzed only the initial errors compared to all other types. However, the errors made to the contingent-uncommon words diverge markedly from the expected pattern. Of the total of 368 errors, including omissions, 282, or 76%, involved the initial parts of the words. These 282 errors are broken down further in Table 7 to show that 107 mistakes involved only the initial part of the words. Also, it is important to note that 243 of the 282 errors involved the children saying the contingent-common form to the contingent-uncommon spelling; that is, kell was said to the written word cell. In summary, then, the nature of the errors was atypical to one set of words and below we shall discuss the implications of these findings for our original hypotheses.

(Insert Tables 6 and 7)

Discussion

Two alternative hypotheses were germane to this experiment. The first stated that for children contingent spelling patterns would require more processing time than those which can be translated to sound without concern for the environments of the letters. The second hypothesis was that the contingent forms become higher order units as a results of the invariant relationship of the letter group to sound. Two types of contingent spelling patterns were contrasted with control words. The contingent-uncommon forms were words whose initial letters could be decoded only by taking into account the second letter of the sequence. The contingent-common form also followed this pattern but were more frequent in English than the former list. Our results indicate that the time taken to decode and say the words on the various lists differed. The contingent-common and non-contingent words were decoded with about equal speed and both of these types of words were read considerably more rapidly than contingent-uncommon words. These results hold even when the instances were limited to those words read correctly.

To this point neither of the hypotheses are clearly confirmed. As a matter of fact, the most conservative interpretation is that these children are responding to a single pattern which they had learned (the common one) and were having difficulty with the pattern with which they had little experience. In other words, the results fit the ubiquitous finding that reaction time is related to the frequency of the word.

The nature of the errors that children made raises some interesting speculations. Reading errors are rarely made at the beginnings of words. Yet the most usual form of errors made to the contingent-uncommon word occurred in this position of the word. In fact, children transposed the contingent-common pronunciation to these other forms. This means that the children had not formed a higher order unit of the initial two letters in the word but were responding only to the initial letter.

It seems to us that the children were over-trained on one form and without the contrastive introduction of the complementary form were not able to derive the higher order unit. Our pedagogical recommendation, therefore, would be that the two contrastive forms be introduced simultaneously to the child.

Summary

Spelling-to-sound correspondences may be direct or contingent. In the latter case, pronunciation of one letter depends upon its environment (cent vs. cant). The present study compared response latencies and errors to three intermixed lists of twelve words each. The first list contained words beginning with c, g, or k in some of their less common pronunciation, (e.g., cell, gem, gnaw, and knelt). The second list contained words also beginning with c, g, and k, but in their more common pronunciations, (e.g., colt, gum, grab, ketch). The third list contained words differing from the first list only in having initial letters with invariant spelling-to-sound correspondences, (e.g., dell, hem, flaw, dwelt). Longer latencies and more errors to the first list but no differences between the other two lists were observed in a sample of fifty-four children drawn in equal numbers from the second, third, and fourth grades. Results were most clear-cut for fourth graders. Analysis of the errors showed that most of the errors made on the first list consisted of giving the more common pronunciation of the first letter.

References

Biemiller, A. J. & Levin, H. Studies in oral reading, II. Pronounceability, Mimeo. Cornell University, 1965.

Gibson, Eleanor J., Pick, Anne, Osser, H. & Hammond, Marcia. The role of grapheme-phoneme correspondence in the perception of words. Amer. J. Psychol., 1962, 75, 554-570.

Hockett, C. F. Analysis of English spelling. In Levin, H., et al. A basic research program in reading. Final Report on Project No. 639 to the U.S. Office of Education, 1963.

Jensen, A. R. Spelling errors and the serial position effect. J. educ. Psychol., 1962, 53, 105-109.

Levin, H. & Biemiller, A. J. Studies of oral reading, I. Words vs. pseudo words. Mimeo. Cornell University, 1965.

Marchbanks, Gabrielle & Levin, H. Cues by which children recognize words. J. educ. Psychol., 1965, 55, 57-61.

Venezky, R. L. A study of English spelling-to-sound correspondences on historical principles. Unpub. Ph.D. Thesis, Stanford University, 1965.

Table 1

Words and Thorndike-Lorge Frequencies

Contingent-uncommon	Frequency	Contingent-common	Frequency	Non-contingent	Frequency
CELL	212	CARL	21	DELL	12
CELT	2	COLT	120	BELT	260
CENT	700	CAN'T	?	BENT	250
GEM	109	GUM	31	HEM	30
GERM	28	GALL	15	TERM	340
GENT	7	GOLF	1	DENT	5
GNAT	15	GRAB	57	BRAT	10
GNAW	110	GLEN	22	FLAW	7
GNASH	28	GLAND	17	TRASH	5
KNELT	83	KETCH	6	DWELT	121
KNIT	137	KICK	214	FLIT	62
KNOT	115	KILN	20	BLOT	19

Table 2

Mean Verbal Reaction Times in Seconds

by Contingency, Grade, and Letter Group.

	Initial Pattern Type			
	Contingent		Non-contingent	All Words
	Uncommon (CENT)	Common (COLT)	(DENT)	
All Words and grades	4.17	3.13	3.01	3.44
Grade				
4th	2.66	1.88	1.70	2.08
3rd	4.16	3.14	3.16	3.49
2nd	5.68	4.38	4.19	4.75
Letter Group				
c-vowel	3.46	2.64	2.20	2.77
g-vowel	3.85	2.49	3.00	3.11
g-consonant	5.15	3.49	3.24	3.96
k-consonant	4.21	3.91	3.61	3.91

Table 3

Mean Percent Errors by Initial Pattern Type, Grade and Letter Group

		Initial Pattern Type			
		Contingent		Non-contingent	All
	N	Uncommon	Common		
All words and grades	54	52.0	28.9	29.5	36.8
Grade					
4th	18	35.1	13.9	15.3	21.4
3rd	18	55.6	34.7	33.8	41.4
2nd	18	65.3	38.0	39.3	47.5
Letter Group					
c-vowels	54	46.9	24.1	19.1	30.0
ge-vowel	54	54.9	21.6	25.9	34.2
g-vowel	54	64.8	35.2	24.1	41.4
k-vowel	54	4.14	34.6	48.6	41.6

Table 4

Mean Latencies for Correct Responses and Number of Correct Responses

X Initial Spelling Pattern, Grade and Word Group

	Initial Spelling Pattern							
	Contingent				Non-contingent		All	
	Uncommon		Common					
	no. correct	mean latency	no. correct	mean latency	no. correct	mean latency	no. correct	mean latency
All words	285	2.64	446	2.35	436	2.28	1167	2.42
Grade								
4th	144	2.22	166	1.56	183	1.65	513	1.51
3rd	97	2.41	142	2.28	143	2.32	382	2.34
2nd 1/	44	3.52	118	3.21	110	2.88	272	3.17
Group								
c-vowel	83	3.00	121	1.72	125	2.34	329	2.35
g-vowel	66	1.95	125	1.99	124	1.97	315	1.97
g-consonant 1/	49	3.58	102	2.46	119	2.20	270	2.75
k-consonant	85	2.34	98	3.11	75	2.89	258	2.78

1/ No correct second grade responses for g-consonant contingent-uncommon.

Table 5

Approximate[1/] Mean Latencies by Initial Pattern Type and Letter Group Among Fourth Grade Subjects Reading Words Correctly.

	Contingent-uncommon	Contingent-common	Non-contingent
c plus vowel	3.18	2.25	2.44
g plus vcwel	3.36	2.79	3.27
g plus consonant	5.85	2.92	2.93
k plus consonant	4.44	2.52	3.14
All words	3.73	2.44	2.90

[1/] Values vary slightlv from those presented here depending on which comparison is being made. Not all subjects read every subgroup correctly. Values presented here represent estimates of latencies based on the maximum number of observations used in any comparison. Overall differences in estimates of mean latencies by initial pattern type never differed by more than 0.2 seconds.

Table 6

Number of Errors of Different Types X Initial Patterns

	Initial Pattern			
	Contingent		Non-contingent	All
	Uncommon	Common		
All errors	368	200	210	778
Initial errors	282	29	9	320
Opposite Conting. Form	243	10	--	--
Other	29	19	--	--
Non-initial Errors	27	138	168	333
Ommissions	59	33	33	125

Table 7

Location of Error X Initial Pattern Type

	Initial Pattern			
	Contingent		Non-contingent	
	Uncommon	Common		
Error on initial spelling pattern only.	107	11	2	120
Error on initial and non-initial spelling patterns.	175	18	7	200
Error on non-initial spelling patterns only.	27	138	168	333
	309	167	177	653

Appendix A

Summary of Analysis of Variance-Latencies

Source	SS	df	MS	F	p
Grades	101.2330	2	50.6163	6.735	.005
Error 1	383.2878	51	7.5150		
Types	28.3182	3	9.4394	21.453	.001
Types X Grades	5.7142	6	.9524	2.164	.10
Error 2	67.3090	153	.4400		
Contingencies	24.4929	2	12.2465	23.416	.001
Contingencies X Grades	2. 6397	4	.6599	1.262	ns
Error 3	53.3756	102	.5230		
Contingencies X Types	8.8905	6	1.4817	3.394	.005
Contingencies X Types X Grades	2.7880	12	.2323	.532	ns
Error 4	133.6006	306	.4360		
Words within	64.5702	24	2.6904	7.032	.001
Residual Error	486.6673	1272	.3826		

Summary of Analysis of Variance-errors.

Source	SS	df	MS	F	p
Grades	24.0751	2	12.0376	5.019	.01
Error 1	122.3104	51	2.3982		
Types	4.6723	3	1.5574	10.623	.001
Grades X Types	1.4311	6	.2385	1.627	.20
Error 2	22.4244	153	.1466		
Contingencies	22.5473	2	11.2737	42.350	.001
Contingencies X Grades	.3508	4	.0877	.329	ns
Error 3	27.1574	102	.2662		
Types X Contingencies	11.2181	6	1.8697	13.932	.001
Types X Contingencies X Grades	.7850	12	.0654	.487	ns
Error 4	41.0525	306	.1324		
Words Within	25.3456	24	1.0561	9.034	.001
Residual Error	148.6543	1272	.1169		

Preliminary Draft 146
November, 1965

Studies of Oral Reading:

IV. Homographs vs Non-Homographs

Harry Levin and Boyce L. Ford[1]

Cornell University

There are various classes of cues, or discriminative stimuli, in words which signal the manner in which the spelling is decoded to sound. One-to-one relationships between the letters or letter groups and sound are the simplest instances. For many words, the values of the letters are dependent on spelling patterns involving the letter(s) in question as well as other letters in the word. In turn, the contingencies may be adjacent to or remote from the letter, e.g. knive, cent, mate (Levin & Biemiller, in preparation). For some words, the written word itself is inadequate for decoding so that the reader must use cues usually in the sentence but always outside of the word (Levin, in press). These words are homographs: written forms that in themselves may be decoded to two or more words. For example, sow is pronounced either /so/ or /saU/ and present can be pronounced /present/ or /presènt/.

Homographs are particularly useful in studying the influence of "context cues" on reading. The use of context is often stressed in the pedagogy of reading as one set of responses children might make to understand what they are reading. Presumably, teachers advise the children to use the sense of what they are reading to decode a difficult word. It seems that the children are taught to use their knowledge of dependencies in the language to decode and make sense of words that elude them.

[1]The study was supported by funds from the U.S. Office of Education. We wish to thank Mary Beckwith for her help with all aspects of the study.

If the child can read "He ate bread and _____", it is not difficult to guess that the next word is butter. On the other hand, it might be "jam" and the child with knowledge only of context-based guessing, to take an invidious view of this teaching procedure, would often be wrong. On the other hand, if the reader can bring to bear a variety of skills, such as facility with spelling-sound correspondences as well as sensitivity to context, the task of reading is advanced.

It is obvious that homographs permit us to study the pure role of context in decoding and we have designed several studies to show this role. Homographs alone may be decoded in several ways. This process of decoding should take longer than the decoding of non-homographs. Further, the latency of decoding should reflect the reader's greater familiarity with one over the second possible rendition of the homograph. In a sense, the latency of decoding should be shortened if one form is much closer to the top of the deck than the other. When the two forms are side by side in the deck, the latency of translating the written word to sound should be extended -- a kind of cognitive conflict between two equal but competing tendencies. In any case, though, we expect homographs to take longer to read than single translation forms.

The reading of unadorned homographs gives us the baseline against which to compare this reading when other cues are available. The second study, then, studies homographs in grammatical frames and the third involves reading a discourse containing a member of the ambiguous forms.

METHOD

Subjects. The 36 Ss, half male and half female, were high school students taking part in the Cornell University Summer Advanced Placement Program. This program is designed for students with high aptitude who are about to start their senior year in high school. The selected students attend the University's regular six week summer session.

Most of the records available to us contained the following information: I.Q. scores, high school grade average, Preliminary Scholastic Aptitude Test (PSAT) scores, age, and, at the end of the summer, the grades received for the summer courses at Cornell. The sample is described in Table 1.

Table 1. Description of Subjects

	Male	Female	Overall
Age	16.87	16.85	16.86
I.Q.	130.73	137.06	134.66
PSAT			
Verbal	57.00	57.13	57.07
Math	57.00	59.31	58.23
High School Grade Ave.	91.93	93.44	92.76
Cornell Grade	78.27	79.89	79.15

Word Lists. The stimuli consisted of 48 words, 24 homographs each paired with a non-homograph control word. Each pair was matched for (1) number of letters, (2) number of syllables, (3) same initial consonant, and (4) frequency as defined by the Thorndike and Lorge G Scale. The word pairs are presented in Table 2. The 48 words were randomized and presented to Ss, though in the same order for all Ss.

Table 2. Word Pairs, Thorndike & Lorge G Scale Frequencies, and Latencies

Homograph			Non-Homograph		
	Freq.	Latency		Freq.	Latency
BOW	50	.759	BUD	30	.658
PROTEST	50	.683	PARTNER	32	.663
TEAR	100	.738	TASK	50	.641
OBJECT	100	.625	OFFICE	100	.747
PERFECT	50	.762	PACKAGE	38	.626
UNIONIZED	1	1.091	ULTIMATUM	2	.83[illegible]
MINUTE	100	.709	MANAGE	50	.678
WIND	100	.639	WOOL	50	.610
LEAD	100	.698	LAST	100	.60[illegible]
ARITHMETIC	9	.692	ACCURATELY	8	.737
DOVE	19	.682	DOCK	16	.633
INVALID	9	.708	ILLEGAL	5	.6[illegible]7
BASS	7	.799	BURR	4	.6[illegible]9
PRESENT	100	.732	PICTURE	100	.656
SOW	26	.727	SHY	21	.580
DOES	100	.694	DARK	100	.639
REFUSE	100	.654	RATHER	100	.618
LIVE	100	.662	LAMP	50	.618
READ	100	.655	ROAR	50	.586
DESERT	50	.777	DAMSEL	36	.703
WOUND	50	.719	WRECK	41	.639
INTIMATE	24	.810	IMPRISON	12	.750
RECORD	100	.664	RABBIT	43	.603
PERMIT	50	.716	PLANET	34	.673

Experimenters. The study was conducted by two Es, one a male (E_1) and the other female (E_2). Each E tested nine male and nine female Ss.

Procedure. Each word was typed on a 2 x 2 slide in upper case pica type and presented with a Kodak Carousel Projector. Ss were tested individually and their responses were tape recorded. The Ss were seated ten feet from the screen and three feet directly behind the projector. The lens of the projector was six feet from the screen so that the image on the screen was 2 1/2 inches high. Ss held the microphone into which they spoke.

The subject was given the following instructions: "I am going to project some words on this screen. The words will be the type you can use in everyday conversation. As a word appears, please say it." The subject was then shown two demonstration words (MAN & SAY) which were displayed at the pace used in the experiment. If S had no questions at this point, the 48 word list, was presented and completed without interruption. E controlled the appearance of each slide.

The tapes were later played through a rectifier into a Brush Recorder. Auditory input from the recording activated a recording pen producing a visual representation on graph paper. When no sound was present the pen drew a straight line. By measuring the length of this line between onset of the visual stimulus[2] and onset of the subjects vocal response, it was possible to determine the latency to each stimulus word. All false starts, vocal segregates, etc. were included as a part of the response latency. Latencies were recorded in milimeters and later transformed according to the following formula: $x = \log (y - 10)$,

[2]Onset of stimulus was indicated by a characteristic sound made by the projector when a slide dropped into place.

where x = transformed score and y = latency in milimeters. Following analysis, latencies were converted into seconds by the formula; x = .04 (y), where x = latency in seconds and y = latency in milimeters.

RESULTS and DISCUSSION

The transformed scores were first analysed according to the classifications, Experimenter, Sex of Ss, and Homograph--Non Homograph. Our pretest data were unclear as to whether the reaction times of boys or girls was faster. Since there was only one E of each sex we cannot attribute the experimenter effects to the sex of the E, although, as will be seen below, the interaction between E and sex of Ss is suggestive.

Table 3. Mean Latencies in Seconds

	$Male_E$	$Female_E$	$Male_E$	$Female_E$	Overall
Homo	.785	.694	.738	.680	.724
Non	.720	.634	.661	.634	.662
Overall	.752	.664	.699	.657	.693

Subject Means

	Homo	Non	Overall
Male	.740	.677	.708
Female	.709	.647	.678

Experimenter Means

	Homo	Non	Overall
Male	.762	.690	.726
Female	.687	.634	.660

Table 3 presents the mean reaction times divided in various ways and Table 4 summarizes the analysis of variance. Latencies to all types

Table 4. Summary of Analysis of Variance

Source	df	ss	ms	F
BETWEEN	35	3966.20		
Subj. Sex	1	338.00	338.00	7.567**
Exp. Sex	1	2005.60	2005.60	44.891***
Subj. x Exp.	1	193.30	193.30	4.327*
error (b)	32	1429.30	44.67	
WITHIN	36	13535.30		
Homo	1	1404.50	1404.50	3.705*
Homo x Exp	1	0	0	
Homo x Subj	1	.20	.20	
Homo x Exp x Subj	1	24.60	24.60	
error (w)	32	12130.80	379.09	
TOTAL	71	17501.50		

* $p < .05$

** $p < .01$

*** $p < .001$

of words were briefer for the female than the male Ss and were faster also with the female compared to the male E. Further, the influence of the female E on the rapidity of boys' and girls' responses was roughly equal, whereas boys were slower than girls with the male E.

One is tempted to think of the rapidity of response in a test situation as an index to achievement concerns in the S (Gallwey, 1958). However, sex differences in need achievement are not at all clear. Secondly, the difference between boys and girls may effect the oft noted superiority of girls on verbal tasks.

Most important, it will be noted that the latencies in reading homographs is significantly longer than when reading the control words.

There are no significant interaction between the types of words and either the E or the sex of Ss. Since our focus of concern is with the latencies of reading homographs we collapsed the Experimenter classification and recalculated the analysis of variance which is summarized in Table 5. The main effects attributable to the sex of S and to the nature of the word are significant.

Table 5. Summary of Analysis of Variance

Source	df	ss	ms	F
BETWEEN	35	2735		
Subj. Sex	1	666	666.00	10.94**
error (b)	34	2069	60.85	
WITHIN	36	4138		
Homograph	1	2016	2016.00	33.18***
Subj. X Homo.	1	56	56.00	
error (w)	34	2066	60.76	
Total	71	6873		

** $p < .01$

*** $p < .001$

The purpose of this study, it will be recalled, was to test the sensitivity of verbal reaction times to reading words which, with no additional information, can be read in two different ways. Clearly, this circumstance increases decision time substantially, about .06 seconds.

Reaction times to homographs may be influenced by the frequency of incidence of each of two versions. The limiting case would be alternate responses equal in frequency which should lead to conflict and long latencies (Cf. Sears and Hovland, 1941). When one fom is strongly dominant, there may be little choice and consequent rapidity of response, as if it were not a homograph. Relative frequency was calculated as the differences in the proportions of each of the two responses to the stimulus word. If read was given as /rEd/ 80% of the time and /rid/ 20%, the relative frequency score was .60. These scores were correlated with the mean latency for each word. Across the 24 words, the correlation is .28, which is not significant. In one of our pre-tests, this correlation was around .90, but that list contained only monosyllables of the type bow, read, and tear and no homographs such as permit or refuse. When we select the 11 words in the present list which are the same form as the pre-test list, the correlation between relative frequency and latency is .59, $p < .05$. It appears, then, that the relative availability of the two forms of the homograph influences its latency, but only for monosyllabic words. This characteristic of homographs is discussed below.

The 24 homographs contain three distinct types, two of which merit closer concern. Eleven monosyllabic words change the phonemic value of their vowels in the two possible pronunciations: /bo/ vs /baU/, or /tIr/ vs /tEr/. Another group involve a stress shift as well as a vowel change: /prətést/ vs /prótest/, or /Ínvalld/ vs /Invǽlld/. One word, permit, shifts stress but does not change vowels, and will not be considered in this analysis.

We would expect that the polysyllabic homographs will require longer latencies than the simple words because of the double operation. However, since they are longer words, the longer latencies may be attributable to length. For example, the longer control words paired with the complex homographs required longer times to read than the shorter control words ($t = 3.219$, $p < .01$).

To control for length, the latency of each control word was subtracted from the latency of the homograph. The difference scores for the 12 polysyllabic words is still longer than the differences of the monosyllables ($t = 1.920$, $p < .05$). That is, it takes longer to read homographs involving a stress shift plus a vowel change than to read those requiring only a vowel change, holding constant the length of the words.

The first purpose of this study was to demonstrate that homographs took longer to read than matched written words that decode to only one spoken word. Our results clearly indicate that this is so. In addition, the latency in reading monosyllabic homographs is longer as the two response tendencies approach equality. Further, homographs in which the two forms differ by a stress shift as well as a vowel change require longer verbal reaction times than words differing by only a vowel change.

REFERENCES

Gallwey, Mary O. A study of decision time in children. Unpublished doctoral dissertation, Cornell University, 1958.

Levin, H. The psychology of reading. In J. S. Bruner (ed.) Toward a theory of instruction. U.S.O.E. Monograph, in press.

Levin, H. and Biemiller, A. J. Studies of oral reading. III. Contingent versus non-contingent initial spelling-to-sound correspondences. Cornell University, in preparation.

Sears, R. R. and Hovland, C. I. Experiments on motor conflict. II. Determination of mode of resolution by comparative strengths of conflicting responses, J. exper. Psychol., 1941, 28, 280-286.

Studies of Oral Reading

V. Homographs in Grammatical Frames

Harry Levin, Boyce L. Ford and Mary Beckwith[1]

Cornell University Indiana University

In an earlier study (Levin and Ford, 1965) we demonstrated that homographs took more time to be read than did other comparison words. Further, monosyllabic homographs depend on the relative strengths of the two response tendencies: those with more nearly equal strengths require longer times to decode. Homographs are distinguished as words that cannot be definitively decoded to one or another sound form without evidence external to the word itself. Hence, the present study provides an additional word that gives the part of speech membership of the homograph. We ask several questions. What are the relationships between time to decode homographs compared to control words when both are in similar grammatical frames? Does the part of speech make a difference? Are homographs with additional cues read more rapidly than homographs without cues (as in the first study)? Finally, how does the grammatical cue influence the reading of the non-homograph control words?

METHOD

Subjects. The 80 subjects participating in this study were drawn from a group of high school students participating in the Cornell University Summer Advanced Placement Program, which is designed to accomodate high aptitude students who are entering their senior year in

[1]This research was supported by a contract with the U.S. Office of Education.

high school. The students selected attend the regular six week summer session at Cornell University.

Selection of these students is based upon the University's normal admission procedures using criteria available from the applicants' high school records. The majority of records contained the following information: I.Q. scores, high school grade average, aptitude scores (PSAT, both verbal and math), and age. The information for the subjects of this study is summarized in Table 1. In addition to the admission data, grades received for the summer courses at Cornell work are also presented.

Table 1. Description of Subjects

	Male	Female	Overall
Age	16.82	16.84	16.83
I.Q.	136.88	136.19	136.50
PSAT			
Verbal	62.86	60.69	61.75
Math	61.18	59.17	62.61
High School Grade Ave.	93.13	92.59	92.83
Cornell Grade	83.13	77.59	80.08

Stimuli. The word lists are the crux of this experiment and since their construction was somewhat complicated we shall describe the process in detail. We started with 24 homographs, each of which was used twice, as a noun and verb, noun and adjective, or verb and adjective. This brings us to 48 homograph words. Each homograph was matched with a control (non-homograph) word which was matched with its paired homograph on the following characteristics: (1) number of letters, (2) number of syllables, (3) same initial consonant, (4) frequency as defined by the Thorndike and Lorge G Scale, and (5) part of speech. One other constraint was placed on the control words. If possible, the words selected could function as two parts of speech in the same manner as their matched homographs. For example, the homograph DOVE can be either a noun or verb depending on the frame in which it is placed. Its control word should also be both a noun and verb, and the word DOCK fills this bill. On the other hand BASS can function as a noun or adjective, but we could find no control word that satisfied the five primary criteria and also acted as two parts of speech. In this case, BASS was paired once with BURR (noun) and once with BUFF (adjective).

We now have two sets of 24 pairs; 48 pairs or 96 words in all. Each member of each pair was placed in the context of a single preceding frame word which signalled its part of speech. For nouns, the frame word THE was used and for adjectives, VERY. Three frame words were used for verbs: TO, HE, THEY. HE occured with three pairs (WOUND-WRANG, READ-RODE, DOES-DREW) while THEY was used with the pair, DOVE-DOCK. All other verbs had the frame TO.

Two lists were formed from the 48 pairs so that each list contained 24 homographs and their yoked controls. A list contained a homograph in one frame only; the same homograph in another frame appeared in the second list. Finally, the two lists were matched on the incidence of nouns, verbs, and adjectives. The two lists are given in Table 2 in the order in which they were presented to the Ss.

Experimenters. The study was conducted by two experimenters, each of whom tested a male and female subgroup on each of the two lists.

Test Procedure. Each stimulus word was typed on a 2 x 2 slide in

Table 2. Word Pairs, Thorndike & Lorge G Scale Frequencies, and Latencies*

List I						List II					
	Homograph			Non Homo			Homo			Non Homo	
	Freq.	Latency		Freq	Latency		Freq	Latency		Freq	Latency
THE DOVE	19	.587	THE DOCK	16	.614	THEY DOVE	19	.633	THEY DOCK	16	.684
VERY BASS	11	.860	VERY BUFF	5	.686	THE BASS	7	.827	THE BURR	4	.689
VERY INVALID	9	.683	VERY ILLEGAL	5	.676	THE INVALID	9	.655	THE IMPRINT	3	.701
TO READ	100	.522	TO ROAR	50	.549	HE READ	100	.687	HE RODE	50	.576
TO PROTEST	50	.598	TO PACKAGE	38	.593	THE PROTEST	50	.660	THE PACKAGE	38	.584
TO LEAD	100	.557	TO LOAD	50	.561	THE LEAD	100	.694	THE LOAD	50	.574
TO RECORD	100	.582	TO RANSOM	10	.620	THE RECORD	100	.599	THE RANSOM	10	.660
VERY MINUTE	100	.702	VERY MODERN	100	.567	THE MINUTE	100	.597	THE METHOD	100	.561
THE TEAR	100	.715	THE TEAM	50	.598	TO TEAR	100	.689	TO TEAM	50	.591
THE POLISH	7	.667	THE PATROL	8	.571	TO POLISH	28	.621	TO PATROL	8	.623
THE WOUND	50	.586	THE WRECK	41	.562	HE WOUND	50	.777	HE WRANG	9	.680
VERY LIVE	100	.616	VERY LEAN	50	.584	TO LIVE	100	.586	TO LEAN	50	.576
TO WIND	100	.706	TO WALK	100	.545	THE WIND	100	.662	THE WALK	100	.559
THE ARITHMETIC	9	.631	THE ANTAGONIST	8	.759	VERY ARITHMETIC	9	.905	VERY AGGRESSIVE	7	.604
THE OBJECT	100	.608	THE OFFICE	100	.588	TO OBJECT	100	.652	TO OPPOSE	50	.633
TO INTIMATE	24	.985	TO IMPRISON	12	.657	VERY INTIMATE	24	.682	VERY IMPERIAL	21	.723
TO PRESENT	100	.642	TO PICTURE	100	.619	THE PRESENT	100	.670	THE PICTURE	100	.607
THE DESERT	50	.807	THE DAMAGE	32	.607	TO DESERT	50	.771	TO DAMAGE	32	.631
TO PERFECT	50	.669	TO POSSESS	50	.635	VERY PERFECT	50	.664	VERY PRIVATE	50	.616
THE PERMIT	50	.742	THE PARDON	36	.635	TO PERMIT	50	.650	TO PARDON	36	.604
THE REFUSE	100	.782	THE RAMBLE	7	.568	TO REFUSE	100	.575	TO RAMBLE	7	.561
THE BOW	50	.739	THE BUD	30	.625	TO BOW	50	.651	TO BUD	30	.663
HE DOES	100	.604	HE DREW	50	.570	THE DOES	100	.978	THE DUST	50	.634
TO SOW	29	.698	TO SKI	6	.568	THE SOW	29	.832	THE SKI	6	.724

* latency for 2nd member of each pair only

upper case pica and presented with a Kodak Carousel Projector (Model 550). Ss were tested individually and tape recordings were made on a Tandberg Model 3B-F. The standard accessory microphone was used.

During the presentation of stimuli the subjects were seated ten feet from the screen and three feet directly behind the projector. The front of the projector was six feet from the screen which made it possible to project an image 2½ inches in height. Subjects held the microphone about eight inches from their mouths. The experimenter sat on a line even with the projector and about three feet to the left of the subject.

Once the subject was seated he was given the following instructions:

> "I am going to project some words on this screen. The words will be real. They will be words that you have probably used or could use in everyday conversation. As each word appears, please say it. The words will be presented in pairs with one word following the other. For example this could be a pair. (At this point THE was presented and then one second later MAN was projected). Your task will be to say each word as it appears keeping in mind that the words go together. I'll project two pairs of words this time. Please say each word as it appears. (At this point TO was projected and as soon as the subject had completed his response, SAY was projected. As soon as the response to SAY was completed VERY was projected and then BIG.) The pace throughout the list will be the same as it was with the two pair I just projected. Remember, all you have to do is say each word as it appears, keeping in mind that the words go together in pairs.

When certain that the S understood the instructions, E proceeded through the list without stopping (except for a projector tray change half way through the list).

The tapes were later played through a rectifier into a Brush Recorder. Auditory input from the tape recording activated a recording pen producing a visual representation on graph paper. When no sound was present the pen was inactive and traced a straight line on the recording paper. By measuring the length of the straight line between onset of visual stimulus[2] and onset of the subjects vocal response, it was possible to determine the latency to each stimulus word. All false starts, vocal segregates, etc. were included as a part of the response latency. Latencies were recorded in millimeters and later transformed according to the following formula: $x = \log (y - 10)$, where x = transformed score, and y = latency in millimeters. Following analysis, latencies were converted into seconds by the formula; $x = .04 (y)$, where x = latency in seconds, and y = latency in millimeters.

RESULTS

The data of this study were analysed in several ways. The first analysis of variance included these classifications: Sex of S, E, List, and Homograph vs Non-homograph. As in our first homograph study, we found that girls responded more rapidly than boys. The overall mean for boys is .674 seconds and girls, .644 seconds ($F = 7.899$; 1/72 df; $p < .01$). Likewise, there was a substantial effect attributable to the E ($F = 9.38$; 1/72 df, $p < .01$).[3] Since Sex of Ss and the two Es are represented in our various groups without bias and since they are both variables in which we are not particularly interested nor which we can explain, we collapsed the subsequent analyses to the following independent variables: List, Homograph, and Part of Speech.

There were, it will be recalled, two separate lists arranged so that any given S responded to only one. Words within part of speech were distributed randomly between the two lists, so that if <u>to bow</u>

---

[2] Onset of stimulus was indicated by a characteristic sound (and later visual pattern) made by the projector when the slide dropped into place.

[3] Our impression is that the Es differed in the rate at which they presented the slides, that is, in the intertrial intervals and it is not surprising that this overall tempo can "push" the reaction times.

appeared on one, the bow would appear on the second. Were these lists adequately matched? Neither the main effect attributable to list nor any interactions between lists and the other variables reach statistical significance (F = 1.779; 1/78 df).

We may turn to the main intents of this study. First, does the speed of reading homographs differ from non-homographs, when both are presented in grammatical frames? The relevant means are presented in Table 3. The overall latencies to homographs is .697 seconds, to control

Table 3. Mean Latencies (Seconds) in Responding to Two Types of Words and Three Parts of Speech.

	Noun	Verb	Adj.	Overall
Homo	.700	.661	.730	.697
Non	.617	.604	.643	.621
Homo + Non	.659	.634	.686	.659

words, .621 seconds. The difference is highly significant (F = 147.811; 1/78 df, p <.001). Further, in the same table, it can be seen that there is a substantial effect attributable to the part of speech signalled by the frame word. Verbs are responded to more quickly than nouns which are faster than adjectives (F = 25.347, 2/156 df, p <.001). All three comparisons reach statistical significance. Some conjectures about the reasons for this finding will be discussed below.

The critical test of the effects of contextual frames on decoding is to compare homographs, as well as the control words, with and without frames. This involves the comparison of mean latencies in the earlier study (without frames) with the present study. First, we can see that the two groups of Ss are similar by comparing the data in Table 1 of that study with Table 1 above. Both groups were bright high school juniors who were in an advanced placement program. Their intelligence and achievement test scores are very similar.

The relevant means from the two studies are given in Table 4.

Table 4. Comparison of Single and Framed Words

	Single	In Frames	t	p
Homographs	.724	.697	1.271	>.10
Non Homographs	.662	.621	2.632	<.01
Overall	.693	.659	3.191	<.001

In general words are read more quickly when they are in frames than as single words. The breakdown of this overall difference is interesting, however. Words which have only one decoding possibility are read substantially more quickly when they are preceded by a context word. On the other hand, and most important for our purposes, is the finding that homographs are decoded only slightly more rapidly when additional cues are furnished, but the difference does not attain a statistically acceptable level.

DISCUSSION

Several findings in this study merit further discussion. Words preceded by a verb marker, usually to and in several cases he or they, are read more quickly than words preceded by the or very. In turn the words are read faster than very words. Two reasons occur to us. Verb forms are more frequently the base forms for derived nouns and adjectives then the other way around. Further, we would guess that adjectives are more frequently derived from nouns than the converse. We are saying, in effect, that readers are responding to the frequency of base rather than derived frequency, although we admit to no direct test of this notion.

Our second explanation depends on what Feigl has called "a promissory note on the past". In language learning, the infinitival form, to _____, is often used to introduce words and indeed appear in dictionaries in this form. It is probably the most practiced grammatical frame. Further, nouns are marked by <u>the</u>, whereas adjectives are unmarked. We are suggesting, therefore, that such language experiences have made the frames differentially available to our Ss, who were scholastically advanced and not unlikely have had language training.

The results are quite interesting when we compare the effects of grammatical frames on the speed of decoding both homographs and control words. The additional cues substantially speed up the reading of regular words but have minimal effects on homographs. This finding appears to mean that the grammatical cue is insufficient to reduce the indeterminancy inherent in decoding homographs, In other words, the single frame word in helpful in reading determinate forms but not useful enough when the form is strongly indeterminate, as are homographs. A single word marking the form class of a word is, after all, a small cue compared to what may exist in the context of a word to be decoded within a discourse. A subsequent study will treat variations of semantic environments on decoding target words.

<u>A Note on the Relations between Frequency and Latency</u>

There have been regular reports in the literature that the lexical frequencies of words and the times it takes to read them are substantially and negatively correlated (See, for example, Fraisse, 1964). The reasons for this finding are not immediately apparent to us. The frequencies are generally determined by word count as in the Thorndike-Lorge List (1944). Since these would generally represent frequency of words read, it is difficult to see why commerce with them should determine the reaction time to reading the word aloud. By this reasoning, on the other hand, one can understand the lowering of visual thresholds for recognizing frequently read words.

The frequency of the lexical word tells us little about the nature of the spelling-sound correspondences in the word. If the process were a left-to-right sequence of decoding the latency should be determined by the spelling-sound correspondences which we have found to be the case in an earlier study (Biemiller and Levin, 1965). On the other hand the correlations do exist and have been replicated. They imply that reading the word aloud involves some overall scanning and recognition prior to decoding. If the word is recognized on the basis of the first scan, it is said; if not recognized, it is decoded.

In the present study, thirteen control words appeared in each of the two lists, half of the time as verbs, half as nouns. The latencies of these words were correlated with their Thorndike-Lorge frequencies. For nouns this correlation is .211, for verbs, .258 (both calculations are Kendall's Tau). These relationships are obviously not statistically significant and are contrary to the usual finding when these two variables are related.

It must be remembered that the usual frequency-latency studies involve words presented in isolation. In this case, they followed a grammatical cue word. We are tempted to speculate that the usual findings occur when there is some indeterminacy about the word. The frame, it will be recalled, speeds up the reading of control words, and the new speed is obvously independent of the word's lexical frequency.

We can anticipate one caveat to these findings. Perhaps, it might be argued, the context word occuring as it does before the target word creates a warm up or response set which pushes the reaction time. Contrary to this argument, we can find no speed up at the end of the lists nor does the context word appreciably affect the homographs.

SUMMARY

The verbal reaction times to reading both control words and homographs whose parts of speech membership were signalled by preceding cue words were studied. The following findings emerged:

1. Control words are read more quickly than homographs.
2. Verbs are read more rapidly than nouns which are more rapid than adjectives.
3. Control words are read more rapidly in a grammatical frame compared to the single word condition; homographs are not.

REFERENCES

Biemiller, A. J. & Levin, H. Studies in oral reading. II. Pronounceability. Unpublished mimeo.

Fraisse, P. Relations entre le seuil de reconnaissance perceptive et le temps de reaction verbale. Rev. psychol. Franc. 1964, IX, 77-85.

Levin, H. & Ford, B. L. Studies in oral reading. IV. Homographs vs. non-homographs. Unpublished mimeo.

Thorndike, E. L. & Lorge, I. The teachers word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Preliminary Draft
August, 1966

Studies of Oral Reading[1]

VI. Words with Digraph Spelling Patterns

Andrew Biemiller and Harry Levin

Cornell University

In English spelling, the usual case is that a single letter relates to a phoneme, although, depending on a variety of conditions, a letter may signal various sounds, or may not signal any sound at all. Less often, a group of letters may correspond to a single sound (phoneme). A subset of this condition are digraph spellings of single phonemes. Some English digraphs and their phonemic equivalents are SH /s/, CH /c/, TH /e/, NG / /, AI /e/, OA /o/. For our purposes digraphs are compared with common clusters, e.g., CL /kl/ or SP /sp/. The point is that the sound corresponding to the digraph is <u>not</u> predictable from either of the two letters alone, whereas the sounds related to the cluster may be "blended" from the individual letter-sound correspondences. It follows, then, that if a word containing a digraph spelling were presented in two successive parts such that th digraphs were broken (C HIP), the task of reading the word would be more difficult than the conditions where either (a) the digraph spelling was preserved on a unit (CH IP), or (b) where the cluster was broken (C LIP), or (c) where the cluster was preserved (CL IP). In fact, this study makes these comparisons.

The results of this experiment are also applicable to a ubiquitous discussion about the nature of the process of decoding print. In an earlier study (Levin & Biemiller, 1965) we suggested that a child shown a word to

---

[1] This research was supported by funds from the U. S. Office of Education.

read aloud, decodes the word, matches it to an auditory memory and says the word in time proportional to the match between the initial decoding and the stored image. Other studies have also invoked auditory-image-mediation (Smith, 1965; Conrad, Freeman & Hull, 1965). There is, however, contrary evidence indicating that the mediation may be visual (Kaplan, Yonas and Shurcliff, 1966; Gibson & Yonas, 1966).

In light of the earlier research, the rapidity with which a word presented in sequential pieces is read aloud permits us to make inferences about the way in which the input has been processed. If a digraph has been broken and the subject is going through a verbal loop, it should be difficult to integrate the whole word. For example S + HIP /s + hIp/ seems intuitively difficult to change to /sIp/. One might conjecture that visual imagery would be necessary to solve this problem. On the other hand if the pieces of the word are integrated visually, that is, each piece put in visual storage sequentially and the total visual image responded to (read with an inner eye), fractionated digraph spellings should not be more difficult than other split clusters.

METHOD

The stimuli were four letter words with various characteristics described below. Each word was presented in two parts via a Kodak Carousel slide projector. First, a blank slide was presented to the S. Following a verbal "ready", two slides were presented at maximum speed for this projector. This gave presentations of approximately 0.4 seconds for each slide with a 1.0 second interval between the two. The two slides were followed by another blank slide.

Latencies were timed with a Hunter electronic clock. The clock was stopped by the verbal response through a voice key in the circuit. The latencies recorded were the times for the total sequence from the presentation of the first slide to the response. The time for the presentation of the stimulus is constant for all stimuli, so the variable increment is the time from the final blank slide to the response.

Stimulus Materials. The total list contained 24 words, made up of three sub-sets containing eight words each and these eight comprised four matched pairs of words. The first sub-set's words represented variations in the intial spelling so that one word of a pair had a digraph spelling and its matched word an initial consonant cluster (chap/clap). We shall call this sub-set the "Initial List." The second, or Medial List, contained four pairs of words with medial digraphs and medial vowel clusters (fail/fall). The third, Final list had four pairs of words with variations in the last two consonants (fish/fist).

Each pair of words had one with a digraph spelling and its mate with a cluster. Since each word was presented under two break conditions --preserving as violating the digraph or cluster--there were 48 stimulus items. The two words in each pair were graphically identical except for the second letter in the two letter spelling pattern which is the focus of our concern. With these restrictions we made some attempt to match pairs for frequency on the Thorndike and Lorge (1944) Juvenile List. All words and their frequencies appear in Appendix A.

The list was presented in four different orders to control for serial effects.

Subjects. A total of 48 children from the second, third, and fourth grades of the Dryden, New York, Central School were tested. Eight boys and eight girls were drawn from each grade level, an equal number from each of two classrooms at that level. In each class, Ss came from the top two reading groups.

Procedure. Subjects were tested individually. Each subject was seated in front of a ground glass screen; the projector was approximately two feet away on the opposite side of the screen. The subject was told, "I'm going to show you some words on this screen." (At this point, the projector was turned on with a blank slide in place to direct S's attention to the screen.) "The words are broken into two parts. You tell me what the word is. Now I'm going to show you some practice words so you can get the idea." E showed a practice word. If the child could not read it, the E would suggest looking at another word. If after two words, the child was still unable to read words presented in this manner, the first two words were repeated again, slowly. If necessary, the child was told what the first word was. With four exceptions, all second grade, who were dropped from the sample, all children were able to read the four practice words by the end of this training.

Throughout the testing, if the child made no response in ten seconds, the experimenter would suggest going on to the next word. This suggestion was always taken.

RESULTS

The effect of breaking digraphs compared to clusters would be reflected in an interaction between the type of word and the nature of the break. The reason for this is that two major factors affected latencies in this study. The first is whether a digraph pattern was broken. The second is simply where in the word, regardless of spelling pattern, the break occurred. We assume that latencies for the two break conditions on the "cluster" control words estimate the impact of the location of the break. (This assumption may be partly wrong. See discussion.) Examination of mean latencies for non-digraph words in Table 1 indicates that at least in the Initial pattern condition there is a "break" effect. The average latency is 0.69 seconds less under the B_1 condition (in which the break occurs between the first and second letters) than under the B_2 condition, (in which the break occurs between the second and third letters.) There is a similar though smaller break effect in the Final condition.

Table 1

Mean Latencies for Digraph and Cluster Words Under Different Break Conditions

	Initial		Medial		Final	
	B_1	B_2	B_1	B_2	B_1	B_2
Digraph	4.87	4.89	4.77	4.74	5.19	4.85
Cluster	4.73	5.42	4.65	4.66	5.16	5.33

B_1= Breaks digraph and cluster spelling patterns.

B_2= Does not break spelling patterns.

Further examination of Table 1 brings out the effect of breaking digraph spelling patterns. Although we have just observed that our subjects showed considerably shorter latencies on the Initial pattern group non-digraph words (e.g., sped, clap) under the break condition which split the first and second letters, no difference associated with position of break in digraph words was found. If one accepts the assumption that latencies for the non-digraph words estimate the impact of breaks, we would now argue that the latency for the Initial digraph words under the first break condition ought to have been about 0.66 seconds shorter than it was. This may be taken as an estimate of the effect of breaking digraphs.

The procedure described above may be summarized as follows:

1. We subtract the mean latency for digraphs under break condition 2 (B_2 - not breaking digraph) from the mean latency for digraphs under break condition 1. This removes the general latency contribution for these words leaving as a remainder the latency reflecting a break effect and the latency reflecting an interaction.

2. We perform the same operation on the non-digraph latencies.

3. We now subtract the remainder for non-digraphs from the remainder for digraphs. This removes any common break effect, and leaves only the latency attributable to the interaction. If there is no interaction, this value will, of course, be zero.

Applying this procedure to the data presented in Table 1, we get the results presented in Table 2. This interaction is significant at the .05 level ($F = 3.481$ for 3 and 18 dif.) Full analysis in Appendix B). Note that there is clearly no effect for Medial digraphs.

Table 2

Mean Latency Contribution in Seconds

from Breaking Digraph Spelling Patterns

Initial	Medial	Final
.67	.04	.51

There was a substantial main effect attributable to grades. Fourth grade children had a mean overall latency of 3.87, third graders of 4.73, and second graders, 6.21. There was no significant interaction of grades x breaks x digraph/cluster. ($F = 0.71$ for 6 and 36 df.) However, the means still bear examination.

Table 3

Mean Latencies by Grades for Digraph and Cluster Words

Under Different Break Conditions

		Initial		Medial		Final	
		B_1	B_2	B_1	B_2	B_1	B_2
2	Digraphs	6.32	6.23	5.95	6.14	6.38	5.82
	Clusters	5.82	6.89	5.76	6.00	6.57	6.69
3	Digraphs	4.25	4.60	4.65	4.43	5.38	4.59
	Clusters	4.38	5.29	4.42	4.47	4.97	5.33
4	Digraphs	4.04	3.84	3.69	3.65	3.80	4.14
	Clusters	4.00	4.07	3.79	3.50	3.95	3.97

The data in Tbble 3 may be analyzed in the same way as the material in Table 1, in order to yield latencies attributable to breaking Digraph spelling patterns. These are shown in Table 4.

Table 4

Mean Latency Contribution in Seconds

From Breaking Digraph Spelling Patterns

	Initial	Medial	Final	Means Initial & final only)
Grade 2	1.15	.04	.67	.91
Grade 3	.56	.27	1.15	.85
Grade 4	.28	-.25	-.31	-.01

A rough interpretation of this table might be that the grade 4 means essentially represent variation around a zero effect, whereas the Grades 2 and 3 means (for Initial and Final groups) represent an effect of about 0.90 seconds. This point will be pursued further in the discussion.

All other main effects and interactions were non-significant. (See Appendix B for complete analysis.)

DISCUSSION

The following results merit additional discussion: (a) breaking digraph spelling patterns appear to influence the latencies of reading words for second and third grade children, but not for fourth graders; and (b) breaking medial digraph spellings has no effect on latencies.

To understand the first finding it may be useful to conjecture about developmental changes in reading. In its usual initial training, reading is closely bound to oral language. Children start to read by reading aloud. In several earlier studies, we discussed a reading strategy in which children, by applying spelling-to-sound correspondence rules, decoded the written word to sound "in their heads", matched the sound to a stored auditory schema, and finally said the word after a longer or shorter latency depending on the closeness of the match. We should point out that the printed word was available to the child throughout this process.

If reading continued to involve such channeling through sound one could never read quickly or efficiently. At some point in learning to read and with some materials, the reader processes the printed information visually, This mode permits rapid sampling of the text which if converted to sound would yield a disconnected, although possibly comprehensible, message.

If we assume that as the child has more experience with reading he relies less on auditory processing and more on visual, the suggestion that the digraphs are read as quickly as the clusters, regardless of the point of break, appears sensible. An attempt to decode the broken digraphs to sound would lead the reader into a morass from which he can extricate the word only by recalling it visually and as a whole. By comparison, if he stores the first piece as visual memory and then adds the second fragment to it, none of the displays should be more difficult than any other.

Our finding that cluster words show break effects could be troublesome. One might argue, for example, that breaking digraphs has no effect in the Initial condition. Instead, one could hold that there is simply a cluster x break interaction.

We think that both effects may occur. If one attempts to articulate the cluster words it does seem that they are more difficult under the break condition in which the pattern remains unbroken. However, inspection of results in the Final condition (see Table 1) in which there was only a small cluster break effect, indicates that there is a substantial digraph break effect in the predicted direction.

It will be noted that the cluster break effect is also lessened for grade 4 children, further supporting our explanation of the differences between grade 4 children and the others.

There were no effects on latency attributable to medial digraph spellings. Several reasons are germane. The initial and final digraphs and clusters were consonants; the medial digraphs were vowels. It may be that the spelling-to-sound correspondences formed to vowels are different in kind than to consonants, as are the perceptions of these two kinds of sounds. Any vowel spelling is associated with so many sounds, that readers may depend for their decoding upon additional sources of information such as their consonantal environments or the probabilities of sounds associated with the spelling. Also, it should be pointed out that there were no double vowel clusters for comparison with the digraphs.

Another reason for the absence of effects in the medial position is the demonstrated impotence of this part of the word to aid in its recognition (Marchbanks & Levin, 1965). Rather, children recognize words on the basis of their initial or terminal characteristics and we have again demonstrated that tampering with these parts of the words slows their reading.

SUMMARY

Words containing digraph spellings (sh, ai, ng) were presented in two parts, either preserving the intactness of the digraph (sh ed) or breaking the two letters (s hed). Comparison words had common clusters in their spelling (sled). The Ss were 48 children drawn from the second, third, and fourth grades. The dependent variable was latency measured from the presentation of the first word fragment to the onset of the oral response.

Results indicated that second and third graders took nearly one second longer to read words whose initial and final digraphs were broken than they did to read words whose digraphs were presented intact. The effect did not occur for fourth grade children. There was no latency effects attributable to breaking medial, vowel digraphs.

The findings are explained in terms of auditory or visual processing of the stimuli.

REFERENCES

Biemiller, A. J., & Levin, H. Studies of oral reading, II. Pronounceability. Mimeo paper, 1965.

Conrad, R., Freeman, P. R., & Hull, A. J. Acoustic factors versus language factors in short-term memory. Psychon. Sci., 1965, 3, 57-58.

Gibson, E. J., & Yonas, A. A developmental study of the effects of visual and auditory interference on a visual scanning task. Mimeo paper, 1965.

Kaplan, G. A., Yonas, A., & Shurcliff, A. Visual and acoustic confusability in a visual search task. Mimeo paper, 1966.

Levin, H., & Biemiller, A. J. Studies of oral reading, I. Words versus pseudo words. Mimeo paper, 1965.

Marchbanks, G., & Levin, H. Cues by which children recognize words. J. of Educ. Psychol., 1965, 56, 57-61.

Smith, W. M. Visual recognition: facilitation of seeing by hearing. Psychon. Sci., 1965, 2, 157-158.

Thorndike, E. L., & Lorge, I. The teacher's word book of 30,00 words. Bureau of Publications, Teachers College, Columbia University, New York, N. Y., 1944.

APPENDIX A

		Thorndike-Lorge "J" Freq.	Mean Latency in Seconds					
			2nd		3rd		4th	
			B_1	B_2	B_1	B_2	B_1	B_2
Initial	chap	45	6.36	7.86	4.81	4.61	4.73	4.04
	clap	100	5.72	6.83	4.22	5.53	3.60	3.42
	chew	49	6.36	5.29	4.14	5.04	3.64	3.76
	crew	210	5.69	8.58	4.17	6.22	3.93	4.49
	shop	483	6.52	4.93	3.95	4.81	3.76	3.69
	slop	4	4.96	6.09	4.36	4.24	3.89	3.71
	shed	203	6.04	6.86	4.10	3.94	4.05	3.88
	sped	62	6.90	7.05	4.76	5.18	4.57	4.68
Medial	fail	410	6.65	6.65	4.14	4.24	3.42	3.43
	fall	M	4.20	4.97	3.92	4.08	3.49	3.50
	hail	250	6.97	5.51	5.11	5.05	4.12	4.03
	hall	700	4.48	5.53	3.92	4.56	3.46	3.51
	boat	600	5.17	5.37	4.30	4.14	3.70	3.35
	bolt	123	7.78	7.07	4.79	4.74	3.57	3.40
	soak	60	6.01	7.05	5.07	4.31	3.54	3.80
	sock	14	6.57	6.42	5.05	4.49	4.66	3.61
Final	sang	219	6.26	6.89	5.24	4.13	3.70	3.90
	sand	415	5.54	5.81	4.05	5.10	3.45	3.49
	hung	320	7.05	5.69	5.44	5.46	3.53	4.09
	hunt	430	5.91	5.91	4.13	5.05	3.49	3.18
	fish	700	5.26	4.80	5.04	3.81	3.43	3.90
	fist	57	7.14	7.14	6.08	5.73	4.06	4.05
	hash	1	6.96	5.92	5.82	4.36	4.56	4.66
	hasp	0	7.71	7.89	5.63	5.45	4.80	5.15

APPENDIX B

Analysis of Variance

Source	d.f.	SS	MS	F
Initial-Medial-Final (IMF)	2	4.5025	2.2514	1.329
Contingency-Noncontingency (C)	1	.4118	.4118	1
IMF x C	2	.7655	.3827	1
error 1	18	30.4876	1.6937	
Breaks within IMF (B/IMF)	3	1.6005	.5335	1.9239
C x B/A	3	2.8602	.9534	3.4381
error 2	18	4.9913	.2773	
Grades (G)	2	134.8375	67.4187	198.2902
G x IMF	4	1.3268	.3317	1
G x C	2	.1411	.0705	1
G x IMF x C	4	.4802	.1200	1
error 3	36	12.2406	.3400	
G x B/A	6	1.5129	.2521	1
G x C x B/A	6	1.2004	.2001	1
error 4	36	10.1010	.2805	

Error 1 = Words within IMF and C

Error 2 = Breaks x Words in IMF and C

Error 3 = Grades x Words in IMF and C

Error 4 = Grades x Breaks x Words in IMF and C

Preliminary Draft
January, 1967

Studies in Oral Reading:

VII. Homographs in a Semantic Context[1/]

Boyce L. Ford & Harry Levin

Cornell University

A homograph is a written word with at least two possible pronunciations each with its particular meaning depending on the context in which the homograph appears. For example, the homograph sow can be read as either /sow/ or /saU/ depending upon whether the context requires a verb meaning the placing of seeds in the ground, or a noun meaning a female pig. When words like sow are presented without a context and subjects are asked to read them aloud, the dual meanings and sounds are reflected in longer reaction times. It takes people longer to read the homograph than a comparable non-homograph word (Levin and Ford,1965.) When the homograph is placed in a grammatical context, e.g., the sow or to sow, subjects are able to respond to the word faster (Levin, Ford, and Beckwith, 1965), indicating that grammar is a potent cue in the resolution of the ambiguity.

Several researchers have investigated the effects of preceding semantic contexts on the reaction times to words. Ford (1952) anticipated a reduction in the latency of free associations to a word if a synonym to that word had previously been presented. However, his results indicated

1/ This research was supported by funds from the U.S. Office of Education.

that the synonym context was associated with a delay in association, as compared to a no context condition. In a later study, Cofer and Shepp (1957) found a reduction in the perceptual recognition time when the test word had been preceded by a synonym. The contradictory results in the above studies could be a funct. n of different response measures.

Our purpose in the present study is to follow-up the preceding study (Levin, Ford, and Beckwith, 1965) in which we demonstrated the facilitating effect of a grammatical context. This study investigates the effects of semantic contexts on the reaction time to the reading of both homographs and non-homographs.

Method

Subjects. 44 female Cornell undergraduates volunteered for the experiment. All Ss were tested by one male E.

Stimuli. The basic stimuli for this study consisted of 17 homographs with a matched non-homograph control word for each form of a homograph. The control words were matched with the homographs on the following characteristics: (1) number of letters, (2) number of syllables, (3) initial consonant, and (4) frequency as defined by the Thorndike and Lorge G scale.

Following selection of control words, each member of each homograph-non-homograph pair was placed in the context of a single preceding frame word. The frame words were either synonyms of the test words or a word within the same general response class. Examples of synonyms used as frame words are, TINY-MINUTE, COIL-WIND, and OWN-POSSESS; examples of context within the same response class as the test words are, ROBIN-DOVE, CRY-TEAR, and RUN-WALK.

Two word lists were compiled from the 34 homograph non- homograph word pairs. A list contained a homograph in one form only; the other form appeared in the second list. When completed a list consisted of 17 homographs and 17 non-homograph controls all placed in separate semantic frames. This made the total number of words in each list 68.

Two other word lists were constructed for control purposes. In a previous study (Levin, For, and Beckwith, 1965), we found a faster latency for words presented in grammatical frames. It is possible that the first member of a pair, i.e. the grammatical frame, served as a ready signal for the test word and was instrumental in lowering the latency. To control for this possibility, the test words in one of the semantic lists were placed in a meaningless frame, consisting of one of the following consonants: N, L, W, X, F, S, M, or H.

The second control list contained the same test words as were used in the consonant frame list, but no context was provided in this list. The test words were presented in isolation.

In summary, four word lists were constructed each consisting of 17 homographs and 17 matched non-homograph control words. Two of the lists contained the test words placed in a meaningful context (MFC1 and MFC2). A third list contained the same test and control words as were used in MFC2 but with a context consisting of a "meaningless" consonant (MLC). The last list contained the same homographs and non-homographs as were used in lists MFC_2 and MLC but with no context (NC).

Test procedure. Each stimulus word was typed on a separate 2 x 2 slide in upper case pica and presented with a Kodak Carousel Projector. Ss were assigned to one of the four word lists in the order in which they appeared for the experiment. All testing was done individually and tape recordings were made of the complete session.

The test words were presented in random order with each homograph and non-homograph immediately preceded by the appropriate frame word. Order was the same for all lists. A constant interstimulus interval of three seconds was used in word list NC while two different interstimulus intervals1/ were used in the other three lists. A two second interval was used between pairs. Prior testing had disclosed that when stimuli were presented at a constant interval, Ss found it difficult to keep track of the pairs. The two and three second intervals enabled Ss to read the words in pairs and still left them enough time to respond without being pressed.

During the presentation of stimuli the Ss were seated ten feet from the screen and three feet directly behind the projector. The front of the projector was six feet from the screen which made it possible to project stimulus words 2 1/2 inches in height. The experimenter sat on a line even with the projector and about three feet to the right of the S.

Once S was seated, the following instructions were given:

> I am goint to project some words on this screen. The words will be real. They will be words that you have probably used or could use in everyday conversation. As each word appears, please say it. The words will be presented in pairs with one word following

1/ Two interval timers in series were used to activate the projector.

the other. For example this could be a pair. (At this point MALE was presented and then two seconds later MAN was projected.) Your task will be to say each word as it appears keeping in mind that the words go together. I'll project two pair of words this time at the pace that will be used throughout the remainder of the list. Please say each word as it appears. (At this point TELL was projected and two seconds later SAY was projected. After SAY had been displayed for three seconds during which S had responded, HUGE the frame word for the next pair was presented. After two seconds BIG was presented.) The pace throughout the remainder of the list will be the same as it was with the two pair I just projected. Remember, all you have to do is say each word as it appears, keeping in mind that the words go together in pairs.

For Ss assigned to the NC list the following instructions were given: "I am going to project some words on this screen. The words will be real. They will be words that you have probably used or could use in everyday conversation. As each word appears, Please say it." S was then shown three demonstration words MAN, SAY and BIG. S was also informed that the pace was the same as would be used throughout the remainder of the list.

When certain that S understood the instructions, E proceeded through the list without stopping.

The tapes were later played through a rectifier into a Brush Recorder. Auditory input from the tape recording activated a recording pen producing a visual representation on graph paper. When no sound was present the pen was inactive and traced a straight line on the recording paper. By measuring the straight line between onset of visual stimulus2/ and onset of S's vocal response, it was possible to determine the latency to each stimulus word. All false starts, vocal segregates, etc. were included

2/ Onset of stimulus was indicated by a characteristic sound (and later visual pattern) made by the projector when the slide dropped into place.

as part of the response latency. Latencies were recorded in millimeters and later transformed according to the following formula: $x = \log(y-8)$, where x = transformed score, and y = latency in millimeters.

RESULTS

It will be remembered that two word lists with a meaningful context were included in this investigation. Both lists were the same except for the form of the homograph included and some control word which were different because of matching difficulties. The adequacy of our homograph-non-homograph matching was evaluated by an analysis of variance, with the classifications, MFC1 vs MFC2 and Homograph vs Non-homograph (Cf. Table 1 for relevant means). The only significant effect is that attributable to Homographs vs Non-homograph ($F = 14.437$; 1/32 df, $p < .001$).

This significant homograph effect replicates our earlier findings and also suggests that our matching procedures were reliable. The absence of a difference between MFC1 and MFC2 suggests that our selection of a meaningful context for each form of a homograph and its control word, was equivalent for both meaningful lists. Because of this equivalence and since the same homograph forms were used in the MFC2, MLC and NC word lists, the NFC1 list was dropped from further analyses.

The effect of context was initially evaluated using two analysis of variance classifications, word type (homograph and non-homograph) and context. There were three levels of context, meaningful (MFC), meaningless (MLC), and none (NC). Again as in previous studies, non-homograph are responded to faster than homographs ($F = 8.73$; 1/32 df, $p < .01$).

The effect attributable to context was also significant ($F = 9.92$; 2/64 df, $p < .001$) with MFC latencies lower than those in the MLC condition which in turn were lower than the NC latencies. The interaction between the two main effects was not significant ($F = 1.08$; 2/64 df). (Cf. Table 1 for relevant means.)

Following the above analysis, it was discovered that the mean log. latencies for the present NC condition were almost the same as the means for the same condition used in a previous study (Levin and Ford, 1965). In fact, the homograph means were the same while the non-homograph means were 91.88 in the present study and 91.30 in the earlier study (Cf. Table 1). The earlier isolation study had been conducted concurrently with our previous grammatical context study (Levin, Ford, and Beckwith, 1965) and had in fact been used as a control for the effects of grammar. Since the two earlier studies had been considered comparable and since the earlier isolation study and the present NC condition appeared to be comparable, it seemed reasonable to compare the grammatical context with the three context conditions used in the present investigation. This post hoc comparison is further supported by the fact that the two earlier studies simultaneously drew Ss from the same subject pool and one subgroup from each study consisted of females tested by the same E of the present study.

Two other factors supply additional support to the comparability of the two types of context. First, although the previously mentioned subgroup of 11 females in the grammatical frame study received 24 test homographs and control words, the same 17 homograph forms that were used in the present investigation appeared among the previous 24 and in the same order as they appeared in the present study. Only two control words were different and these satisfied the same criteria of selection. Secondly, rank-order correlations of latencies between the earlier investigations and the present study were all significant. The correlation between the MFC homographs and the same words previously placed in a grammatical context was .49 ($p < .05$). For non-homograph control words the correlation was .68 ($p < .01$). The present NC condition and the previous isolation condition correlated .64 ($p < .01$) for homographs and .69 ($p < .01$) for the 17 non-homographs.

Table 1

Means (Log.)

	MFC1	MFC2	MLC	NC	G 1/	I 2/
Homo.	98.12	95.35	97.06	100.88	91.79	100.88
Non-h.	86.65	87.35	91.29	91.88	86.80	91.30

1/ Grammatical context (Levin, Ford & Beckwith, 1965)
2/ Isolation (Levin & Ford, 1965).

Weighing both the pros and cons of a post hoc comparison, we felt justified in comparing the earlier grammatical context condition with the present context conditions. The supplemented data were analyzed by analysis of variance using the same model as was previously used, the only change being 4 levels of context instead of the previous three. As expected the results were the same as were found in the previous analysis.

Non-homographs are again responded to faster than homographs ($F = 8.96$; 1/32 df, $p < .01$). Type of context was also significant ($F = 11.89$; 3/96 df, $p < .001$) while the interaction between the two main effects does not reach a significant level ($F = 1.06$; 3/96 df). The means, as would be expected from the non-significant interaction, are ordered in the same direction for both homographs and non-homographs (cf. Table 1). For both word types the shortest latencies are under the grammatical (G) condition with MFC next in order and followed by the MLC context which is in turn followed by the slowest condition, NC. The Tukey method was used to test the significance of the differences among the appropriate means (cf. Table 1). These differences and the results of the Tukey test are presented in Table 2.

Table 2

Log. Differences[1/] Among the Four Context Means

Homograph	G	MFC	MLC	NC	Non-homograph	G	MFC	MLC	NC
G	--	3.59	5.30*	9.12**	G	--	.59	4.53	5.12*
MFC		--	1.71	5.53*	MFC		--	3.94	4.23
MLC			--	3.82	MLC			--	.59
NC				--	NC				--

1/ A difference of 4.77 is necessary to attain the .05 level of significance while a difference of 5.84 is significant at .01.

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

It is immediately apparent that either of the meaningful contexts are sufficient to reduce the decoding latency to homographs. Both meaningful contexts are superior to no contexts while only the G context is effective when compared to homographs placed in a meaningless context (MLC). The MLC condition is not significantly different from the NC condition, but the difference is large enough to be suggestive. The same situation holds for the difference between the two meaningful context conditions. Although not significant, the difference is suggestive.

When non-homographs are considered, grammar is the only context that significantly reduces the latency. There is obviously no significant difference between the two meaningful conditions nor is there any difference between a meaningless context and the no context condition. Neither of the meaningful contexts are significantly different from the meaningless context but both approach statistical significance and are in fact significant when the less conservative Dunnett test is used. The MFC - NC difference is also significant using the Dunnett test.

The differential effects of the various context conditions on homographs and non-homographs was assessed by conventional t-tests. Although the interaction between context and word type was not significant, the graph presented in Figure 1 does suggest a differential effect between context and the two word types. For example, the slope of the homograph curve between the MFC and G conditions appears to be steeper than the slope for non-homographs. This in fact is the case ($t = 2.14$; 16 df, $p < .05$) which suggests that grammar more effectively reduces the latency to homographs than to non-homographs. The difference between the two curve slopes between the isolation and CP conditions was also tested. The difference does prove to be significant ($t = 2.46$; 16 df, $p < .05$) suggesting that homographs are more affected by the single consonant context than are the non-homograph control words.

DISCUSSION

It is apparent that both grammatical and semantic contexts facilitate the reading of words. The results from the meaningless context condition demonstrate that this facilitation does not result from a "priming" effect resulting from any preceding stimulus. The preceding stimulus must be meaningful, either grammatically or semantically.

The semantic context used in this experiment is comparable to the synonym context utilized by Cofer and Shepp (1957). As mentioned earlier, these investigators demonstrated facilitation in the recognition of the second member of a word pair if the first member was a synonym to the word to be recognized. On the basis of this result Cofer and Shepp inferred that "facilitation of recognition will occur in stimuli other than a practiced stimulus if there is some direct associative and/or meaningful connection. We feel that our semantic context proved to be facilitating for essentially the same reason. It will be remembered that either synonyms or words in the same general response class as the test word were used as context words.

Although "associative and/or meaningful connections" are apparent in the semantic condition, these relationships are less obvious in the grammatical condition. An associative mechanism does not seem applicable to grammatical facilitation. Cofer (1965) has commented on this problem as it is related to the free-recall of categorized adjective-noun word pairs and concludes that "we need more information."

We agree with Cofer but also feel that Braine's (1963 and 1966) work on contextual generalization is related to this problem. Braine finds that ten-year old subjects, after a relatively brief exposure period to an artificial language, are able to generalize the correct temporal position of the words used in the language. When Braine asked his subjects why they gave the correct response, a common reply was "that it sounded right." Although the subjects had never heard the particular utterance before, they were able to respond correctly if the temporal position of the words remained constant.

This is precisely the case in our grammatical frame condition. Subjects were presented with word sequences that maintained the temporal order normally found in English. The pairs, if you wish, "sounded right", the pairs conformed to common words sequences.

In summary, this study has shown that single words can be read faster if they are preceded by another semantically or grammatically related word. The same results were found with both homographs and non-homographs, with a greater effect demonstrated using homographs. The facilitating effect of semantics was discussed in terms of "associative and/or meaningful connections", and the effect of grammar was discussed in terms of familiarity with normal English word order.

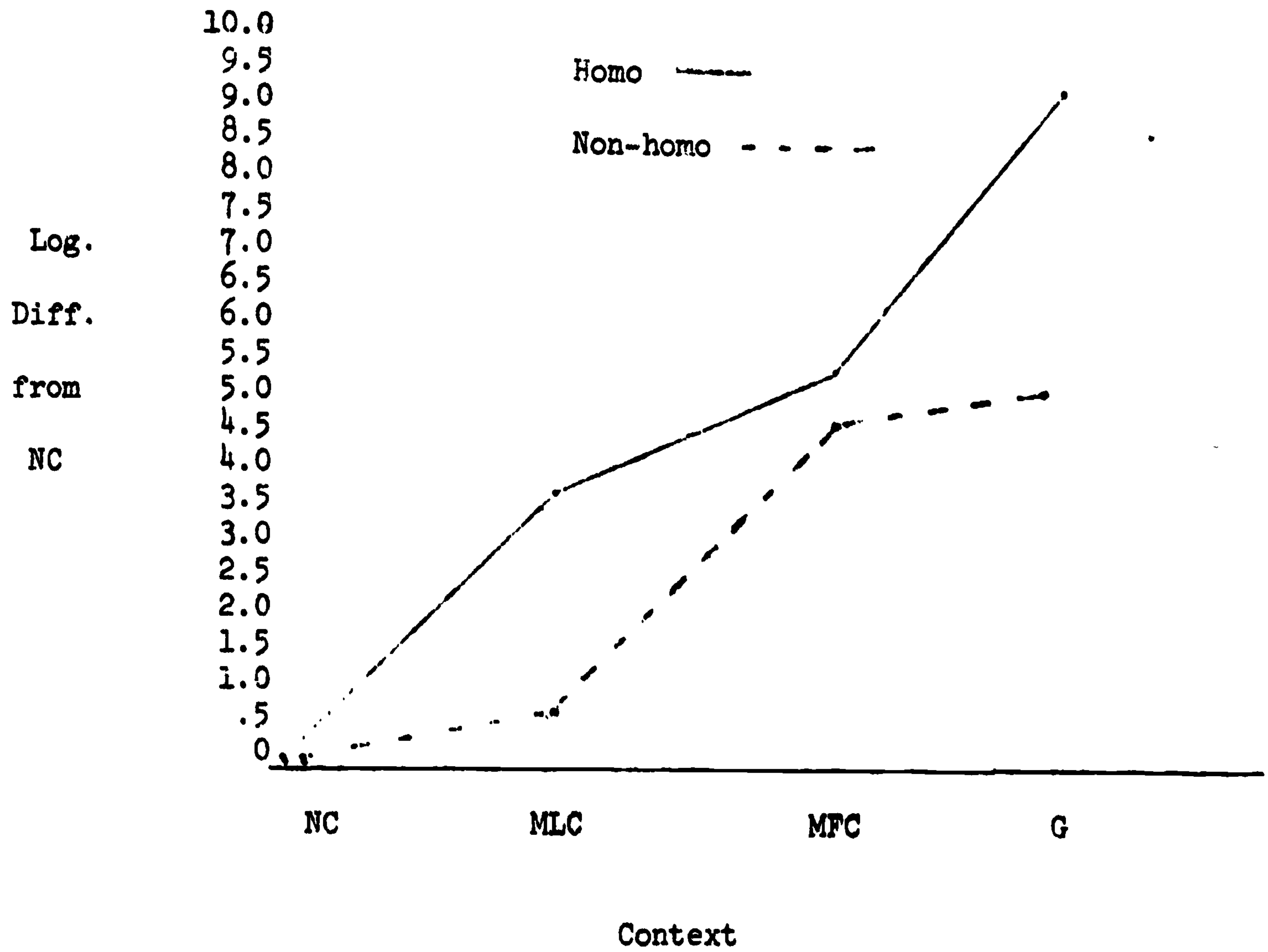


Figure 1.

Logrithm differences of each context condition from the NC condition.

REFERENCES

Braine, M. D. S. On learning the grammatical order of words. Psychol. Rev., 1963, 70, 323-348.

Braine, M. D. S. Learning the position of words relative to a marker element. J. exp. Psychol., 1966, 72, 532-540.

Cofer, C. N., & Shepp, B. E. Verbal context and perceptual recognition time. Percept. mot. Skills, 1957, 7, 215-218.

Cofer, C. N. On some factors in the organizational characteristics of free recall. Amer. Psychologist, 1965, 20, 261-272.

Ford, T. J. An investigation of mediated generalization among synonyms along a gradient of similarity of meaning. M.A. thesis, Univer. of Maryland, 1952

Levin, H., & Ford, B. L. Studies in oral reading, IV. Homographs vs. non-homographs. Unpublished mimeo.

Levin, H., Ford, B. L. & Beckwith, Mary. Studies in oral reading. V. Homographs in grammatical frames. Unpublished mimeo.

Thorndike, E. L., & Lorge, I. The teachers word book of 30,000 words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

PRELIMINARY DRAFT
September, 1966

Studies in Oral Reading:

IX. Sentence Structure and the Eye-voice Span[1]

Harry Levin and Elizabeth Ann Turner
Cornell University Harvard University

In reading aloud, the eye-voice span (EVS) is the distance, usually measured in words, that the eye is ahead of the voice. Interest in the EVS extends to the end of the last century. One consistent finding is that EVS tends to increase with age (Buswell, 1920; Tinker, 1958), and that the EVS is readily affected by the difficulty of the reading material (Buswell, 1920; Anderson, 1937; Fairbanks, 1957; Huey, 1922; Stone, 1941; and Tinker, 1958). The more difficult the reading material the shorter the EVS. Similarly, reading rate and EVS increase with more structured or constrained materials (Lawson, 1961,; Morton, 1964a, b). Thus the EVS would be shorter for a word list than for sentential material or, said another way, the greater the redundancy of the material the longer the EVS.

There is contradictory evidence as to whether the position within a line has any effect on the EVS. Buswell (1920; 1936) found no effect of position within a line. Quantz (1897) and Fairbanks (1937), however, both reported that the EVS was longest at the beginning of a line, medium length in the middle of a line, and shortest at the end.

1. This research is supported by funds from the U.S. Office of Education. We wish to thank Eleanor Kaplan and Peter Gamlin for their help with this study.

Fairbanks (1937) found, nevertheless, that the length of the EVS was more dependent on the difficulty of the reading material than on the position within a line. Both Buswell (1936) and Fairbanks (1937) found that position within the sentence affected the EVS; that the EVS was longest at the beginning of the sentence and shortest at the end. Buswell reported this effect for good readers only, while Fairbanks found it with both good and poor readers. As subjects were reading paragraphs in most instances and since little statistical analysis was presented, it is unclear as to how position in sentence and position in line were separated. Also, if we assume that sentences are more constrained, in general, at the ends than at the beginnings, several of the above findings are contradictory.

With good and poor readers selected by a standardized reading test, Buswell (1920) found that good readers had longer EVS's and read more rapidly. Morton (1964a), using reading rate as the criterion for good and poor readers, found that the EVS for good (fast) readers is longer than for poor (slow) readers. Quantz (1897) also reported that the higher the reading rate, the longer the EVS. Thus fast rate and long EVS seem to go hand in hand. This would seem to be due to the fact that most of the time spent in reading is in fixation or pausing, and the fewer the pauses, the more rapidly a subject reads and the more he sees in one fixation pause. In an investigation of the number of times his Ss'EVS's ended at a phrase boundary, Schlesinger (1965) did not find a difference in frequency of times good or poor readers stopped at the end of phrase units.

A different sentence content was used for each of the light-out positions. Thus there was a total of eight two-word phrase sentences, 12 three word active sentences 12 three-word phrase passive sentences. and 16 four-word phrase active sentences. In addition, there were eight structureless word lists, making a total of 56 presentations.

Sentences were constructed with enough phrase units in them so that there would always be at least ten words in the sentences beyond the light out position. Starting with the sixth grade Ss, each of the critical sentences was embedded in a paragraph of four sentences. The critical sentence occurred an equal number of times in the first, second, third, and fourth sentence position. For the second and fourth graders, the paragraphs contained two sentences and the critical sentence occurred in either the first or second position. The light tended to be turned out toward the beginning of the line so that there would be at least ten words remaining in the critical sentence on that line for any one given sentence. A random order of presentation of sentences was used the same order was used for each subject.

Two similar sets of sentences were used. One set was made up with the vocabulary of the second grade reader and was used with second and fourth grades; another set was made up with the vocabulary of a sixth grader and was used with the sixth grade and all older subjects.

<u>Recognition Lists</u>. Recognition lists were made up for half of the sentences in each of the four sentence types. Each recognition list consisted of five content words taken from the final part of the sentence, starting three words beyond the light-out position, and five confusion words, one for each of the five content words.

Apparatus. A wooden box (24 X 18 X 12) with a slanted front surface in which there was a one-way mirror was used to present the stimulus sentences. It was so designed that S could only see through the mirror to read the sentences on the cards when the light inside the box was turned on. A micro-switch was used to activate the light when E released the micro-switch, the light inside the box turned off. A timer was also connected to the apparatus so that the clock started when the light inside the box was turned on and stopped when the light inside the box was turned off. Thus the timer measured the time taken by a S to read from the beginning of a passage to the light-out position.

Procedure. The S was seated in front of the apparatus and was told to focus on a red dot on the one-way mirror which designated the point where the beginning of the paragraph would appear when the light was turned on. The S was told to read at his normal rate or the rate at which he would read a storybook out loud. When the light inside the box was turned on, S began to read the passage in front of him out loud; when the light went out, S was told to report all the words he had seen beyond the word he was saying when the light went out. All the words reported by the S were recorded. The time taken by the S to read to the point at which the light was turned out was also recorded in order to get a measure of the S's reading speed. Readers were divided into two groups, slow and fast, on the basis of whether their scores were below or above the median of the distribution of words/second read.

When there was a recognition list for the sentence, the S was shown each word of the list individually and was asked if he thought he had seen it. Both correct and incorrect recognitions were noted.

Scoring. The number of consecutive words which each S reported having seen beyond the light-out position was recorded for each sentence as a measure of his eye-voice span for that sentence. In addition, the number of times S reported having read ahead to the end of a phrase boundary versus to a non-boundary position was recorded.

RESULTS

Length of Eye-voice Span. A comparison was made between the mean length of the EVS on the unstructured word lists (mean span = 2.195 words) and the mean length of the EVS on all the structured sentences (mean span = 3.910 words). The difference in the EVS for these two types of reading materials is significant at $p < .001$ ($t = 6.17$, $df = 38$, two-tailed test). Table 2 shows the mean EVS for each of the six grade

Table 2

levels, for slow and fast readers, and for the three lengths of phrase units. From this table it is evident that the sixth grade subjects are a deviant group. They were the first age group tested and were an unselected sample of sixth grade readers in a rural New York school. However, it was evident very early in the testing, that a child who was an average sixth-grade reader in this school really read poorly, for our purposes. As these Ss had extreme difficulty in reading the material presented (all with a fifth grade vocabulary), they were excluded from the analysis. The remaining subjects were divided into two groups on the basis of the experimental material which they read: the second and fourth grades were grouped together and the eighth and tenth grades plus the college students were grouped together. Thus two three-way analyses of variance with the classifications, grade, reading speed, and chain length, were carried out on the length of the EVS. In the analysis of the second and fourth grades, the main effect of subject-grades was significant at the .01 level ($F = 26.86$, $df = 1/16$). Fourth graders had longer spans than second graders. There was also a significant difference between slow readers ($F = 5.00$, $df = 1/16$, $p < .05$). Number of words in a chain or phrase unit was not significant. However, the interaction between grades and chains was significant ($F = 3.42$, $df = 2/32$, $p < .05$). Figure 1a shows that the EVS for second graders is relatively constant for all chain

---

Figure 1a

---

lengths. Fourth grade subjects, on the other hand, show a tendency for the EVS to maximize on the three-word chains.

In the analysis of variance of the older subjects (eighth and tenth grades and adults), grade level or age was again significant ($F = 6.53$, $df = 2/24$, $p < .01$). From Table 2 it is evident that the adult subjects had longer eye-voice spans than the high school subjects. The difference in length between the mean EVS of the eight and tenth graders was not great but the eighth grade mean span was slightly longer than the tenth grade mean span. In contrast to the results with the younger subjects, the main effect of number of words in a chain or length of chain was significant in the analysis of the eye-voice span of older subjects ($F = 14.67$, $df = 2/48$, $p < .01$). Figure 1b shows that the EVS for all

Figure 1b

three older grades tends to maximize on the three-word chains, the same as it did with fourth grade subjects. No significant interactions were found.

A second set of analyses of variance was carried out to determine the effect of sentence voice—active versus passive—on EVS. The other factors investigated in these analyses were subject grades, speed of reading (slow-fast), and whether the light was turned out before or after the first major immediate constituent division of the sentence (light-out position). Again second and fourth grades and eighth and tenth grades plus adults were analyzed separately, because of the use of different stimulus materials for these groups. In the analysis of the younger subjects, the main effect of subject grade-level was significant ($F = 40.63$, df = 1/16, $p < .01$) as was the main effect of reader speed (= 7.63, df = 1/16, $p < .01$). Sentence voice and light-out position were not significant ($F = .20$, df= 1/16, n.s. and $F = .48$, df = 1/16, n.s., respectively). The interaction between voice and light out position was significant, however, ($F = 5.61$, df = 1/16, $p < .05$). Figure 2a shows that the EVS tended to be longer <u>before</u> the verb in the passive sentence and longer <u>after</u> the berb in the active sentence. No other

Figure 2a

interactions were significant.

For older subjects, the main effects of subject grade and reader speed were significant ($F = 9.17$, $df = 2/24$, p .01 and $F = 25.12$, $df = 1/24$, $p < .01$, respectively). Although the effect of voice was not significant at the younger grade levels, it was found to be significant for older subjects ($F = 22.65$, $df = 1/24$, $p < .01$). The mean EVS was longer for the passive than for the active sentences. The light-out position, before or after the verb, was not significant ($F = 2.62$, $df = 1/24$, n.s.). Again, the only significant interaction was between voice and light-out position ($F = 27.83$, $df = 1/24$, $p < .01$). Figure 2b shows that the eye-voice span

---

Figure 2b

---

is longer before the verb in the passive sentence and after the verb in the active sentence.

<u>Unit versus Non-unit reading</u>. In order to test whether there was a significant tendency for subjects to read in phrase groupings, the number of times each subject read to the end of a phrase unit on each of the sentence types was recorded. This score was corrected for any tendency of subjects to read to phrase boundaries only when their modal EVS took them there. The number of times a subject read to the end of a phrase with his modal EVS (with his two consecutive, most frequent EVSs) was subtracted from the overall total. If it were the case that subjects tended to read to the end of phrase units only when their modal EVS ended there, then the sum of all the scores computed as described above would be zero. The overall mean number of times subjects read to phrase boundaries, over and above the times they read to phrase boundaries with their modal EVS, was 8.20 which was found to be significantly greater than

zero ($t = 16.73$, $df = 59$, $p < .001$).

Another way of testing the hypothesis that subjects tended to read to the end of phrase boundaries more frequently than to non-boundary positions is to compare the number of times subjects read to phrase boundaries with the number of times subjects read to non-boundary positions divided by the number of chances to read to non-boundary positions in the phrase. Thus, in the case of the three-word phrase, there would be one chance to stop at a phrase boundary for every two non-phrase boundaries. Thus the number of times subjects stopped at a phrase boundary would be compared to the number of times they stopped at the non-phrase boundary divided by two. In an overall comparison of boundary versus non-boundary reading, it was found that subjects read to boundaries significantly more times than to non-boundaries ($t = 22.75$, $df = 8$, $p < .001$).

A two-way analysis of variance including all five grade levels did not show either subject grade level or the number of words in a chain to have a significant effect on the frequency with which subjects read to phrase boundaries. A t-test did, however, show a significant difference between the number of times the older subjects read to phrase boundaries as compared to the number of times second graders read to phrase boundaries ($t = 2.66$, $df = 18$, $.01 < p < .02$). Another t-test indicated a significant difference between slow and fast readers ($t = 2.84$, $df = 8$, $p < .01$), with fast readers reading in phrase units more frequently than slow readers.

In addition to the 492 times subjects read to phrase boundaries when the phrase boundary was <u>not</u> at the end of the modal EVS, subjects changed the sentence structure or the last word read 107 times in such a way as to make a phrase boundary. Thus, for example, if the final stimulus phrase was '... next to the house,' the subjects might read '....next-door.'

The number of times subjects were able to correctly recognize words beyond the point at which they stopped reading minus the incorrect recognitions was 578 times. An analysis of variance was carried out on the recognition scores for all five grade levels of slow and fast readers on the four chain types. Both the main effect of subject grades and reader speed were significant ($F = 7.62$, $df = 4/12$, $p < .01$ and $F = 22.6$, $df = 1/12$, $p < .001$, respectively). The main effect of length of phrase was not significant. It is evident from Figure 3 that the number of times

Figure 3

subjects recognized words correctly which were beyond their stopping points increased from second grade through adults for both slow and fast readers.

SUMMARY AND DISCUSSION

The results of the present study support the hypothesis that subjects tend to read in phrase units. Subjects read to phrase boundaries a significant number of times over and above the number of times their modal EVS took them to the end of a phrase. This suggests that readers have an elastic span, which stretches or shrinks far enough to read to phrase boundaries. There was no difference in the number of times subjects read to phrase boundaries on the four different types of sentences: thus, the finding that subjects read to phrase boundaries cannot be a function of the facilitative effect of a particular chain length. The finding that older subjects read to phrase boundaries more times than did the second grade subjects, suggests that beginning readers tend to read more perceptually, more word by word, than do older subjects.

The phrase-unit reading hypothesis was further supported by the observation that subjects, not infrequently, made up unit or phrase endings so that they stopped reading at the end of a completed unit even if they had not actually seen the end of the phrase boundary on the printed page.

Fast or good readers read to the end of phrase boundaries more often than did slow or poor readers. Thus good readers seem to be processing more in terms of units or phrases and their EVS seems to be more adaptable to the structure or content of the reading material. The slow readers, like the beginning readers, may be reading more in terms of what Anderson and Swanson (1937) call "perceptual" factors, i.e. they tend to be reading every word individually and taking advantage of the contextual constraints.

The findings with respect to the length of the EVS tend to support earlier findings. The EVS did tend to increase with age. There was, however, a slight inconsistency in this trend in that eighth grade subjects tended to read with longer spans than did the tenth graders. As she subjects volunteered from their study halls for a reading experiment, there was no control as to the ability of the readers obtained. Therefore, this deviation from the tendency for the EVS to increase with age is probably due to a biased sample at both grade levels.

Another finding which confirms previous research is that the EVS for unstructured or word list material is significantly shorter than for structured sentences. The fact that this difference exists suggests that all readers, both slow and fast, must take advantage, to some degree, of the contextual constraints of the material they are reading. The more structured the material is, the less the subject has to focus, in detail, on every letter or every word in its entirety, as he reads across the page.

The older subjects recognized more words beyond the point at which they stopped reading than did the younger subjects. This suggests that the older subjects seem to be taking in more of the peripheral material, beyond the EVS, than do the younger subjects. It has been shown that older subjects have longer spans, but now it seems that in addition to these longer spans they have more highly developed peripheral vision or that their memory for material which has been picked up peripherally decays more slowly than that of younger subjects. The difference between the recognition scores of slow and fast readers was also highly significant. Fast readers recognized significantly more words beyond their EVSs than did slow readers. Thus the more skilled and experienced reader can pick up more information peripherally.

An interesting finding which is difficult to explain is that subjects at all grade levels, except the youngest grade---the second grade--tended to have the longest EVS on the three-word phrase sentences. If it is the case that subjects are reading in terms of one phrase or two phrases grouped together, they would make more fixations with shorter spans ahead in getting through a two-word phrase sentence than in getting through a three-word phrase sentence. However, it should then be the case that the reader would have fewer fixations and longer spans still in getting through a four-word phrase sentence. This was not the case; the EVS is longest on the three-word phrase sentences and shorter on both the two-word phrase sentences and the four-word phrase sentences. A possible explanation might be that if the subject is "phrase-reading" in a four-word phrase sentence and the light is turned out before or after the first word in a phrase, he would have to read either four or eight (or three or seven) words in order to read in phrase units. The seven or certainly the eight words would most likely be well beyond the subject's EVS, <u>and</u> beyond the elastic limits of the EVS. Therefore, the subject would probably tend to shrink his EVS one or two words and read only to the end of the nearest phrase boundary. As the longest two phrase sequence in a three word phrase sentence would be six words (much nearer the modal EVS of most subjects), it would be more likely that the subject's EVS would stretch to take in the two phrase sequence rather than shrink to take in just the two or three word sequence. In the two-word phrase sentence, the longest sequence for two phrases would be four words. The fact that the tendency for the longest EVS to be found on three-word phrase sentences was absent in the second grade in addition to the fact that second graders read to phrase boundaries less frequently than did older subjects, suggests that beginning readers are less sensitive to the grammatical and semantic structure of what they are reading than are older subjects.

Another factor which had a significant effect on the EVS was sentence voice. There was no significant difference between the lengths of the EVS on the active and passive sentences for second and fourth grade subjects, but there was a significant effect for older subjects. In the latter cases, the EVS tended to be longer for passive voice sentences than for active voice sentences. There was, in addition, a significant interaction between position of light out (before or after the verb) and sentence voice for both sets of subjects. In active sentences, the longest EVS tended to occur _after_ the verb, while in passive voice sentences the longest EVS tended to occur _before_ the verb. In both cases, the part of the sentence in which the longest EVS occurred, was in the part containing the object. In the active sentence the order of elements is "actor--verb--object" and in the passive sentences the order is "object--verb--actor." In an investigation of the uncertainty of various elements in active and passive voice sentences, Clark (1965) found that in both the active and passive sentences, the verb was significantly more constrained by the object than by the actor and that the object was significantly more constrained by the verb than by the actor. This suggests a possible explanation for the differences in length of EVS before and affer the verb in active and passive sentences. The longer EVS occurs in both sentence types in that part of the sentence where the mutual constraints of the elements are strongest. This interpretation would agree with the findings of Morton (1946a) and Lawson (1961) who found that the EVS increased with an increase in contextual constraint.

References

Anderson, I. H. Eye-movements of good and poor readers. Psychol. Monog., 1937, 48, 1-35.

Anderson, I. H. and Swanson, D. E. Common factors in eye-movements in silent and oral reading. Psychol. Monog., 1937, 48, 61-69.

Buswell, G. T. An experimental study of eye-voice span in reading. Suppl. Educ. Monog., 1920, no. 17.

Cattell, J. McK. Mind, 1889.

Clark, H. H. Some structural properties of simple active and passive sentences, J. verb. Learn. verb. Behavior., 1965, 4, 365-370.

Fairbanks, G. The relation between eye-movements and voice in oral reading of good and poor readers. Psychol. Monog., 1937, 48, 78-107.

Huey, E. B. The psychology and pedagogy of reading. New York: MacMillan, 1922.

Lawson, Everdina. A note on the influence of different orders of approximation to the English language upon eye-voice span. Quart. J. Exp. Psychol., 1961, 13, 53-55.

Morton, J. The effects of context upon speed of reading, eye-movements and eye-voice span. Quart. J. Exp. Psychol., 1964a, 16, 340-354.

Morton, J. A model for continuous language behavior. Lang. Speech., 1964b, 7, 40-70.

Quantz, J. O. Problems in the psychology of reading. Psychol. Monog., 1897.

Schlesinger, I. M. Sentence Structure and the Reading Process. In press, 1966.

Tinker, M. A. Recent studies of eye-movements in reading. Psychol. Bull., 1958, 55, 215-231.

Table 1

Positions of Light Out for Two-word Phrases

/ 1 2 3 4		5 6 7 8
1 / 2 3 4		5 6 7 8
1 2 / 3 4		5 6 7 8
1 2 3 / 4		5 6 7 8
	First Major IC	
1 2 3 4		/ 5 6 7 8
1 2 3 4		5 / 6 7 8
1 2 3 4		5 6 / 7 8
1 2 3 4		5 6 7 / 8

/ = light out position

Table 2

Summary of Mean Eye-voice Spans for Experimental Conditions

Reader Speed:	Second Grade		Fourth Grade		Sixth Grade		Eighth Grade		Tenth Grade		Adults	
	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast	Slow	Fast
Chain Length												
2-word chain	2.70	3.62	4.18	4.30	1.97	2.98	3.45	4.53	3.45	4.18	4.74	5.05
3-word chain	2.73	3.63	4.57	4.87	2.47	3.42	3.62	5.33	4.14	4.60	4.67	6.17
4-word chain	2.89	3.65	4.21	4.32	2.03	3.11	3.16	5.02	3.51	3.81	4.49	4.99
Mean by Grade	3.19		4.41		2.66		4.18		3.95		5.02	

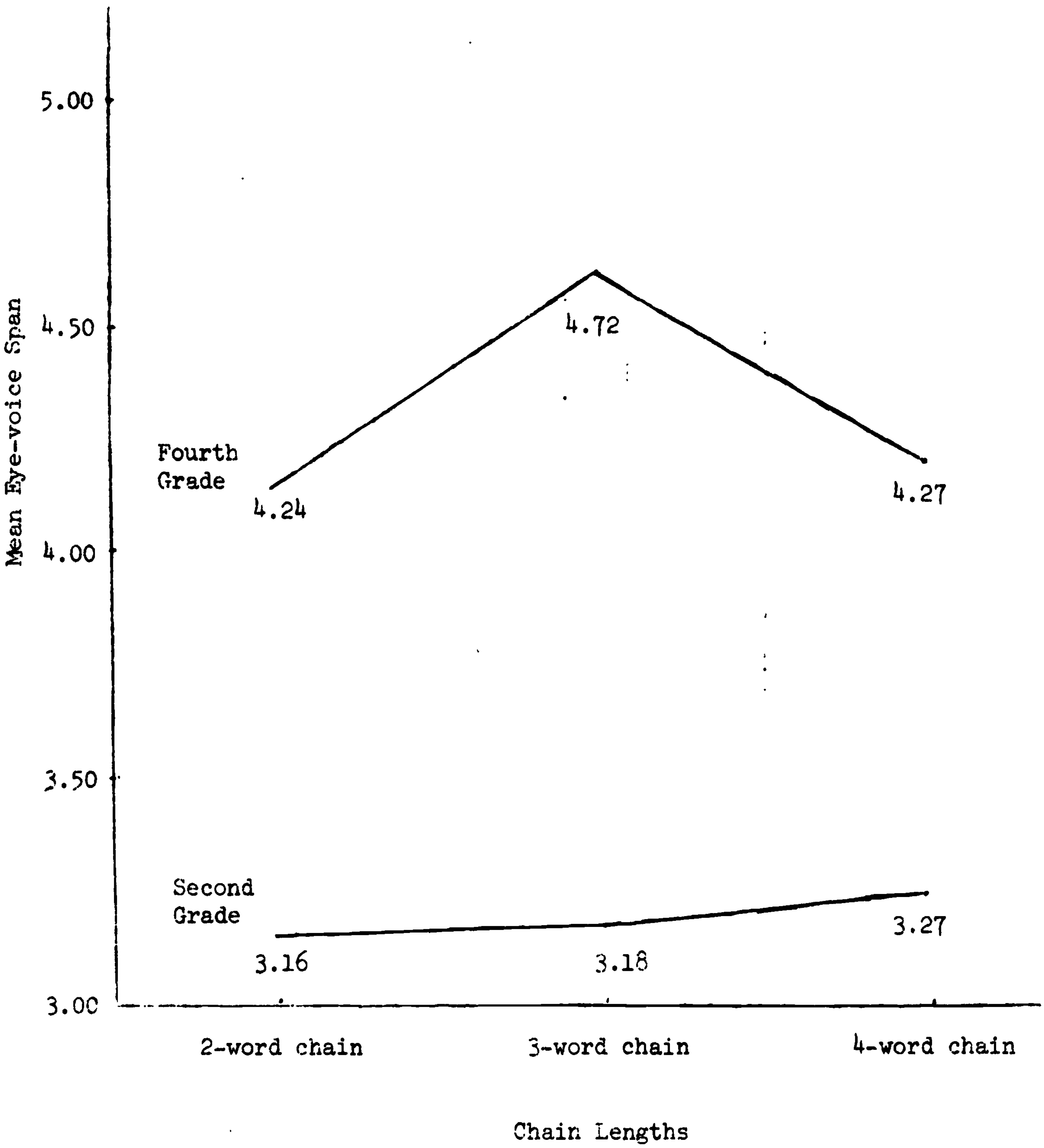


Figure 1a. Second and Fourth Grade Mean Eye-voice Spans for All Chain Lengths

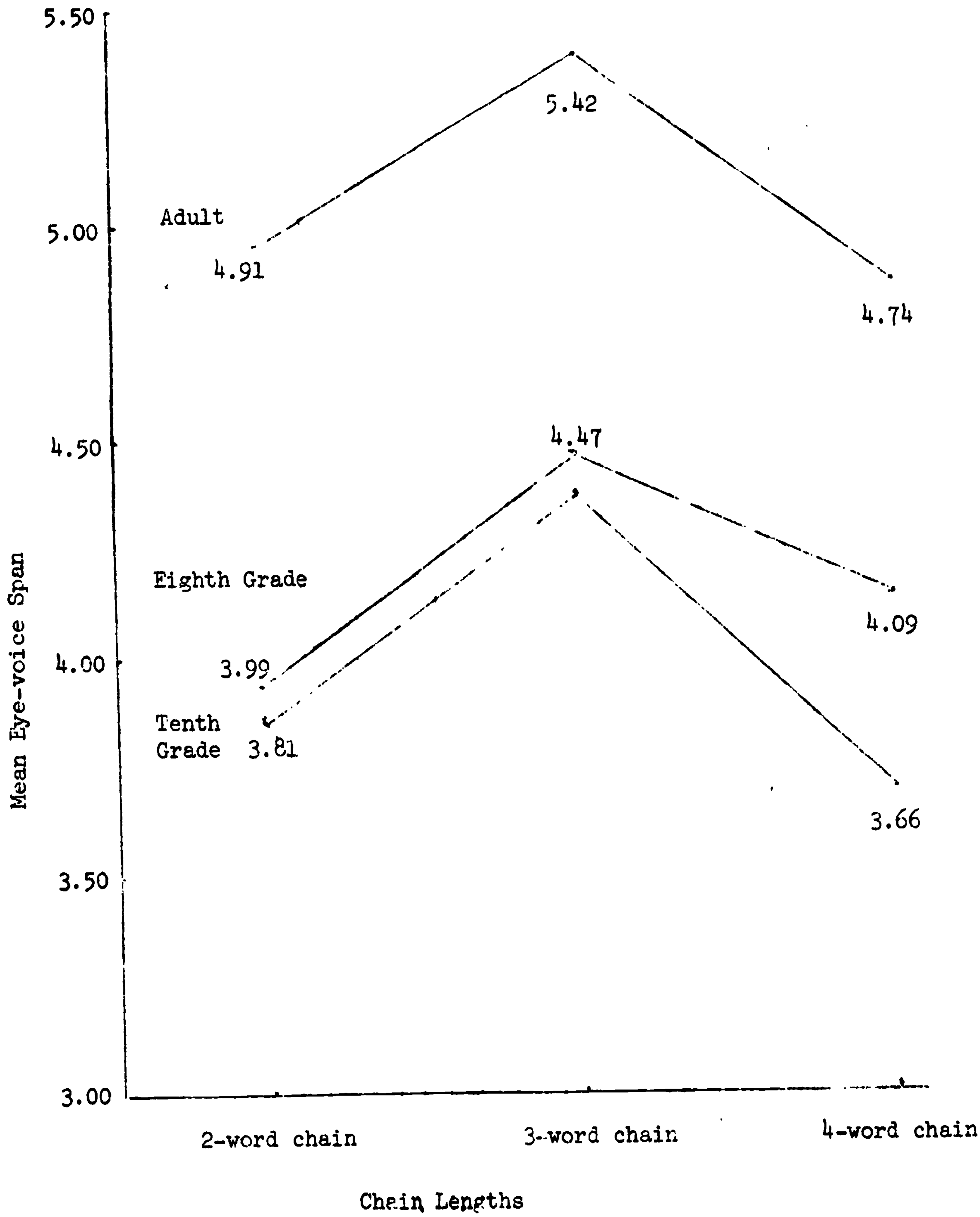


Figure 1b. Eighth, Tenth, and Adult Level Mean Eye-voice Spans for all Chain Lengths

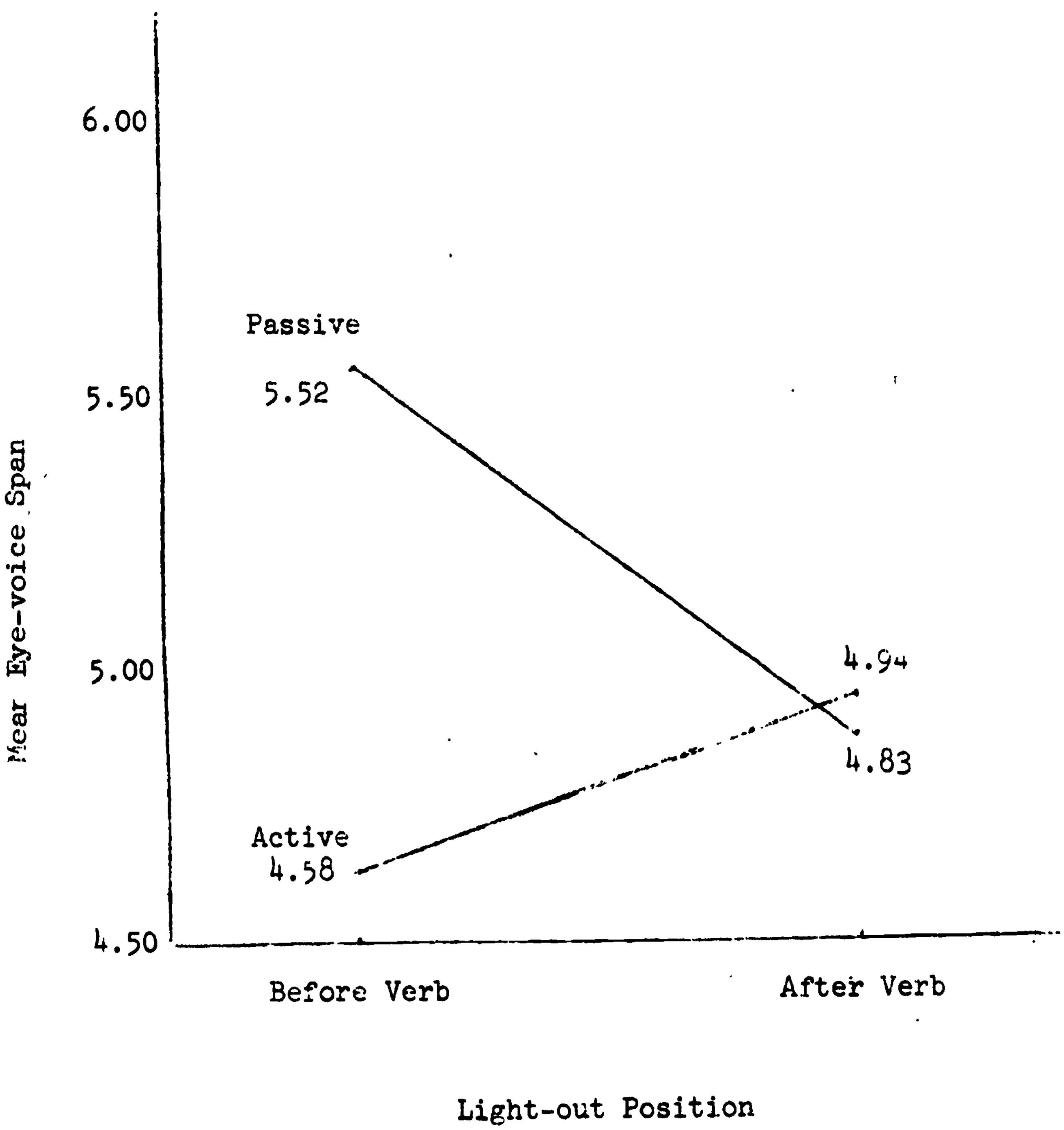


Figure 2a. Second and Fourth Grade Mean Eye-voice Span for Light-out Position by Voice.

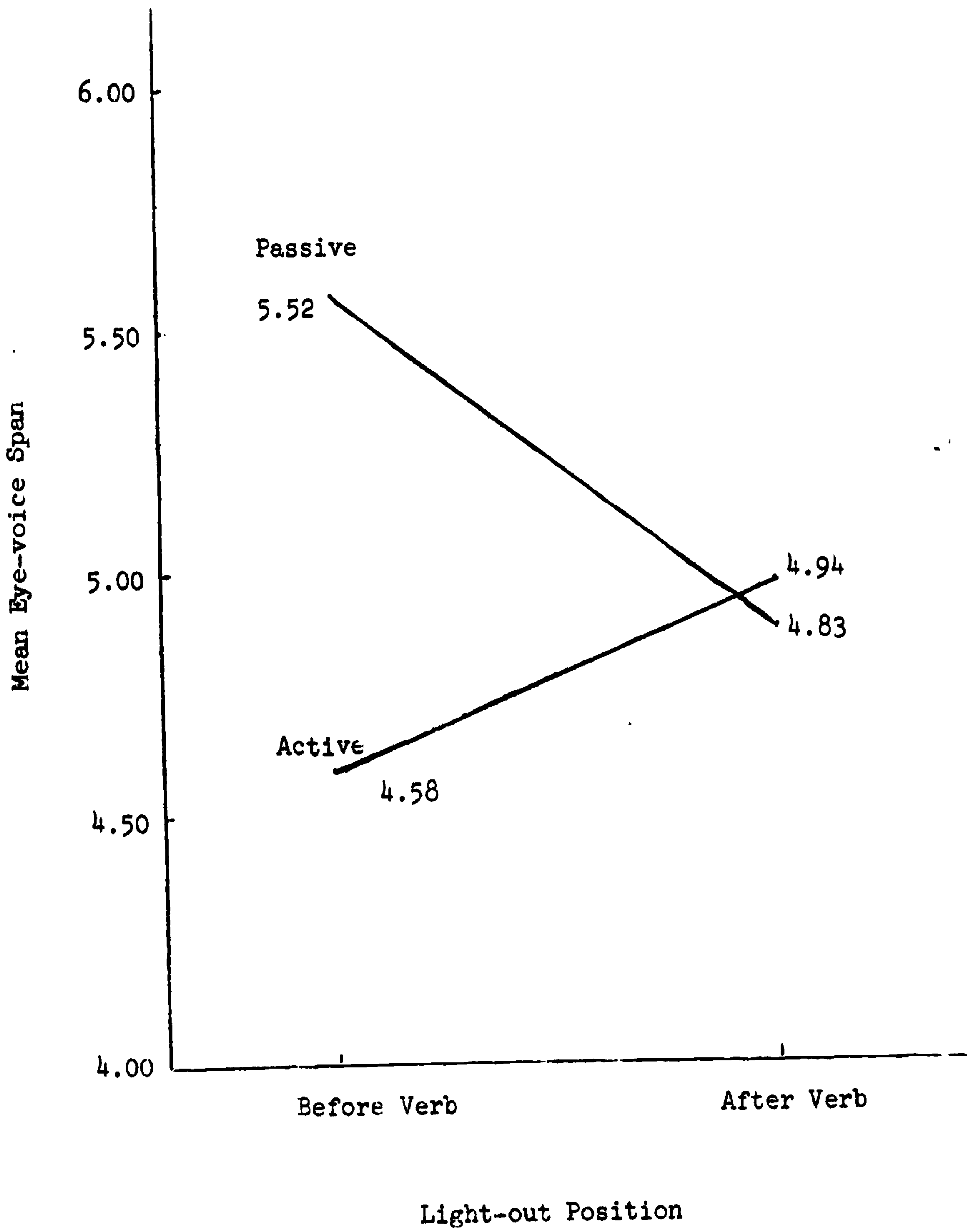


Figure 2b. Eighth, Tenth, and Adult Level Mean Eye-voice Span for Light-out Position by Sentence Voice.

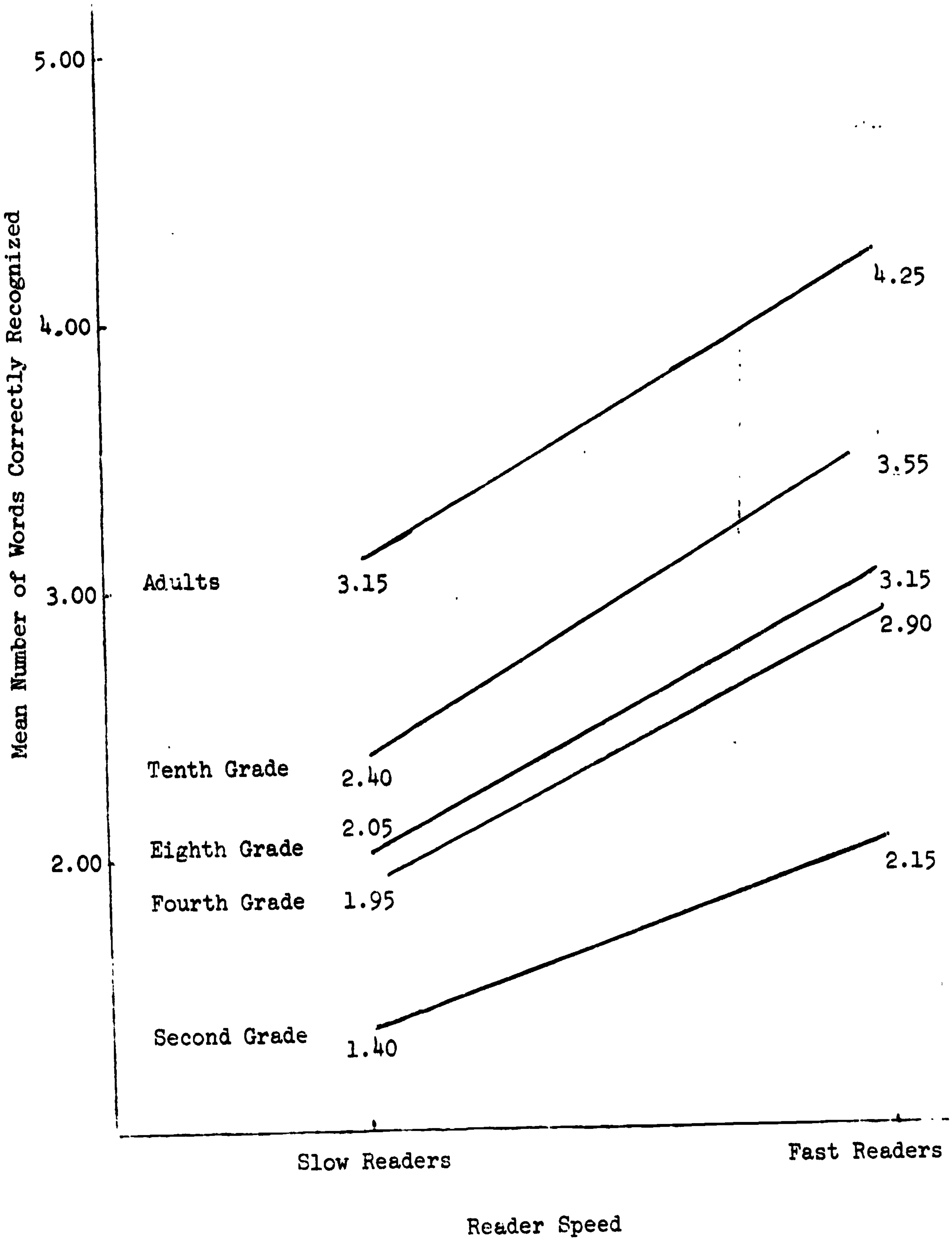


Figure 3. Mean Number of Words Correctly Recognized for All Grade Levels of Slow and Fast Readers

Preliminary Draft
September, 1966

Studies of Oral Reading

X. The Eye-Voice Span for Active and Passive Sentences

Harry Levin and Eleanor Kaplan[1]

Cornell University

Recent studies (Schlesinger, 1966; Levin and Turner, 1966; Morton, 1964; Lawson, 1961) have suggested that there is a strong relationship between the structural properties of written material and the nature of the decoding process in oral reading. They have argued that the span of the eye ahead of the voice (EVS) represents a unit of decoding.[2] Schlesinger (1966) and Levin and Turner (1966) proposed that this unit is in part determined by the syntactic structure of the reading material. The results of both studies indicate that the size of the EVS is not constant, but varies in accordance with the size of intra-sentence constituents. Lawson (1961) and Morton (1964) found that the size of the EVS was directly related to the information content of the written material. That is, the EVS increased with increasing orders of statistical approximations to English. Since the manipulation of statistical approximations is equivalent to the manipulation of sentence redundancies, it seems likely that the EVS should also be sensitive to the manipluation of other structural variables which have been shown to affect within sentence constraints.

1. This study was supported by funds from the U.S. Office of Education. We acknowledge with thanks the contribution of computer facilities by the Bell Telephone Laboratories, Inc., Murray Hill, N.J.

2. A more complete historical survey of the literature on the EVS is given in Levin and Turner (1966).

Active and passive sentence forms were selected for study since there is both psychological and linguistic evidence showing that the within sentence constraints differ between the two types of sentences. Clark (1965) in an examination of the contingencies among sentence parts found the uncertainties associated with the actor, verb and object and the constraints between them to be different for active and passive sentence forms. By having Ss complete sentences from which certain grammatical forms, e.g. actor, verb and object, had been deleted, Clark found that there was less uncertainty and higher grammatical constraint associated with the passive form. These findings were also supported by later studies (Clark, 1966; Roberts, 1966). The latter demonstrate that recall for different sentence parts is dependent upon these uncertainties and contingencies.

We expect that the EVS will be longer in the more highly constrained sentence form, the passive, than in the less constrained form, the active. Levin and Turner (1966) found longer EVSs overall for passive than for active sentences, across various age groups. More specifically, the EVS should increase in size when the reader reaches information which specifies that the sentence is a passive. This should occur when the "eye" reaches the "by" phrase in the passive sentence. There is no reason to believe that there is information for this decision prior to the beginning of this phrase.

Secondly, it is predicted that the EVS will vary in accordance with both the location and size of the constituents within the sentence, regardless of sentence type. This is in keeping with Schlesinger's (1965) and Levin and Turner's (1966) results which demonstrate that the EVS is most often extended to the end of within-sentence constituent boundaries.

The previous data, then, indicate that the EVS is longer for constrained or predictable written materials and that passive sentences are more constrained than active ones. The meshing of these two sets of findings provides the hypothesis for the present study.

Method

Subjects. Eighteen college students attending Cornell University served as subjects.

Experimental Sentences. Two different phrase lengths and two different sentence types were selected for study, which in combination comprised the four sentence types used:

Active sentences composed of 4-word phrases.

Passive sentences composed of 4-word phrases.

Active sentences composed of 5-word phrases.

Passive sentences composed of 5-word phrases.

The 4-word sentences contained 19 words, broken up into five phrases or constituents: 4 words, 4 words, 3 words, 4 words, 4 words. The 5-word sentences contained 18 words divided into four phrases or constituents: 5 words, 5 words, 3 words, 5 words. The short 3-word phrase represented the "by" phrase in the 4- and 5-word passive sentences, and a short prepositional phrase in the 4- and 5-word active sentences. The sentences were constructed so that the first half of both active and passive sentences were identical. For example:

Passive: The cute chubby boy was slowly being wheeled by the maid along the pebbled lane to the quaint store.

Active: The brash tall man was certainly being loud at the rally of the new group on the main campus.

Differences in the size of the EVS could thus be attributed to the presence or absence of the passive form. Each sentence was embedded in a separate paragraph of either four or five sentences. The sentences within each paragraph were unconnected in order to prevent inflation of the EVS by S's ability to guess succeeding words on the basis of context. Since exploratory data indicated that Ss seemed to scan the first line before beginning to read aloud, the experimental or target sentence was never first. The target sentences were positioned so that there were at least three words preceding the critical word on the same line and at least eleven succeeding it for the 4-word sentences and 8 for the 5-word sentences. Results from an exploratory investigation indicated that these distances were sufficient. Few Ss extended their EVS to the last word in the sentence. Also, sentences extended beyond this limit of 19 or 18 words were too artificial.

In addition, 10 paragraphs made up entirely of lists of unrelated words were included in order to ascertain the relative contribution of syntactic structure to the variation in the EVS. Finally, an additional 20 paragraphs were included as fillers. The first sentence in these paragraphs was treated as the target sentence in order to encourage Ss to attend equally to all sentences. These sentences were not included in the data analysis. Thus there was a total of 142 paragraphs. Ss were assigned to 1 of 6 different random presentation orders.

Procedures. Paragraphs were exposed on a small ground glass rear projection screen directly in front of the S, who was postioned so that he could scan the lines with minimal head movement. The size of the letters projected on the screen was approximately equivalent to that found in texts. A fixation point indicating where the beginning of each paragraph would appear eliminated the problem of having the S search the screen each time a new paragraph was exposed. After-images were prevented by minimizing the light contrast.

When the paragraph appeared on the screen the S was instructed to begin immediately to read out loud at his normal reading rate. When the S reached a predetermined critical position in the target sentence, the E triggered a switch closing the shutter on the projector. When this occurred the S was instructed to report all the words he had seen but not yet read aloud. This was used as a measure of the EVS. A full record was made of where the S's voice was when the shutter was closed, by simultaneously recording the S's voice on one track of a stereo tape and the shutter click on the other track. The time taken to read to the critical position was recorded on a Hunter KlocKounter which was activated by the opening and closing of the shutter. The fast shutter speed eliminated any light decay period.

Critical Positions. There were 6 critical positions each for the 4-word active and 4-word passive sentence types, and 8 critical positions each for the 5-word active and 5 word passive sentence types. There were four sentences for each critical position, making a total of 24 passive and 24 active 4-word sentences and 32 passive and 32 active 5-word sentences. In the 4-word sentences this occurred after the third, fourth, fifth, sixth, seventh and eighth words in the target sentence. In the 5-word sentences this occurred after the third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth words in the target sentence. These positions were selected on the basis of results from an exploratory investigation. Although it would have been informative to include critical positions occurring after the "by" phrase this was not possible because of the length of the experimental session.

Recognition lists. In order to obtain another measure of the EVS a series of recognition lists were constructed one for each target sentence. All content words occurring after the critical word were included on the list. Function words were omitted since such words occurred more than once in a given target sentence making it difficult to know which function word was being referred to. In order to discourage guessing the lists were constructed so that only 50% of the words actually appeared in the target sentence, while the other 50% were words that were preceptually and semantically similar to those which had appeared in the sentence. That is, the first three or four letters in this set of words were identical to those that had actually appeared; in addition, they were semantically similar in that they could be readily substituted for the words that had appeared in the text. After reporting all words seen but not yet read aloud, the S searched through the list of words circling any additional words that he recognized as having appeared in that sentence. The S was informed that only half of the words on the list had actually appeared in the sentence.

RESULTS

Two measures were obtained for each S, (1) EVS_1, the number of correct consecutive words reported after the shutter was closed, and (2) EVS_2, the EVS_1 score plus the number of additional words correctly selected from the recognition lists. In the latter case only the recognition words selected that immediately followed to the right of those already reported were included in the EVS_2 measure. However, since only content words were included on the lists an S was given credit for any intervening function words. For example, in the sentence, "The cute little girl with the long hair..." if the cirtical position occurred after little, and the S reported having seen girl as well as recognizing long and hair, he would be credited with an EVS_2 of 5 words. This procedure was followed since it was assumed that any S who had read the content words would also have read preceding function words. The recognition EVS was not corrected for guessing since the ratio of incorrect to correct recognition word choices was approximately 1 to 1,000.

The EVS_1 scores were computed by taking the mean of the four scores obtained from each S at each of the critical positions. Thus for the 4-word sentences there were 6 EVS_1 scores and for the 5-word sentences there were 8 EVS_1 scores. Due to unpredictable variations in the Ss' reading rates, there were a few instances where the shutter was not closed at the predetermined critical position. In such cases where there were less than four scores per critical position for any one S, the mean of the other three scores was taken as the value of the fourth score.

In all of the analyses to be reported, the original EVS scores (EVS_1) and the EVS plus recognition (EVS_2) scores yield completely redundant findings. As might be expected, EVS_2 score containing both recall and recognition components, is higher than the EVS score alone. In the four word-phrase sentences, the mean EVS_1 is 4.5 words and EVS_2 is 4.9 words; for the five word phrase sentences, the two mean scores are 4.0 and 4.4. It can be seen in Figure 1 that the two EVS scores parallel each other for the various critical positions in both types of sentences studied in this experiment.

Figure 1

Since the second EVS score yielded no new information, the subsequent analyses treat only the recall EVS score, which, incidentally, is the one used in the first study in this series.

Four-Word Sentences

Table 1 presents the results of a three-way analysis of variance with repeated measures for the 4-word sentences. The main effects, Subjects, Position, and Type are significant beyond the .001 level. Only the Position X Type interaction was significant ($p < .001$). The Scheffé test for comparison of treatment means reveals that Positions 1 through 4 are significantly

---

Table 1

---

different from Positions 5 and 6 ($p < .05$). Figure 2 shows a plot of Position X Type for Active and Passive sentences. Active and Passive sentences do not

---

Figure 2

---

differ for positions 1 through 4, but they do for critical positions 5 and 6. There are no differences among the critical positions in the active sentences. Among the passives, however, positions 1 through 4 differ significantly from Positions 5 and 6 at the .01 level. (See Figure 2)

Five-Word Sentences

A similar three-way analysis of variance for 5-word sentences again yielded three significant main effects, Subject, Position and Type. Again only the Position X Type interaction is significant ($p < .001$).

Table 2

The Scheffé test for differences among treatment means shows that Positions 1 through 6 are significantly different from Positions 7 and 8 ($p < .01$).

Figure 3 shows a plot of Position X Type means for Active and Passive Types.

Figure 3

Similar to the 4-word sentences there is no significant Active/Passive Type X Position differences for Positions 1 through 6, while there are for Positions 7 and 8 ($p < .01$). No positions in the Active sentences differ from each other. For the Passive Type, Positions 1 through 6 are significantly different from Positions 7 and 8 ($p < .01$).

The findings, then, are (1) in both sentences composed of four and five word phrases, the EVS span is longer at the two terminal critical positions for the passive than for the active sentences; (2) within the active sentences, the position in which the shutter was closed had no effect on the EVS; and (3) in the passive sentences, when the shutter was closed at the two final positions that we studied, the EVS were larger than critical positions earlier in the sentence. In general, these findings support the major hypothesis that the EVS varies in accordance with intra-sentence contingencies. The results show that the EVS is longer for Passive sentences at that point where the contingencies are highest (Clark, 1965, 1966). Since (1) the first half of both active and passive sentences were identical and (2) the short three-word phrase in the active sentence was a prepositional phrase where the intra-phrase contingencies would not be expected to differ from those within the passive "by" phrase (Aborn and Rubenstein, 1958; Treisman, 1965; Fillenbaum, et al., 1964), these differences must be attributed to the structure of the sentence type as a whole.

Additional support for the notion that the EVS is sensitive to syntactic structure is the finding that the EVS remains fixed for the sentences composed of strings of unrelated nouns and verbs. The EVS obtained from these "sentences" was 2.0 words at all of the critical positions in comparison with longer EVS of real sentences and the variation of EVS across the different critical positions for the passive sentences.

Constituent Boundary Effects

To evaluate the tendency for the EVS to vary in accordance with both the size and location of the constituent boundaries within sentences, the frequencies with which Ss read to boundary and to non-boundary positions were compared. Both scores were corrected for the tendency of Ss to read to a given position only when their most common EVS, their modal EVS, happened to fall there. The number of times a S read to a given point in the sentence with his modal EVS was subtracted from the total number of times he read to that same point. In order to compare the two scores, they were weighted by the likelihood of S's stopping at either boundary or non-boundary positions. For example, the ratio of boundary to non-boundary positions in the four word sentences is 1:3. However, due to the short three word phrase in each sentence this ratio is not constant across all critical positions; it varies from 1:3 to, 1:2. A conservative procedure was adopted by weighting the score by the smaller of the two ratios. Thus for the 4-word sentence the score representing the frequency with which Ss read to boundary positions was divided by 1 and that for the non-boundary positions was divided by 2. The 5-word sentence was treated similarly.

Overall, Ss read to boundary positions 361 times and to non-boundary positions 202.7 times. This difference is significant ($p < .001$; $t=3.98$, 2-tailed).

Concluding Statement

The two findings which emerged from this study are that EVS is longest at that point in passive sentences where the constraints are highest and that the EVS tends to terminate at phrase boundaries. Both results, it seems to us, lend themselves to a common interpretation.

That is, the amount of written material which is picked up and processed depends on the constraints or the predictability of the message. As Clark's (1965) research has shown, passive sentences are highly predictable from the point of the "by phrase." Likewise, the constraints within sentence constituents are stronger than those between constituents. Such a general formulation based on predictability permits us to hypothesize about the size of the EVS span both within and across sentences of various types. These preditions will be tested in subsequent studies.

References

Aborn, M., Rubenstein, H. & Sterling, T. "Sources of contextual constraint upon words in sentences." J. Exp. Psychol., 1959, 57, 171-180.

Clark, H. "Some structural properties of simple active and passive sentences." J. Verb. Learn. Verb. Behav., 1965, 4, 365-370.

Clark, H. "The prediction of recall patterns in simple active sentences." J. Verb. Learn. Verb. Behav., 1966, 5, 99-106.

Fillenbaum, S., Jones, L.V. & Rappoport, A. "The predictability of words and their grammatical class as a function of rate of deletion from a speech transcript." J. Verb. Learn. Verb. Behav. 1964, 2, 186-194.

Lawson, Everdina. "A note on the influence of different orders of approximation to English language upon Eye-Voice Span.' Quart. J. Exp. Psychol., 1961, 13, 53-55.

Levin, H. & Turner, Ann. "Studies in Oral Reading: IX. Sentence Structure and the Eye-Vocie Span." mimeograph. 1966.

Morton, J. "The effects of context upon speed of reading, eye-movements and eye-voice span." Quart. J. Exp. Psychol., 1964, 16, 340-354.

Roberts, K. "The interaction of normative associations and grammatical factors in sentence retention." Paper read at MPA, Chicago 1966.

Schlesinger, I.M. "Sentence structure and the reading process." Unpubl. Monograph, 1965.

Treisman, Anne, "verbal responses and contextual constraints in language." J. Verb. Learn. Verb. Behav., 1965, 4, 118-128.

Table 1. Anova 4-word EVS

SOURCE	SS	d.f.	MS	F
Ss	465.02	17	27.35	12.45*
Pos	58.11	5	11.62	5.10*
T	48.21	1	48.21	15.96*
Ss X Pos	194.41	85	2.28	1.04
Ss X T	51.35	17	3.02	1.37
T X Pos	72.83	5	14.57	5.63*
Ss X T X Pos	220.04	85	2.59	1.17
Within	1423.26	648	2.19	
Total	2533.24	863		

Table 2. Anova 5-word EVS

Source	SS	d.f.	MS	F
Ss	549.49	17	32.32	16.81*
Pos	144.59	7	20.66	11.59*
T	167.60	1	167.60	52.88*
Ss X Pos	212.04	119	1.78	0.93
Ss X T	53.88	17	3.17	1.62
T X Pos	116.26	7	16.61	9.85*
Ss X T X Pos	200.64	119	1.69	0.88
Within	1661.50	864	1.92	
Total	3105.99	1151		

* Significant beyond the .001 level

Pos: Position
T: Type

Figure 1. Comparisons of Recall EVS_1 and Recognition EVS_2.

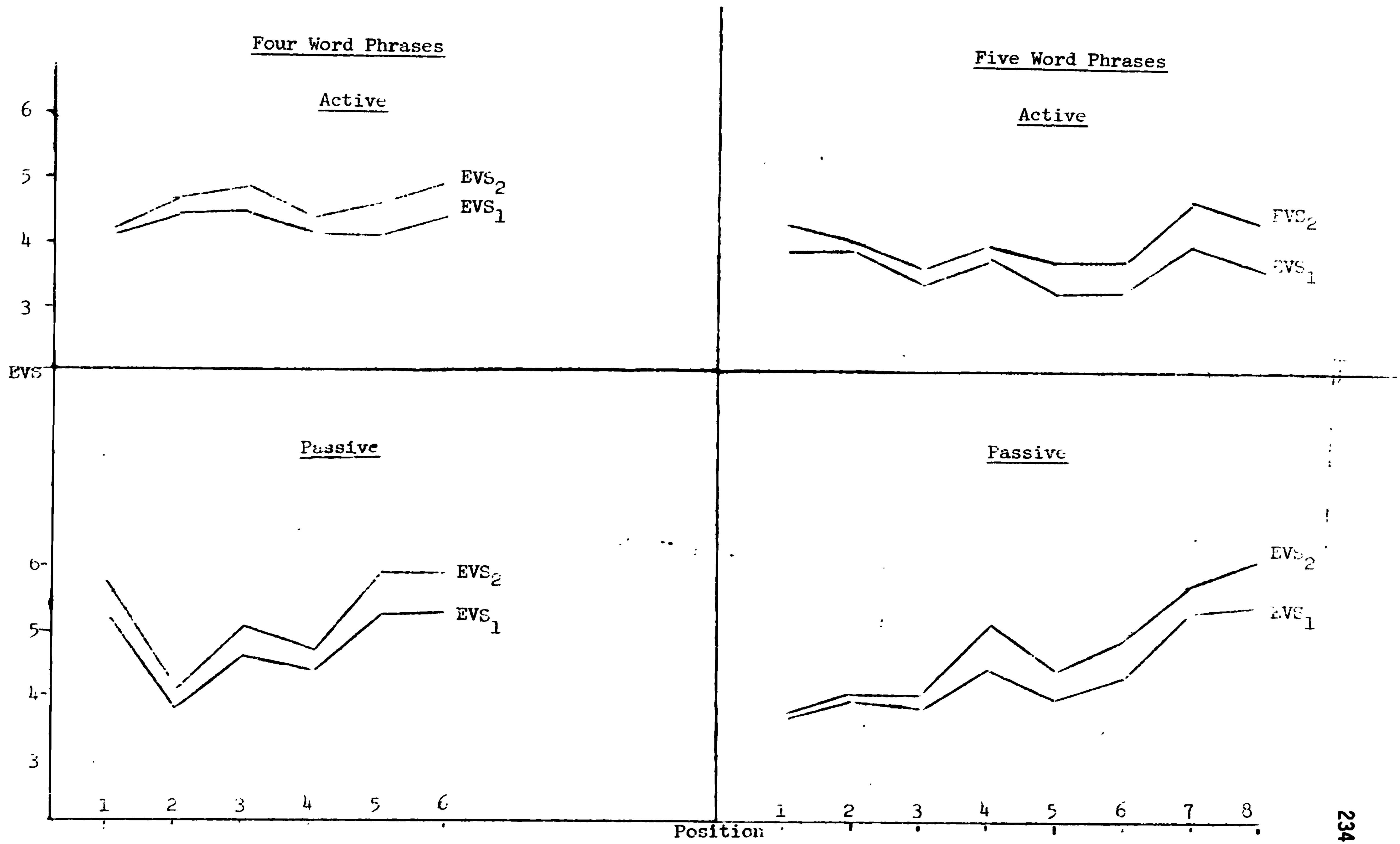


Figure 2. Mean EVS for Various Critical Positions: Four Word Phrase Active [1] and Passive sentences.

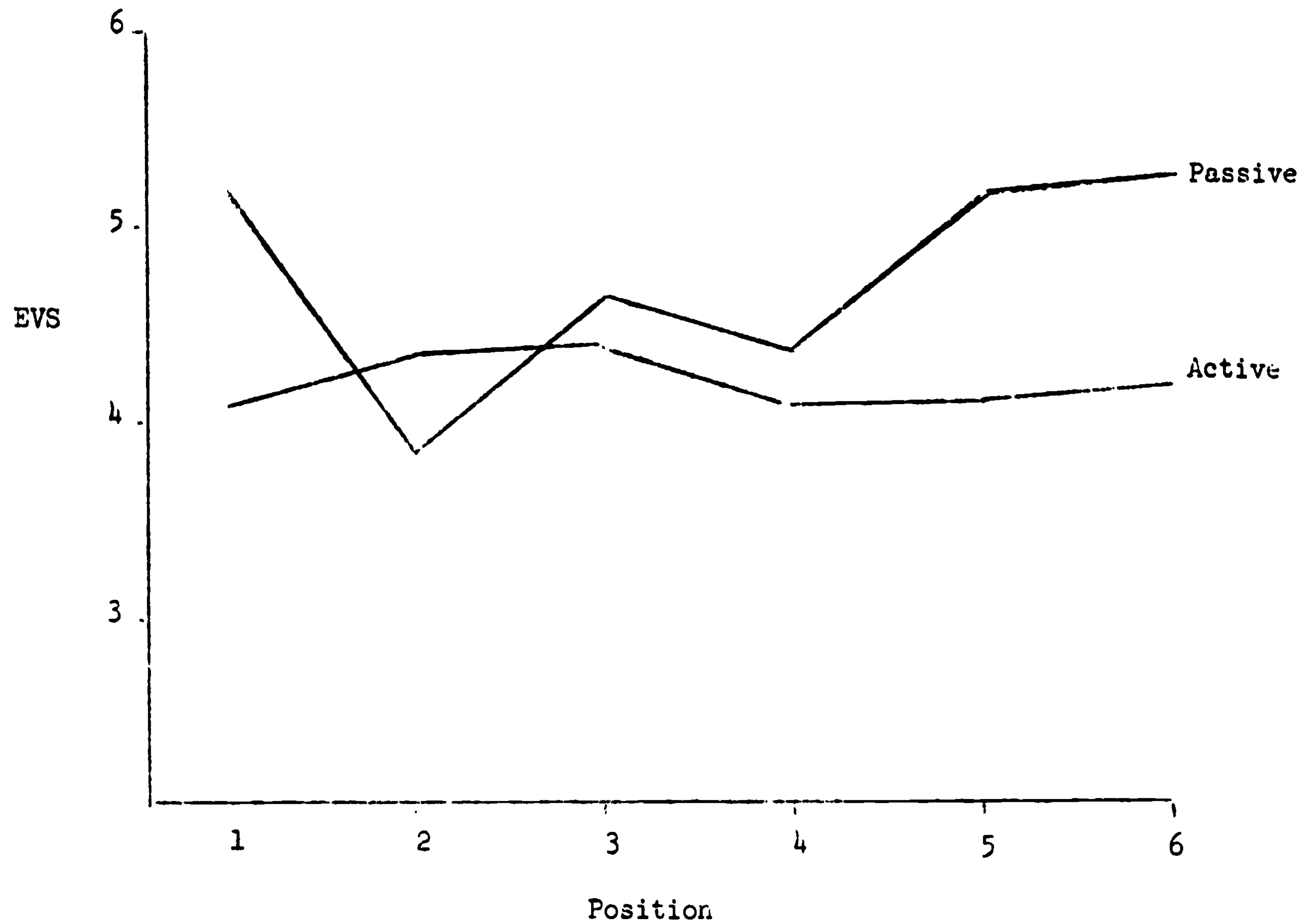
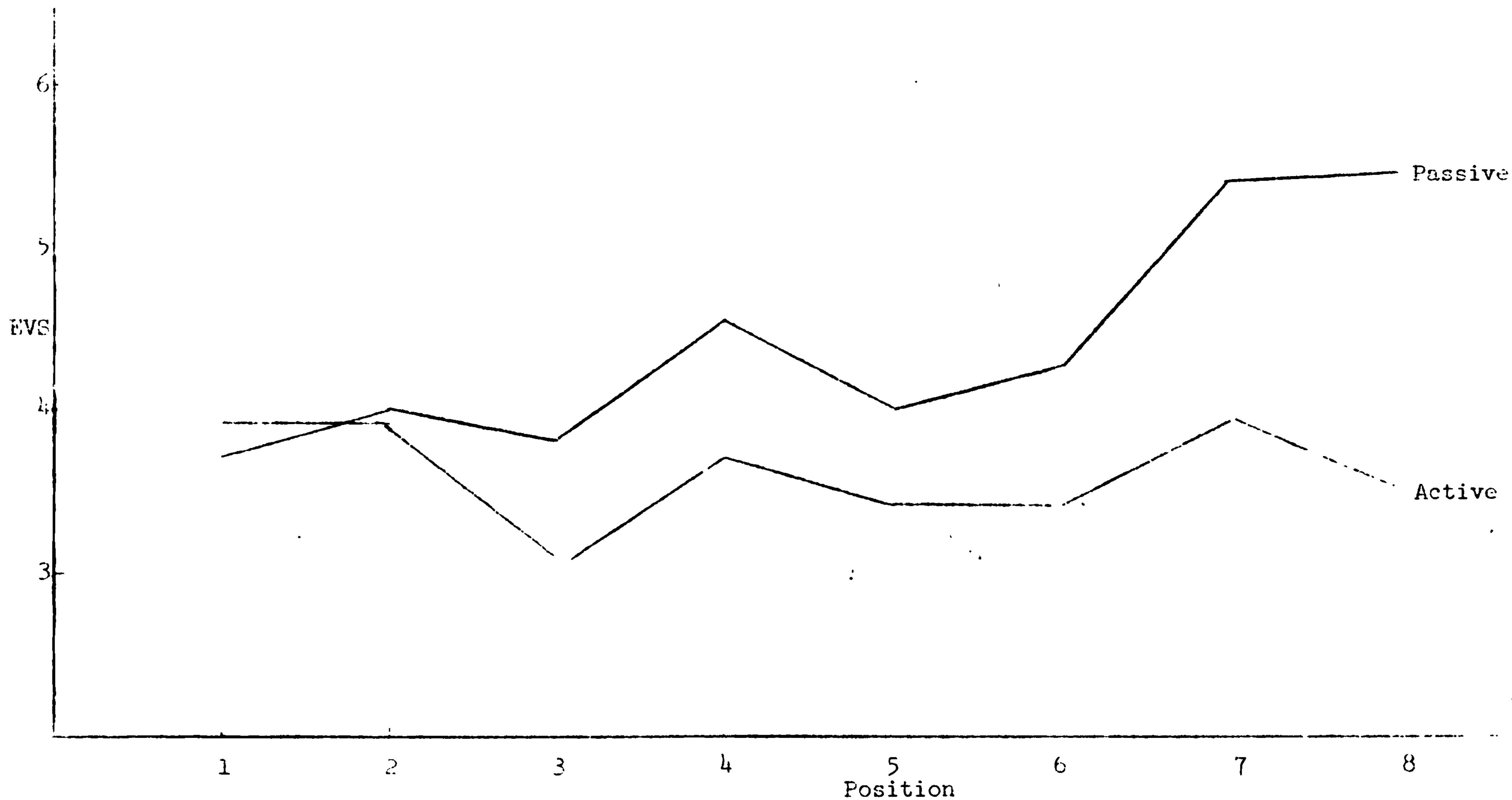


Figure 3. Mean EVS_1 for Various Critical Positions: Five Word Phrase Active and Passive Sentences.



PRELIMINARY DRAFT
June 1967

Studies of Oral Reading:

XI. The Eye-Voice Span: Reading Efficiency and Syntactic Predictability[1]

Stanley Wanat and Harry Levin

Cornell University

ABSTRACT

This experiment studies the relationships between reader efficiency in processing sentences and differences in the deep structure of the stimulus sentences. The efficiency of processing, as measured by Ss' Eye-Voice Span, was found to vary with changes in the deep structure. Comparisons were made between reader processing of pairs of sentences in which the surface structure was the same, but in which the deep structure was different. The Eye-Voice Span measure was found to validly discriminate between sentences with the same surface structure but with differing deep structure. The results were interpreted to very tentatively suggest that the efficiency of reading processing is (1) a function of the 'congruence' or 'constraints' between the surface structure and the deep structure of the sentence, and also (2) a function of the number of structural 'categories' required in the deep structure.

1. This research is supported by funds from the U. S. Office of Education.

This experiment studies the relationships between efficiency in reading and differences in the deep structure of the stimulus sentences. This, the experiment seeks to determine if differences at the deep structure level of sentences with the same surface phrases structured will effect the reader's processing behaviors — that is, (1) "Is sensitivity to deep structure reflected in reading processing?" Further, the experiment seeks to determine what effect the contrasting deep structures assigned to sentences which are superficially similar will have on reading efficiency —— that is, (2) "Given a sentence with a specified deep structure, what can be predicted about the relative efficiency with which it can be processed?"

The distinction made here between 'deep structure' (DS) and 'surface structure' (SS) is that expressed by transformational grammar.2/ Transormationalists maintain that (1) the language user is intuitively aware of differences between DS and SS; and further, that (2) sentences can only be understood through a reconstruction of the sentence's 'structural description', including its DS.3/ This study tests the suitability of the theoretical distinction between DS and SS in explaining the efficiency of information input and processing in reading.

Recently, there have been contradictory research findings about the language user's sensitivity to DS. Mehler ("What We Look at When we Read"; 1966) found in observing Ss' eye-fixations for ambiguous sentences: "The structure which differed only at the deep

2/ See, for example Jerrold J. Katz's The Philosophy of Language (1966), and Noam Chomsky's Current Issues in Linguistic Theory (1964) and "Topics in the Theory of Generative Grammar" (1966).

3/ See, for example, Geo. A. Miller and Chomsky's "Finitary Models of Language Users" (1963); and Chomsky's Aspects of the Theory of Syntax, (1965) phrase structure did not show such differences (in eye-fixation patterns as did the surface structure differences.)" However, Blumenthal ("Prompted Recall of Sentences"; 1966) and Blumenthal and Boakes ("Supplementary Report: Prompted Recall of Sentences"; 1967) indicate that "Recall differences correspond to the nature of the underlying grammatical relations."

In this study, the language-user's processing of linguistic material was measured by the Eye-Voice Span (EVS). In oral reading, the EVS is the distance, usually measured in words, that the 'eye' is ahead of the voice. The EVS was selected as the index for processing since recent work indicates that it is senstive to grammatical constraints within the sentence. Schlesinger ("Sentence Structure and the Reading Process"; 1966) states that the EVS "represents a unit of decoding." Further, Levin and Turner ("Sentence Structure and the Eye-Vice Span"; 1966) have found that subjects tend to read in phrase units, Thus, they have shown the EVS to be sensitive to phrase structure. Also, Levin and Kaplan ("The Eye-Voice Span for Active and Passive Sentences"; 1966) found the EVS to vary in accordance with intrasentence constraints — They found that the EVS was related to sentence voice (passive versus active).

METHOD

Subjects Thirty Cornell University freshmen and sophomores, 15 males and 15 females, served as subjects.

Experimental Sentences. Two kinds of passive sentence constructions were selected for this study. (See the Appendix for a list of the test sentences.) In the first type, the 'agent' or 'actor' was included while in the second, the agent was deleted. The sentences were paired so that the surface structure and lexical items were identical for both types of sentences, except that in one case the agent appeared, but in the other the agent was deleted and replaced by a non-agentive form. For example:

(A) His brother was beaten up by the gang.

(B) His brother was beaten up by the park.

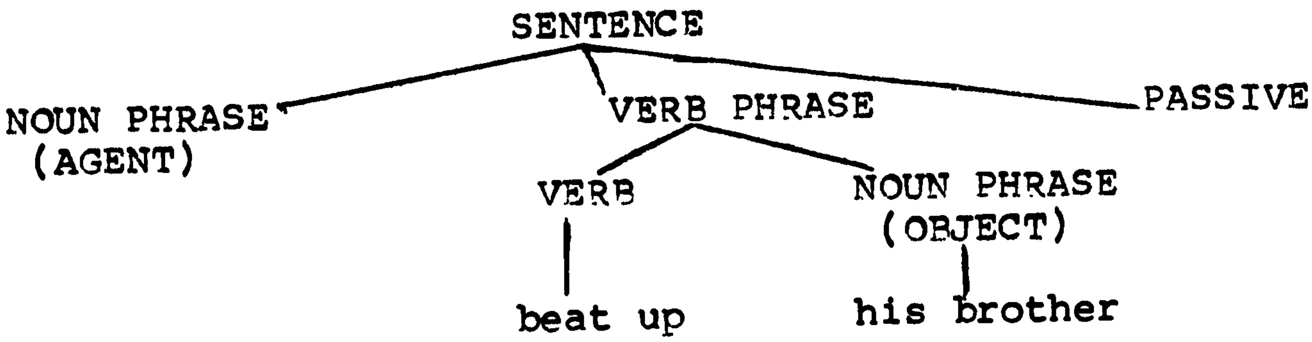
Both sentences A and B have the same surface structure.

[His brother] [was [beaten up]] [by [the gang / park]]

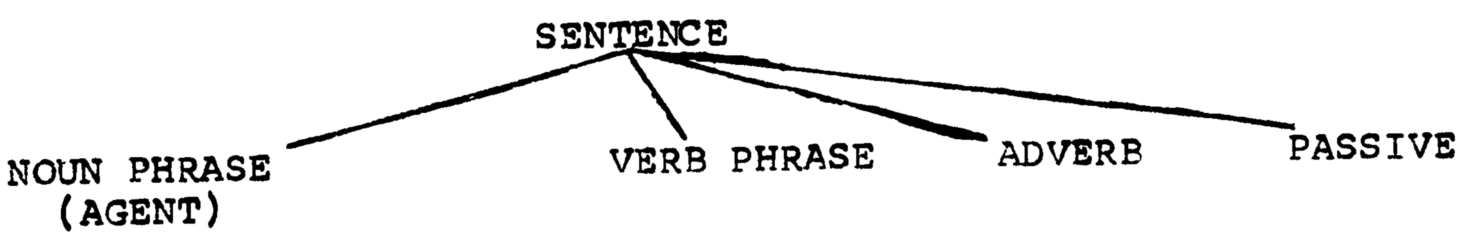
The actual test sentences are longer than these, having approximately 18 words, with about 8 words after the light-out position, to take into account Ss whose EVS might tend to be relatively large.

Both A and B contain the same lexical items except for one item. However, the substitution of "park" for "gang" in this case reflects a change in the DS of the sentence. In A, "gang" is the agent. In both A and B, "his brother" is the object of the verb "beat up", and someone or something else performed the action. Thus, A can be paraphrased as "The gang beat up his brother." Sentence B, on the other hand, cannot be paraphrased as "The park beat up his brother," since "park" is not the agent, but serves to indicate where the beating took place. An appropriate paraphrase would be: "His brother was beaten up near the park." In order for one to understand these two sentences, one must know that "his brother" is the underlying object and that something else is the agent or actor. In the first case, "the gang" is the agent. In the second case, one must know that "the park" is not the agent, and that the action of beating was performed by some agent not specified in that sentence.

It was hypothesized taht sentences of type A would be easier to process than sentences of type B since the DS in A was 'simpler' than in B. The DS for both A and B contain some of the same elements:



The DS in A is simpler for two reasons. First, the NOUN PHRASE — the AGENT NOUN PHRASE — is 'realized' in the surface structure as "the gang". Secondly, sentence type B requires an additional 'category slot' in the DS which is not required by type A. The deecription for the DS level for sentence type 3 would be like the following:



The ADVERB category slot must be realized in the DS because of SS phrase "by the park". Type B is more difficult to process because in addition to requiring a slot for the underlying subject (the AGENT) even though it is not realized at the surface level, it requires a slot for the adverbial phrase which was substituted.

Since both kinds of sentences have the same surface structure, differences in the way in which they are processed can be attributed to differences in the deep structure. Further, since both kinds of sentences were of the passive type, required the 'surface subject' to be understood as the 'underlying object' of some action performed by the (specified or unspecified) underlying level in type B sentences.

Procedure To test the hypothesis that the efficiency of reading processing is related to the deep structure of the sentence, the EVS was used as a measure of language processing in reading. There were eight pairs of test sentences. In each pair, both sentences had the same structural description and the same lexical items, except for one item. Both sentences of each pair were imbedded in an identical context of five other sentences. The sentences in these 'paragraphy' were unconnected so that S's reading of the test sentences would not be affected by intersentential cues.

The test sentences were imbedded within 'paragraphy' so that even though S knew the light would be turned out while he was reading the sentences in each 'paragraph', he did not know which sentence in the 'paragraph' would be treated as the 'critical sentence'. Also, even though every sentence position was treated as the critical positions, none of the test sentences occurred in the first position (sentence position 1), so that S was able to read at least one sentence through completely before reaching a test sentence. In addition, no test sentence occurred in the last position (sentence position 6), so that the experimental results would not be affected by S's ability to predict that the light would be extinguished during his reading of the final sentence in the 'paragraph'.

In order to minimize memory interference and response set, the sentences were divided into two groups, with each group containing the other members of each of the eight pairs:

Set #1 (given at Time 1)	Set #2 (given at Time 2)
A 1	B 1
B 2	A 2
A 3	B 3
B 4	A 4
A 5	B 5
B 6	A 6
A 7	B 7
B 8	A 8

Each subject read all sixteen test sentences, eight at each of two sessions, spaced one week apart. In the first session, S read one member of each pair and in the second session, the other member. In both sessions, Ss were presented with the eight paragraphs containing the critical sentences along with sixty-eight other paragraphs used as fillers between the test items. In both sessions identical filler paragraphs were used, and the order of presentation of all 76 (8 test and 68 filler) paragraphs was the same.

The paragraphs were typed single spaced on 9" x 5" cards so that Ss could read them as they would a regular 9" wide typed page. The cards were viewed one at a time by S. The viewing apparatus consisted of a wooden box approximately 24" x 18" x 12" with a one-way mirror through which S read the paragraphs. The cards were visible to S only when the light within the apparatus was on. The light was controlled by the experimenter who would extinguish the light and thus obscure the stimulus sentence when S reached the predetermined position for each paragraph.

S was instructed to begin reading each paragraph immediately when the light went on. He was told to read aloud at his normal rate, the rate at which he would read a story aloud to someone. He was told that within each paragraph, the light would be turned out so that he would no longer be able to see the text, and that he was to report all the words that he had seen but had not yet had a chance to read aloud. The light was turned out approximately an equal number of times in each of the six sentence positions, and within the sentences, the light was turned out at the beginning, middle, and end. The light-out position was varied to minimize Ss' response set. Of the eight pairs of critical sentences, two each appeared in sentence positions #2, #2, #4, and #5.

In four of the pairs of test sentences, the light was turned out immediately before the critical word reflecting a change in the underlying structure, and in the other four pairs, the light was turned out three words prior to the critical word. Two light-out positions were used because previous work by Levin[4/] has shown that there is a significant interaction between light-out

4/ Levin and Turner, "Sentence Structure and the Eye-Voice Span", 1966; and Levin and Kaplan, "The Eye-Voice Span for Active and Passive Sentences", 1966.

position, sentence structure, and reading processing as measured by the EVS. The light-out position immediately prior to the word reflecting a change in the DS (e.g., "gang" versus "park") was chosen since it was thought that the difference in reader response as measured by the EVS would be greatest at the point where the SS signalled the change in the DS. The light-out positbn three words prior to the critical word was used to determine how the reader's processing would differ at a point where the EVS would have picked up the SS cue, but prior to the point at which his oral reading had taken him to the critical word.

Scoring S had to recall a word perfectly in order for it to be included in his EVS score for a sentence. Thus, if a singular noun were changed to the plural form, or if a verb tense were changed, S's response was not counted. Also, no more than two words skipped could intervene between words actually recalled by S. This scoring procedure was used to minimize any distortion of the data resulting from the possibility that some S might skim to the end of a sentence without 'processing' the middle. Since these procedures required perfect recall of items in the SS and adherence to sequencing at the SS level, the scoring tended to favor surface structure over deep structure. For example, the ordering of structural elements would not necessarily be the same at the DS level as it is at the SS level; hence, scoring procedures which require S's recall to generally follow word order at the SS level are biased against DS. Also, requiring perfect recall of words, and consequently scoring as incorrect any changes in number (singular - plural) reflects a bias in favor of SS, for the 'number transformation' takes place at a low level within the structural description, proximate to the SS level. It is not something which is of high priority in the sentence's DS, since it is an obligatory mechanical process.

RESULTS

There were sixteen scores for each of the thirty Ss. The scores were averaged for each S so that there was a mean EVS score for type A sentences, and a mean EVS score for type B sentences. The scores were further broken down for the two light-out positions. In all cases what was being compared was S's EVS score for two sentences, one of type A, and one of type B, both with exactly the same surface structure, but with differing underlying structure. All the lexical items were identical except that in type A sentences the object of the preposition "by" was the agent (underlying subject), while in type B sentences the object of the proposition "by" was a non-agentive noun telling where or when, not by whom, the action was performed.

When the light was turned out immediately prior to the critical word, the mean EVS score (30 Ss, 4 pairs of sentences apiece) for the agent-included (type A) sentences was 5.81 words, while for the agent-deleted (type B) sentences it was 5.21 words. The difference between the means, 0.60 words, is significant at the .002 level (2-tailed test) and compares closely with the results of a previous pilot study.[5/]

5/ The pilot study was run under essentially the same conditions with 10 Cornell graduate students as Ss and with the same or similar stimulus sentences (6 pairs of sentences apiece). The difference between the means for type A and B sentences was 0.65 words.

When the light-out position was three words prior to the critical word, the mean EVS scores (30 Ss, 4 pairs of sentences apiece) were in the same direction as above (5.13 words for agent-included sentences versus 4.94 words for agent-deleted sentences). As expected, the difference between the means was greater at the point where the SS signalled the difference at the DS level.

DISCUSSION

The experiment indicates that reader efficiency is related to the deep structure of a sentence. Since the sentences tested had the same SS, but differing DS, the difference between the means supports the hypothesis that the efficiency of language processing in reading is not solely dependent upon SS, but is related to DS. The EVS measure was found to be sensitive to differences at the DS level.

In explaining the difference between the efficiency with which these two sentence-types were processed, one might argue that the individual lexical items appearing in the agent-included sentences were easier to process than the corresponding itmes in the agent-deleted sentences. This, however, seems unlikely, since both sets of words are quite common (See Appendix for a listing of these words). The explanation that objects of the prepositional phrases ("by the _______") of the sentences with the greater EVS appear more frequently in that context, and that such phrases are hence a more 'natural' part of the reader's linguistic repertory than the non-agentive phrases seems similarly unlikely. The grounds for saying that the phrase "by the gang" is more 'frequent' and consequently more easily processed than the phrase "by the park" must contend with the counter-argument that the frequency of occurrence of any particular phrase — or sentence, when compared to the indefinite number of possible phrases or sentences, is negligible.

'Naturalness', however, does appear to be a factor in the way these two structural types are processed. When the reader encounters a passive construction, some surface structure manifestation of the AGENT category is 'expected'. When the reader recognizes the verb form as marking the sentence to be a 'passive', the AGENT construction is somehow more 'predictable' or 'natural'. There is a strong correlation between judgments of 'naturalness' and size of EVS with respect to these agent-included versus agent-deleted sentence pairs.[6/]

One might explain both findings (greater efficiency of processing and higher rating on naturalness) in terms of 'predictability of occurrence'. However, the 'predictability of occurrence' would not refer to the frequency of particular lexical items or of particular combinations of lexical items. Rather, 'predictability of occurrence' would refer to the co-occurrence of particular items as they are defined at the DS level. Thus, in terms of these two

---

6/ In a related experiment, 76 Ss were presented with a list of 23 sets of sentences, including 8 agent-included sentences paired with 8 agent-deleted sentences. They were asked to "Decide which sounds better or more natural to you." The agent-included type was judged more natural twice as often as its agent-deleted counterpart.

structural types, what the passive construction makes 'predictable' is a surface structure realization of the DS AGENT category. A noun phrase such as "the gang" can be the NP immediately dominated by the S constituent when the verb "beat up" is given — that is, "the gang" and "beat up (+ OBJECT)" can co-occur as NP and VP of S. This is not the case with "the park" when the verb "beat up (+ OBJECT)" is given as the VP of S.

Predictability would seem to apply to the occurrence of items as they are functionally defined at the DS level, within the context of the selectional restrictions, or constraints, specified by the Ss realizations of the DS category slots. As the reader begins to pick up syntactic cues when he starts processing a sentence, what he has just encountered makes what is about to follow more predictable. (That is, if he has just 'read' the word "the", he would, on the basis of his previous linguistic experience, expect it to be followed by a noun such as "man", rather than by a verb such as "went".) This predictability is a function of the syntactic constraints holding between the lexical items in the sentence. For example, the syntactic cues that the reader could pick up as he begins processing "His brother was beaten up by the ..." might include the following:

1) "was beaten up" signals the passive construction

2) "his brother", since it is the NP preceding the verb in a passive constrction, is the OBJECT of the verb

3) some AGENT, as yet unspecified — i.e., not yet realized in the surface structure —"beat up his brother"

4) the syntactic features determining the verb "beat up" restrict the class of NPs which can co-occur as AGENT to those which have corresponding features — e.g., (+ ANIMATE)

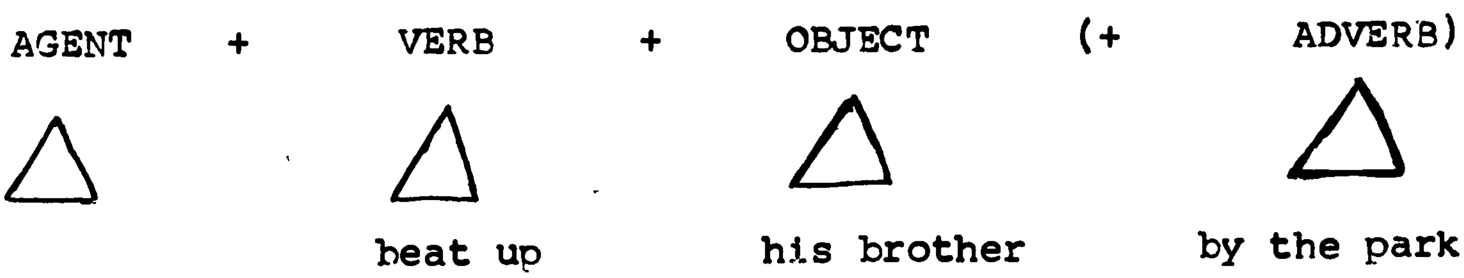
5) the presence of "by" here indicates that the NP functioning as AGENT is likely to follow. That is, the reader's previous experience with the preposition "by" occurring after the verb in a passive construction indicates that the AGENT is likely to be realized as the NP of "by"

Since all eight pairs of test sentences differed in the same respect (agent-inclusion versus agent-deletion), the experiment is tentatively interpreted to suggest that where categories specified in the DS are realized in the SS, reading processing is more efficient. That is, the reader is better able to process sentences in which there is a higher degree of "syntactic congruence" or "syntactic constraints" between the DS and SS levels. Of the two sentences discussed previously, A — "His brother was beaten up by the gang" — is more congruent to the DS

AGENT + VERB + OBJECT

since ther is a closer correspondence between slots realized in the DS and in the SS levels:

Sentence B requires substantially the same DS description:



except that the congruence between the two levels is not as great as with A, since the underlying subject (the agent) is not realized in the SS. This congruence can be spoken of in terms of the syntactic constraints at the two levels of structure.

Thus, the syntactic features of the AGENT category, at the DS level, are realized at the SS level in "the gang", which has the same syntactic features. Hence, the co-occurrence of these features at the two levels established the syntactic constraints or congruence between these two elements. This congruence can be partially explained by the fact that "gang" is marked for the feature (+ ANIMATE), while "park" is marked for the feature (- ANIMATE). Consequently, "park" cannot be a realization of the AGENT category in this case. In both sentences, the relation of VP to S (that is, of VERB + OBJECT to SENTENCE) is the same as "beat up his brother" is to S. Also, the relation of NP (that is, of AGENT) to S, <u>in sentence A</u>, is the same as "the gang" to S. But this relation doesn't hold in B. This is so because the syntactic features specified by the $\{NP\ ;\ S\}$ are realized by "the gang" in A, but not by "the park" in B.

The experiment is also very tentatively interpreted to suggest that efficiency in reading processing decreases with the amount of structure that must be realized in the DS to accommodate the SS. Sentence B above, like all the agent-deletion test sentences, requires another slot at the DS level in addition to those required by the agent-inclusion test sentences. The degree to which the amount of DS - SS congruence facilitates reading processing, or the extent to which extra category slots at the DS level hinder reader efficiency cannot be determined from this experiment. However, this study shows that deep structure does affect processing of linguistic material by readers. Possible explanations for this phenomenon, including the effects of additional structure at the DS level and the effects of DS -- SS congruence, will be tested in subsequent experiments.

LISTING AND EXPLANATION OF TEST SENTENCES

11. We knew that she was seriously injured by the / car (store) that was standing at the end of the alley.

17. His brother was beaten / up by the park (gang) that Fred had pointed out at the other end of town.

22. He had been / shot by the madman (restaurant) now being searched by detectives from the police station.

32. The speeding car was stopped by the / garden (police) where all the spectators were milling around.

35. If the lock was / fixed by the helper (time) you had promised us we would have paid you the full amount.

51. The rooming house that was / close by the University (closed by the University) had both male and female residents.

56. The door was opened by the / man (time) we had called and now we could once again use the garage.

62. We found out that his wife had been / attacked by the message (suspect) that the police brought over to his house.

The slash "/" marks the light-out position for each sentence. There were 4 instances in which the light was turned out immediately prior to the "critical word" (the word which serves to make the object of the by phrase the agent in the passive construction, or the word substituted for it): #11, #32, #51, and #56. In the other 4 instances, the light was turned out 3 words prior to the critical word.

The underlined word is the critical word which serves as agent for the passive construction.

The word in parentheses is the critical word which appeared in the sentence at the second test session in place of the preceding word. For example, in test session #1, each subject was presented the sentence:
"We knew that she was seriously injured by the car that was standing at the end of the alley." At test session #2, the sentence was presented in the following form:
"We knew that she was seriously injured by the store that was standing at the end of the alley."

The numbers preceding the sentences above refer to the sequence of the paragraphs in which each test sentence was imbedded. The order of presentation of the paragraphs (and, therefore, of the test sentences) was the same at both test sessions. The paragraphs in which each member of a pair of test sentences was imbedded was the same.

PRELIMINARY DRAFT
July 1967

Studies of Oral Reading:

XII. Effects of Instructions on the Ey.-Voice Span

Harry Levin and Julie A. Cohn[1/]

Cornell University

As mature readers, we are all at least vaguely aware that reading is not a unitary, static process; that, in fact, we read somehow differently -- although we may not realize exactly how so -- when we are reading different types of material and when reading for different purposes. It is this latter phenomenon which is dealt with in this study, namely the differential reading styles elicited by various mental sets. Instructions to read for certain purposes serve here to produce these different reading attitudes.

One measure currently used to describe the reading processes is the eye-voice span, or EVS. The eye-voice span is defined as the distance, or span of words, that the eye is ahead of the voice in oral reading. After its initial mention by J.O. Quantz in 1897, the phenomenon was studied to a considerable extent throughout the 1920's, when research in reading focused primarily on the eye-movements involved. It was generally believed then that techniques for reading improvement should involve the training of the oculo-motor system, such that the eye-movements of the poorer readers could be made to resemble those of the better readers.

1/This study is based on a Senior Honors Thesis submitted to the Psychology Department by the junior author.

Quantz, without resort to any of the elaborate eye-movement camera recording devices such as were subsequently developed, utilized the simple procedure of quickly slipping a card over the page while the subject was reading, and then recording the number of words spoken after the view was thus cut-off. Recently, renewed interest in the eye-voice span has concentrated not on the eye-movements involved, but rather on the central processes which govern it and upon the syntactic structure of the reading material affecting the eye-voice span as a chunking or decoding unit. The modern experimental apparatuses are modifications and refinements of Quantz's model, similarly cutting-off the subjects's view of the reading matter -- by shutting off the light which had allowed him to see it through a one-way mirror, or turning off the projector which had projected a slide of the passage on a screen -- and asking him to repeat whatever additional words he remembers having seen.

As a flexible, elastic unit, the eye-voice span has been demonstrated to increase and decrease in size due to the effects of certain stimulus and subject variables. Dealing first with the stimulus determinants -- characteristics of the reading material itself -- some contradictory evidence has been reported.

Quantz (1897) and Fairbanks (1937) reported that the position of the cut-off within the line of printed matter exerted a definite and consistent influence; specifically, that the EVS, measured at the beginning of a line is longest, that at the end of the line is shortest, and that central in the line is of intermediate length. Presumably the span is narrowest at the end of the line because the eyes pause to wait for the voice to catch up. Accordingly, Judd (1918) found that in addition to the one pause per word generally occurring in oral reading, there is also this long pause at the end of the line.

Buswell (1920; 1936) on the other hand, did not corroborate this finding; his studies demonstrated no effect of position within a line. Rather, he maintains that the differential EVS's reflect position within the sentence. Fairbanks (1937) and Vernon (1931) feel that both variations can occur. Buswell's good readers and all of Fairbank's subjects exhibited their longest EVS's at the beginning and the shortest EVS's at the end of the sentence. To Buswell (1920):

> The fact that the EVS varies with the position in the sentence is of considerable significance. If the span varied only with the position within the line, as Quantz's study indicated, the determining factors would be entirely mechanical and would be governed by the printed form of the selection. The control of the span, in that case, would be a matter of the mechanics of book construction and would be independent of any teaching factor. But if the span varies with position in the sentence, it is evident that the content of meaning is recognized, and that the EVS is determined by thought units rather than by printed line units. Position in the line may be a minor factor...but the differences due to position in the sentence are much greater.

This is rather a sophisticated statement for 1920. In his concentration on the meaningful sentence as influential on the EVS, Buswell seems to anticipate the recent emphasis on its grammatical determinants. He suggests that the EVS "allows the mind to grasp and interprete" a unit of meaning before the voice must express it, but does not offer evidence, however, to back his hypothesis that the chunks necessarily be meaningful phrase units, as Anderson (1937) also suggests. Therefore, a trend of exploration is currently directed toward the grammatical phrase structure of the sentence itself. Much earlier, in 1897, Catell had deduced from his findings -- that subjects could just as readily recognize tachistoscopically presented words, phrases and short sentences as they could single letters -- that the units of perception could be words, phrases or even sentences. Likewise, Tinker (1958) expressed the opinion that reading units, rather than in terms of spelling or syllabizing, are in terms of word groupings which form perceptual "wholes." Recently Schlesinger (1965), defining the EVS to be a "unit of decoding," predicted that these could be understood with reference to the syntactical structure of the stimulus materials. Accordingly, Levin and Turner (1966) found their subjects to read to phrase boundaries significantly more often than to non-phrase boundaries, suggesting again that the decoding unit be of phrases. At least with their adult subjects, Levin and Turner also revealed that the grammatical sentence-voice produced an effect on the EVS, and this effect -- that passive target sentences yielded significantly longer EVS's -- was replicated by Levin and Kaplan (1966). They accounted for this tendency on the basis of the greater constraint of the passive form. This inference follows from the fact that Lawson (1961) and Morton (1964) had found the length of the EVS to be directly related to the contextual constrain, or information content, of the reading matter, and that Clark (1965, 1966) and Roberts (1966) have shown the passive voice to be more predictable, more constrained, than the active voice.

Still other experimenters have considered additional elements of the nature of the reading material, focusing on the content, rather than structural properties. It has been reported consistently, for example, that the more difficult the reading matter is, the shorter the resultant EVS (Buswell, 1920; Anderson, 1937; Fairbanks, 1937; Huey, 1922; Tinker, 1958). On the other hand, Ballantine (1951) and Morse (1951), dealing not specifically with the EVS, but comparing the eye-movements of children reading selections appropriate to their grade and geared to grades two years above and below their own, found the eye-movements to be quite consistent, not changing in any predictable way with the change of two years in difficulty. Tinker dismisses such findings as resultant from an "unfortunate lack of flexibility" in these subjects, hampering their ability to adjust their pace and procedures in accordance with the difficulty of the passage. This seems to be a weak and rather defensive argument however. If indeed the lack of change in these subjects' eye-movements was due to their lack of flexibility, then their lack of flexibility itself is a valid phenomenon, necessitating acknowledgement and further explanation.

Ledbetter (1947) found differences in the eye-movements in reading of material from various subject areas, although the selection were controlled for length, vocabulary difficulty and sentence structure. That the greatest sources of difficulty were in the reading of poetry and mathematics "seems to point to the conclusion that meanings or concepts present more difficulty to the average student than vocabulary, sentence length or sentence structure, and that certain subject matter fields have inherent difficulties." Yet Tinker accounts for such results with resort to probable differences in the familiarity, difficulty, and purpose of the reading:

> Different types of material, read with the same instructions (set) do not automatically elicit different patterns of eye-movements when the passages are equally difficult. These findings are important. The above findings do not mean, however, that reading for different purposes or reading materials with wide variations in difficulty would not produce variation in oculo-motor behavior.

Assuming content differences to be explained in terms of the degree of difficulty, we may perhaps further explain the effect of the difficulty of a selection in terms of its lesser predictability or constraint; a difficult passage is one which is not highly predictable, due either to its unfamiliar vocabulary or to the obscurity of the concepts dealt with.

Another line of investigation has been directed toward subject determinants of the eye-voice span. It has been consistently demonstrated that the EVS tends to increase with age up to adulthood (Buswell, 1920; Tinker, 1958; Levin and Turner, 1966). Also fairly well established is the fact that good readers generally have a longer EVS than poor ones (Buswell, 1920). Quantz (1897) and Levin and Turner (1966) reported that faster readers have a longer EVS than slow ones: this probably because, since most of the time involved in reading is in fixational pauses, faster readers, with fewer pauses, must take in more information at a single pause. As Morton (1964a) has used reading rate as his criterion for determining good versus poor readers, it seems that the two are inextricably related.

Fairbanks (1937) holds that the difference between good and poor readers are of a central nature, reflecting comprehension as it is modified by the reading attitude, and variations of mental activity as determined by the difficulty the reading matter presents to the subject. After an analysis of their EVS's and errors, revealing that faulty eye-movements cannot have caused the errors, he concluded that the errors must be central in origin and that other peripheral manifestations of reading likewise must have similar dependence upon central processes. This is consistent with Anderson and Tinker who express a similar opinion, that eye-movements in reading vary with central, rather than ocul-motor processes. Says Tinker (1958):

> It is now well established that oculomotor patterns are exceedingly flexible and quickly reflect any variation in the central processes of perception, judgement, comprehension, etc. In other words, it appears that eye-movement patterns merely reflect ease or difficulty of reading performance and degree of comprehension, rather than cause good or poor reading. Versatility in adjustment of reading habits to variation in purpose and materials is one "hallmark" of maturity in reading.

The notion of reading attitude or set (as determined by purpose, for example) has thus been acknowledged by several investigators as a determinant of eye-movements. Most experimental evidence in this area does not deal directly with the eye-voice span, but rather with the recording of various eye-movement measures, primarily in silent reading. However, since some of these measures -- forward shift of the eyes, size of fixation, etc. -- appear to be involved in the eye-voice span, the results of these experiments seem to be highly relevant here.

C. T. Gray (1917) contended that although the length of pauses did not vary in any large degree, the number of pauses and regressive movements did change appreciably with the different types of reading his subjects were asked to do, increasing when the subjects were required to answer questions about the material. Thus there was a demonstrated tendency to read in smaller units when they were to be questionned than when asked to reproduce only a general thought. His results, typical of all subjects, indicate clearly that the reader does differentiate between types of reading and evidently approaches different reading problems with different mental sets. An even clearer statement of similar results is reported by Judd and Buswell (1922) who, finding a general tendency to increase the number and decrease the duration of fixational pauses, in addition to the expected increase in regressions, with detailed reading, infer again that a mental set for close reading is answered by a procedure utilizing smaller reading units, while larger units are employed for more superficial reading. Even greater differences were yielded when the subject was instructed to paraphrase the material; this highly detailed reading set required even smaller units of analysis. Vernon (1931) claims that reading is most irregular when disturbed by conflicting interests or emotional tensions; pressure to read quickly or to learn all the details at a single reading may produce irregularity and confusion in the reader.

Anderson (1937) compared eye-movements utilized by good versus poor readers in response to different instructionally-determined attitude sets for their silent reading. With regard to the mean duration of fixations, the shortest pauses of the good readers resulted from instructions to read for the general idea, while poor readers demonstrated their shortest pauses under the normal condition ("to obtain a moderate knowledge of the text"). On all measures of eye-movements, the instruction to read for the general idea yielded the largest differences between good and poor readers. The size of fixations decreased to almost a common figure under the detailed reading condition. The largest mean forward shifts were found under the general idea condition, followed by the moderate knowledge condition, and smallest in the detailed reading condition, although these measures were consistently greater for good readers than for poor ones. Rate of reading followed a parallel pattern. Although these measures did not deal directly with the eye-voice span; they would seem relevant to it as movements which comprise the scanning of the eyes ahead of the voice, and therefore seem to suggest that similar results would be found with the EVS measure.

The differential results for these different reading conditions illustrate the flexibility of the eye-movement patterns in accordance with the varied mental processes and the fact that the good readers demonstrated a wider variation in habits, especially in the general idea condition, and least for the detailed reading condition seems to imply that good readers are more adaptable than poor ones, and that the norm of the poor reader lies closer to the careful word-by-word reading procedure characteristic of both detailed and immature reading. Laycock (1955) studying the oculo-motor patterns of this flexibility phenomenon in college readers, found that when told to read faster, both groups increased their rate, but the flexible group increased it more. The flexible group likewise increased their average fixation span by 31%, the inflexible group by 17%, and although both groups decreased their duration of fixation, the flexible group saved twice as much time at each fixation.

In sum, evidence to date has concentrated on two main points of focus in investigation of the eye-voice span. Along the lines suggested earlier by Busewell, one trend of exploration has centered on the structural properties of the sentence as these affect the length of the EVS. The eye-voice span has come to be considered as a unit of decoding -- in terms of meaningful phrase units. Contextual constraint, or predictability, has been recognized as an important determinant of the EVS, and one which seems to explain such findings as the increased EVS elicited by difficult material, and the differential EVS with active and passive sentences. Another line of approach has dealt with the effects of subject variables. In addition to concrete factors like age and reading rate of the subjects, investigation have also attempted to deal with the more illusive central processes -- comprehension, perception, attitude set -- as these affect the EVS. It is felt that the flexible oculo-motor patterns result from these central determinants rather than vice versa. Different reading styles are utilized by readers with different reading attitudes -- as determined by the purpose of the reading -- yet the degree of difference depends upon the maturity and flexibility of the reader. Findings on eye-movements, especially in silent reading, suggest that closer, more detailed reading, utilizing smaller chunking units, would therefore exhibit shorter EVS's than normal; whereas broader reading, resembling skimming, involves longer forward shifts and, by implication, longer eye-voice spans.

In the present experiment, the effects of different instruction-imposed attitudinal sets will be investigated as they relate directly to the eye-voice span in oral reading. Utilizing elementary and high school students, a developmental approach may be considered. It is hypothesized that: (1) mental set for detailed reading will yield a decrease in EVS resulting from the smaller-than-normal chunking units, whereas reading for a general idea should yield a longer EVS than normal; (2) the EVS should increase with the grade level of the subjects; and (3) the older subjects would be expected to show a greater range of variation in EVS with the different instructions than the younger readers, due to the greater flexibility of mature reading processes.

'ETHOD

Subjects. The subjects were sixty students from the public schools in Ithaca, divided as follows:

fifteen subjects (seven boys and eight girls) from the second grade at West Hill School;

fifteen subjects (six boys and nine girls) from the fourth grade in West Hill School;

fifteen subjects (nine boys and six girls) from the ninth grade at Boynton Junior High School;

fifteen subjects (five boys and ten girls) from the eleventh grade at Ithaca High School.

The selection procedures varied. At West Hill School, the children were chosen at random from class lists, with the principal eliminating only those who did not have a minimal second grade reading ability, since these children would be unable to read the stimulus materials. The ninth graders had been chosen by the Guidance Department at Boynton Jr. High, as a sample which they felt would be a typical cross-section of their ninth grade reading population -- some were remedial, some advanced, most classified as average readers. At Ithaca High School, the subjects were solicited from two Introduction to Psychology classes. No knowledge of the nature of the experiment was provided, and the selection was made only on the basis of their scheduled free periods.

The introduction of these several selection procedures, of course presents the problem of possible bias in the sample, due to the various characteristics of the subjects. Even in the elementary school, the otherwise random sample may have been affected by the additions and deletions which the principal made. It is even more likely that the sample selected from Boynton was biased; the guidance counselors probably had certain reasons for the selection of these particular students. By soliciting volunteers in the high school, the possibility for a self-selection bias was introduced; and the very fact that the classes asked to participate were confined to psychology classes -- an elective which only certain students take, perhaps based on academic criteria -- may also have rendered the sample non-random.

Although, at least for the first three grade levels, the subjects had been classified by the school officials as good or poor readers, no such classification is utilized in this study. The grade level, taken as a whole, is the only factor considered.

Stimulus Materials. The reading material consisted of twenty-two short selections on each of two grade levels (reproduced in Appendix A in the same from and size of type as used in the experiment), which were typed on individual 5" by 7" cards for insertion in the reading apparatus. The passages were taken either verbatim, or with slight modification, from level A second grade and tenth grade readers as cited. The two elementary grades read second grade level material; the upper grade students read tenth grade matter.

Each selection consisted of a passage of six sentences of meaningful connected discourse. The critical sentence was either the third, fourth, or fifth sentence of the passage, always beginning on a new line, so as to allow a sufficient number of words beyond the light-off position on the single line. Fox example:

The boys/ followed their mother into the store.

(The slash indicates the light out point.) These critical sentences were of consistent grammatical structure: beginning with a noun phrase of either one, two, or three words, followed by a three-word verb phrase, and a three-word prepositional phrase, and, in the higher level selections, still another three-word phrase. The light-off position was always between the subject and verb phrases, so placed in the beginning of the sentence in order to allow either six words (for the younger readers) or nine words (for the older subjects) beyond this point, so that the subjects could exercise their full eye-voice span on this single line. The confounding effects of different within-line and within-sentence positions were thereby avoided.

Apparatus. The experimental apparatus consisted of a wooden box, measuring 24" X 18" X 12", with a slanted top fitted with a one-way mirror. The passages on cards were inserted behind the mirror in such a way that the subject could only see the material if the light inside the box was on. This light was operated by a microswitch; when the experimenter depressed the switch, the light turned on; when he released the switch, the light shut off. A timer was operated by the same switch, thus beginning when the light came on, and stopping immediately as the light was turned off. In this manner, the amount of time of reading, until the light-out point, was recorded.

Procedure. The subject entered the darkened experimental room and the apparatus and procedure were explained and demonstrated. The subject was told that in the middle of the passage, the light will be turned off and he will be unable to see any more of the selection, but he is requested to report whatever words he remembers of the rest of the sentence. The twenty-two selections were then presented, one at a time, in a constant order. For each subject, the first passage was presented under the "normal" reading condition, and was used mainly for demonstration or practice purposes; the results obtained from this first sentence are not included in later analyses. The other twenty-one selections were each preceded by one of three sets of instructions, geared to elicit either a normal reading set, an especially careful reading set, or a set similar to "skimming." Seven of each instructional condition comprise the set. The instructions were as follows (Parentheses denote changes for the older grade subjects):

(1) Now, I'd like you to read me this story (passage) just the way you normally would if you were reading out loud for the teacher (to a friend).

(2) Now, this time, I want you to read me the story (passage) very carefully. Pay close attention to all the details because I'm going to ask you questions about it afterwards.

(3) This time, you don't have to pay such close attention to each (individual) little detail. Instead, I'd like you to just read right through for the general idea of the story.

For all those selections presented under the second condition, one question, on the details of the action, was posed.

The three conditions were arranged in a single random order, but the series began at a different point in the order for each subject, so that the same condition did not fall consistently with the same critical sentence. For example, the order began as follows: 1, 2, 1, 3, 3, 2... etc. Thus, the first subject began his first paragraph under condition one, and continued on through the series; the second subject began his series with the first selection under condition two, and finished the set with condition one, and so on. No order effect would thus be expected. In this manner each subject served as his own control, reading seven selections in each of the three experimental conditions.

After completing the series of passages, the subject returned to his classroom and sent in the next scheduled student.

Scoring. The number of consecutive words which the subject reported having seen (i.e. the amount of the sentence he completed beyond the light-out point) was recorded for each critical sentence. This measure was operationally defined as the eye-voice span. Note was made of the condition under which each of the selections was read, and of the sentence position of the critical sentence in the passage. The time until the light-out point was likewise recorded, later to be converted to a rate score in terms of number of words read per second.

RESULTS

The basic hypothesis testing the responsiveness of the EVS to the three types of instructions was tested by an analysis of variance with the classifications, grade and condition. The means and summary ANOVA are given in Table 1. Both main effects, grade level and instructions, are significant; the interaction is not. Instructions

Insert Table 1 near here

to read carefully resulted in the shortest EVS (3.69 words), normal reading next and skimming in the longest.

The grade means are a bit more complicated. Second graders have the shortest EVS and eleventh graders the longest. The fourth and ninth grades reverse the expected order. It should be recalled that the second and fourth graders read second grade materials and the two older grades read tenth grade prose. This suggests that the lower than expected EVS score for the ninth graders was due to the relative difficulty of their reading materials.

There were different numbers of boys and girls in the various grades. The possibily suggested itself that the grade level differences in EVS were due to the larger proportion of girls in the upper grades. In fact, the overall EVS for girls is larger than for boy ($t=1.88$, $p < .10$). Consequently a second ANOVA was calculated which included "sex" as a classification. In order to equalize the cell frequencies the ninth and eleventh grades were combined. Again, the main effects of grade and instruction conditions are significant; sex means differ at the 10% level. No interactions are noteworthy. We conclude, then, that the findings are not attributable to the different representations of boys and girls at the various grade levels.

Reading Rate and EVS. Because of the previous finding that EVS and rate are positively correlated, (Levin and Turner, 1966) it was expected that such relationships would hold in the present experiment. Thus rank order correlations between EVS and rate were performed for each instructional condition among the subjects of the second, fourth and ninth grade. (Time scores for the eleventh grade were unavailable due to a mechanical failure of the timing apparatus.) Table 2 shows the

Insert Table 2

rho values for each of these correlations, all of which are significant beyond the .05 level of confidence. At least within the condition and grade, then, the eye-voice spans and reading rates of the subjects are highly correlated -- the faster readers exhibiting the longest EVS's in each condition.

It would be expected, on the basis of this consistent co-variation, that instructional condition would produce in the reading rate a comparable effect with that produced in the eye-voice span: i.e. that the careful condition would decrease the rate of reading, and the general idea condition increase the rate beyond the normal rate. Surprisingly, a comparison of the average reading rates yielded for each condition by grade (Table 3) does not show such effects. Although

Insert Table 3

there seems to be an overall increase in rate from the second to fourth grades, the rates are fairly constant despite instructional differences.

DISCUSSION

The main concern in this study, that of instruction-induced motivational set upon the eye-voice span, appeared as expected from the similar results of eye-movements studies of silent reading by Gray, Judd and Buswell, and Anderson. The shorter EVS obtained from the careful reading condition seems to indicate that the subject uses smaller chunking units, focusing more closely and exclusively on each word. In terms of the eye-movement measures utilized by these other experimenters, the increased number of fixations, the small forward shifts, the increased number of regressions may all be seen as involved in the decreased EVS length. As Anderson describes the process:

> In order to satisfy a more severe requirement of comprehension, a more deliberate and careful scrutiny of the reading material was necessary; this was immediately reflected in the eye-movements.... The irregularities of eye-movements found in this reading situation are not comparable to the erratic and inconsistent eye-movement behavior of the immature reader. The peripheral signs in this situation are rather a characteristic pattern which reflects an increasing dependance upon the reading material and a peculiar mode of attack necessary to satisfy the requirement of comprehension.

One other factor which may have entered as a contributing cause of the shortened EVS under this instructional set is that of increased tension or anxiety. Vernon maintains that any such pressure (as here may have occurred due to the interrogations to follow the reading of the selections) may result in more erratic and confused reading processes, which would decrease efficiency. However, since the questions were handled very casually, without scoring, and almost jokingly if the subject appeared nervous, it is suspected that any such effect would have been minimal.

In the condition where the subject's purpose was to read for the general idea, the opposite tendencies were exhibited, as was predicted from the earlier reports. The eye seemed to scan well ahead of the voice, producing a general forward tendency. Subjects were noticed to exhibit less regression and repetition and to be less bogged down with individual troublesome words. Without any anxiety, subjects in this condition could perhaps read more freely. Anderson understands the differential reading habits to be a function of the different number of kinds of cues necessary to satisfy the requirements of a particular reading situation. In this case,

> since the objective cues necessary to reproduce the general idea are fewer than those required in the other reading situations, the effective reader will show considerable independence of the printed page. The subsequent reduction in the number of fixations indicates that the comprehending activities are enhanced and contribute more to the successful realization of the task than would a careful and deliberate exposure to the reading material.

As there was no mention of a quickened speed in this condition, the instructional effects seem to have been confined to the types of eye-movements that occurred rather than to producing the overall accelerated pace which usually goes hand-in-hand with an extended eye-voice span.

Looking at the mean scores for the normal condition, and comparing these with the mean EVS's obtained under careful and general idea reading sets, it becomes apparent that in the early grades, the normal score is very close to the careful one, whereas, on the high school level, it is the general idea condition which produces results more similar to normal. This seems to suggest the fact that normal reading for the beginning readers entails a process very similar to that which is customarily utilized for detailed reading; that is, the reading is word-by-word, with close attention to the individual word, and without much forward scanning. Contrarily, older readers ordinarily read for the general idea; thus, this attitude is close to normal for them. (Of course, their reading is generally silent rather than oral, but the similarity of attitude, if not an identity of process, between the general idea condition and the usual skimming condition of silent reading, is obvious.) Anderson tends to support this explanation, citing the "inability of poor readers to adopt any other than their everyday reading attitude. Good readers, he continues, "on the other hand, showed their most regular eye-movement patterns in reading for the general idea." Effective reading for the general idea, according to Anderson, is a highly refined skill which presupposes a mastery of certain basic fundamental skills, like word recognition, knowledge of vacabulary and sentence meaning. Unless these elementary skills have been mastered, and can function with a minimum of effort, the reader will fail to make the necessary psychological transition which reading for the general idea requires. In Anderson's good readers, and in the more mature readers of the current study, these skills are well developed and therefore these subjects succeeded in adjusting their reading attitude to comply with the more subtle interpretation demanded of the material. Immature readers, highly engrossed in the elemental concerns, tend to read all the material in almost the same manner.

In sum, there seems to be a notable flexibility of eye-voice span, reflecting changes in the central processes of apprehension and comprehension. Eye-movements appear to be dictated by conscious processes, activated subsequent to the instruction-induced set for the purpose of reading. By implication, and in view of the apparent direction of the relationship between eye-movements involved in the eye-voice span and the conscious attidudinal determinants, it seems that earlier attempts to improve reading through eye-movement training were using the wrong approach.

The results dealing with the rate of reading do not lend themselves to such clear-cut and consistent interpretation. The fact that rates were found to correlate significantly with the EVS of subjects within grade and condition classifications may be construed simply as another replication of the repeated finding that EVS varies directly with the ability of the readers, since Morton utilizes rate as an index of ability. Here, then, the fastest readers in each classification had the longest EVS's.

However, it was not expected, in view of the fact that EVS and rate normally go hand-in-hand, that no significant changes would occur among reading rates for the different instructional conditions. This would seem to indicate that while the instructions did not effect the rate of the verbal enunciation of the passage, the effect was confined to the movement of the eyes, serving to push these further ahead of the spoken word. The rate of oral reading is limited by the rate of articulation.

Any such conclusion cannot, however, be considered definitive due to the several methodological difficulties presented by these rate measures. In the first place, a mechanical failure in the timing apparatus was responsible for the omission of all rate scores of the eleventh grade subjects. Even excluding this group of scores, the reading rates are not very accurate due to inconsistent behavior of the subjects. Often the children would go back and repeat parts or whole sentences, or would make comments to the experimenter, all while the timer was running, thus producing unreal inflations of the rate scores. Other subjects sometimes omitted entire sentences completely, thereby unrealistically decreasing their computed rates.

SUMMARY

In the present experiment, the effects of different instructionally-induced sets upon the eye-voice spans of readers at various elementary and high school levels was investigated. It was hypothesized that:

(1) The set for detailed reading would result in a decrease in EVS length, caused by the smaller-than-normal chunking units utilized in this type of reading; whereas reading for a general idea (skimming) would yield a longer EVS than normal.

(2) The EVS would increase in length with the grade level of the subjects.

Primarily, then, this experiment has demonstrated that different instructions produce significant changes in the reading behavior of school children. The EVS was shown to vary with instructions in the same ways as eye movements, as was indicated in a number of earlier studies.

REFERENCES

Aborn, M., Rubenstein, H. & Sterling, T. Sources of contextual constraint upon words in sentences. J. Exp. Psychol., 1959, 57, 171-180.

Anderson, I. H. Eve-movements of good and poor readers. Psychol. Monog., 1937, 48, 1-35.

Ballantine, F. A. Age changes in measures of eye-movements in silent reading. In Studies in the Psychology of Reading. U. of Michigan Monographs in Education. No 4. Ann Arbor: U. of Mich. Press, 1951, 65 - 111.

Buswell, G. T. An experimental study of eye-voice span in reading. Suppl. Educ. Monog., 1920, no. 17.

Cattell, J. McK. Mind, 1889.

Clark, H. H. Some structural properties of simple active and passive sentences. J. Verb Learn. Verb. Behavior., 1965, 4, 365 -370.

Fairbanks, G. The relation between eye-movements and voice in oral reading of good and poor readers. Psychol. Monog., 1937, 48, 78 - 107.

Gray, C. T. Types of reading ability as exhibited through tests and laboratory measurements. Suppl. Educ. Monog., 1917, 1, no. 5.

Huey, E. B. On the psychology and physiology of reading. Am.Jrnl. Psychol., 1901, XII, 292 - 313.

Huey, E. B. The psychology and pedagogy of reading. New York: MacMillan, 1922.

Judd, C. H. Reading: its nature and development. Suppl. Educ. Monog., 1918, no. 10.

Judd, C. H. and Buswell G. T. Silent Reading, a study of the various types. Suppl. Educ. Monog., 1922, no. 23.

Lawson, E. A note on the influence of different orders of approximation to the English language upon eye-voice span. Quart. J. Exp. Psychol., 1961, 13, 53 - 55.

Laycock, F. Significant characteristics of college students with varying flexibility in reading rate. I - Eye-movements in reading prose. J. Exp. Educ., 1955, 23, 311 - 330.

Ledbetter, F. G. Reading reactions for the eye-movements of eleventh graders. J. Educ. Res., 1947, 41, 102-115.

Levin, H. and Kaplan, E. Studies of oral reading: the eye-voice span for active and passive sentences., 1966.

Levin, H. and Turner, E.A. Studies in oral reading: sentence structure and the eye-voice span., 1966.

Morse, W. S. A comparison of the eye-movements of average fifth and seventh grade pupils' reading materials of corresponding difficulty. In Studies in the Psychology of Reading. U. of Mich. Monog. in Educ. No. 4. Ann Arbor: U. of Mich. Press, 1951, 1 - 64.

Morton, J. The effects of context upon speed of reading, eye-movements and eye-voice span. Quart. J. Exp. Psychol., 1964a, 16, 340-354.

Quantz, J. O. Problems in the psychology of reading. Psychol. Monog., 1897.

Roberts, K. The interaction of normative associations and grammatical factors in sentence retention. Paper read at MPA, Chicago, 1966.

Schlesinger, I. M. Sentence Structure and the Reading Process., In press, 1966.

Tinker, M. A. Recent studies of eye movements in reading. Psychol. Bull., 1958, 55, 215-231.

Vernon, M. D. The Experimental Study of Reading. Cambridge: University Press. 1931

Table 1.

Mean EVS (words)and ANOVA by Grade and Instructional Condition

	Grade			
Instructions	2	4	9	11
Normal	2.95	4.32	3.97	4.64
Careful	2.92	4.04	3.73	4.09
Skimming	3.09	4.59	4.07	4.82

	SS	df	MS	F	P
Grades	62.24	3	20.75	5.83	<.01
Error(b)	199.32	56	3.56		
Instructions	6.14	2	3.07	10.59	<.01
Gr. x Instr.	1.69	6	.28		
Error(w)	32.84	112	.29		
Total		179			

Table 2.

Rank Order Correlations of EVS with Rate

Condition	2nd grade	4th grade	9th grade
Normal	.86	.52	.68
Careful	.84	.75	.80
Skimming	.67	.52	.81

Table 3.

Mean Rates (Words / Sec.) by Grade and Instructions

	*Grade		
Condition	2	4	9
Normal	1.58	2.31	2.59
Careful	1.59	2.27	2.53
Skimming	1.52	2.19	2.69

*Time measure were not available for the 11th grade.

PRELIMINARY DRAFT
September 1967

Studies of Oral Reading

XIII. Filled inter-word spaces and the Eye-Voice Span (EVS)[1/]

Harry Levin and Dalton Jones

Cornell University

When reading aloud, the eyes are usually sampling the text beyond the point of the voice. The distance between the eyes and the voice has been called the eye-voice span (EVS) and the characteristics of this phenomenon have been investigated since the turn of the century. The EVS seems to represent a decoding unit, usually a phrase (Schlesinger 1966; Levin & Turner, 1966). This means that the EVS bellows in and out depending on the phrase structure of the sentence: smaller as the voice is near the end of phrase; longer when the voice is at the beginning of a phrase. Further, the EVS is sensitive to the grammatical constraints within a sentence. For example, the span is shorter for word lists than for sentences and longer for passive than for active sentences (Levin & Kaplan, 1966).

As readers become more skilled, the size of the span increases (Levin & Turner, 1966; Levin & Cohn, 1967). This finding implies that reading involves a strategy of sampling the text so that the EVS reflects the process of scanning the text to pick up the cues that determine the performance of the voice as well as to gather information about what is coming next. By its nature, oral reading involves saying every word

1/This research was supported by funds from the U.S. Office of Education.

so that at one level, the sampling can be said to be 100%. Nevertheless, errors in oral reading imply that sampling is less than total and predications by the reader are made on the basis of information that is picked up. Sampling of greater or lesser density is more obviously the case during silent reading where it is not necessary to translate every word to sound.

We have attempted from several directions, to discover the determinants of text sampling. For example, in a recent study (Hochberg, Levin and Frail) we predicated that spaces between words function as cues for the peripheral search guidance of the eyes. Although the results were at first equivocal, several additional lines of evidence are converging to support this predication.

The present study brings together, the EVS and the filled space studies. If the EVS involves a systematic scanning of text and if such scanning depends on the inter-word spaces, we expect that the size of the EVS should be reduced by filling the spaces.

METHOD

Subjects

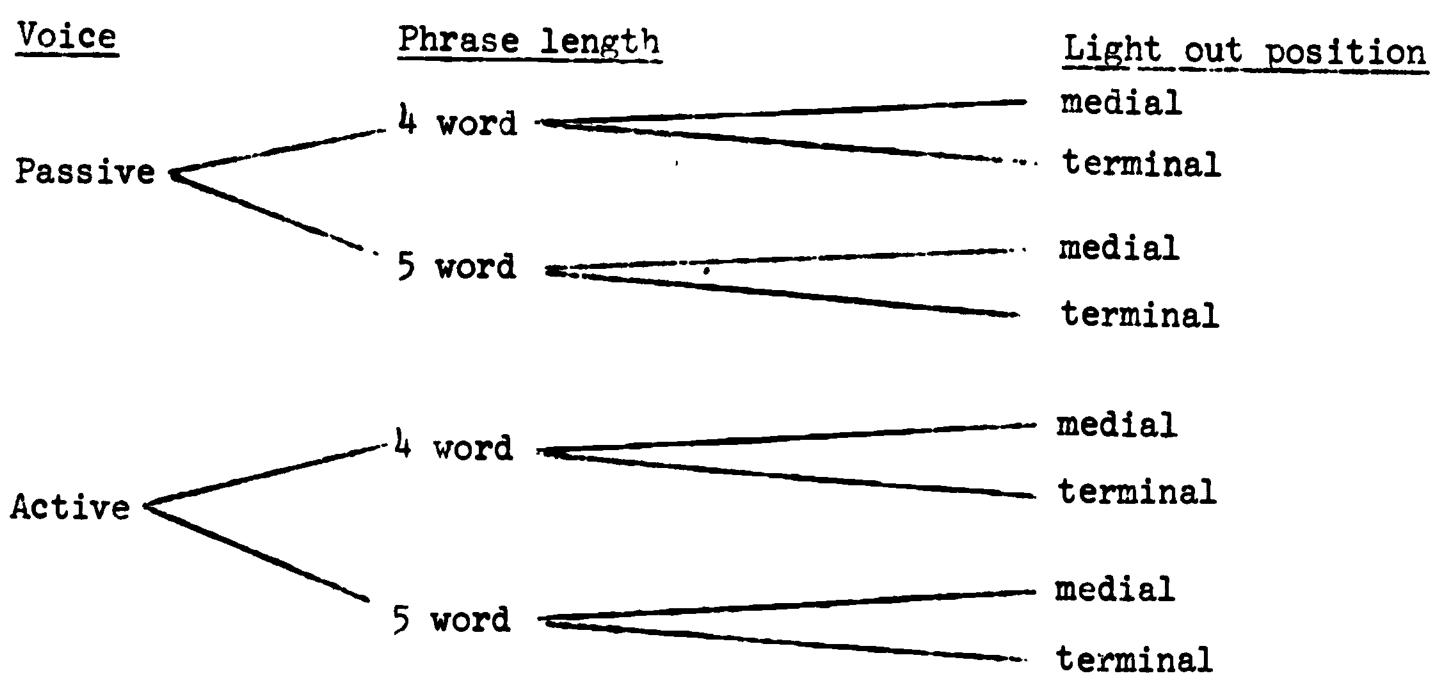
The subjects were 18 undergraduate students from a general psychology course at Cornell University. The Ss were told that the experiment was an attempt to find out what people were doing when they read, not to test how well they read. Ss were instructed that they would be presented with short paragraphs to read out loud and that at some point the light would go out. Their task was to report as much as they could beyond the point where the light went out.

Apparatus

The stimuli were type written cards placed in a small box containing a one way viewing mirror. The experimenter controlled a microswitch, which when depressed illuminated the reading material. At the appropriate point the switch was released, effectively removing the lighted stimulus material.

Stimuli

The stimuli were 32 paragraphs selected from the Levin and Kaplan study. One half (16) of the paragraphs contained an X in every inter-word space. In addition to the 'X'-non 'X' category, each group of 16 paragraphs was classified by voice (active, passive), phrase length (4 word, 5 word), and light out position. The light out position was held constant with regard to the phrase position in the sentence always occurring in the phrase just prior to the "by" phrase in passive sentences and the prepositional phrase in active sentences. The light either went out in the middle of the phrase (after the 3rd word in 5 word phrases- the medial position) or at the end of the phrase (the terminal position). Diagrammatically, the paradigm is as follows:



Expectations

The hypotheses tested were:

1. Removing the inter-word spaces should significantly reduce the EVS, forcing subjects into a more elementary process of word reading as opposed to phrase reading.

*2. Following the suggestion of earlier studies, the passive sentences should yield longer EVSs than active sentences. This assertion is based on the assumption that the highly constrained passive form is more prediatable and facilitates phrase processing.

*3. Subjects should read to phrase boundaries more frequently than not.

*4. The EVS is expected to be longer for the terminal or phrase boundary light out position. When the light goes out at the phrase boundary the S will most likely give back the entire next phrase (if it is not too long). If the light is turned off while he is reading in the middle of a phrase he will tend to report only up to the end of that phrase (unless the phrase is short in which case he will pick up the entire next phrase as well).

RESULTS

A summary of the data for the EVS is presented in Table 1. The mean EVS's ("X" = 1.72 words and "non X" = 4.36 words) were compared in an analysis of variance (Table II), and found to be

*This prediation is based on the non "X" or normally spaced sentences.

significantly different ($p < .001$). This confirms the first expectation that filling in the white spaces between words would reduce the EVS, ostensibly forcing Ss to use a more primitive, probably word-by-word, sampling strategy.

This same analysis of variance showed no significant differences between the passive and active sentences in either the "X" or "non X" conditions. This finding does not conform to earlier studies. In fact, a rank order comparison of the data shows that the longest EVS in the "non X" condition was obtained for the active 4 word phrase sentences with the terminal light out position. The effect of "X" ing in is greatest on this sentence form as well, dropping it from the longest to the shortest EVS.

A graph of the passive vs. active data is presented in Figure 1. The plot of active and passive forms of 5 word phrase sentences indicates a major difference in the terminal light out position and this difference is in the expected direction. The mean for active (3.4 words) and passive (4.4 words) were subjected to a t-test for correlated means and found to be significant at the .001 level (two tailed test). This increase in words reported in the passive form occurring with the highly constrained or predictable part of the sentence is a partial confirmation of the Levin and Kaplan study. Why this should be the case with the terminal light off position for 5 word phrases and not so for 4 word phrases is not readily discernable and serves as a point for further inquiry.

In the "non X" or normally spaced sentences the differences between the Medial (3.95) and Terminal (4.74) light off positions was found to be significant ($F = 10.37 \quad p < .005$). The difference in the mean EVS for four word phrases (4.96) and five word phrases (3.73) was also significant ($F = 23.47 \quad p < .001$). Thus, four word phrases with the light out position occurring at the phrase boundary yielded the longest EVS.

Over-all, Ss read to phrase boundaries 115 times and to nonboundary positions 135 times. When these data are corrected for the likelihood of Ss stopping at a phrase boundary vs. any other non-boundary position, a clear tendency toward reading to within sentence phrase boundaries is apparent. A rank order correlation between the length of the EVS and the number of times subjects read to phrase boundaries yielded a $P = .22$. This correlation is not significant ($t = .83$).

One may ask did Ss tend to read to boundary positions as a result of passive or active sentences, 4 or 5 word phrases, or light off position being in the middle vs. the end of a phrase? These data are presented below and it can be seen that the medial position elicits reading to a phrase boundary more than any other factor.

4 word active		4 word passive		5 word active		5 word passive	
M	T	M	T	M	T	M	T
15	12	18	13	19	12	16	8

Table 1

Summary of Mean EVS for all Conditions

"Non X" = 4.36 | "X" = 1.72

"Non X" = 4.36								"X" = 1.72							
Active = 4.32				Passive = 4.36				Active = 1.51				Passive = 1.94			
Phrase length								Phrase length							
4 word= 5.1		5 word= 3.63		4 word= 4.92		5 word= 3.78		4 word= 1.43		5 word= 1.58		4 word= 2.11		5 word= 1.77	
Position				Position				Position				Position			
M= 4.33		M= 3.89		M= 4.41		M= 3.17		M= 1.82		M= 1.56		M= 1.87		M= 1.53	
T= 5.66		T= 3.41		T= 5.44		T= 4.43		T= 1.09		T= 1.59		T= 2.34		T= 2.00	

Table II

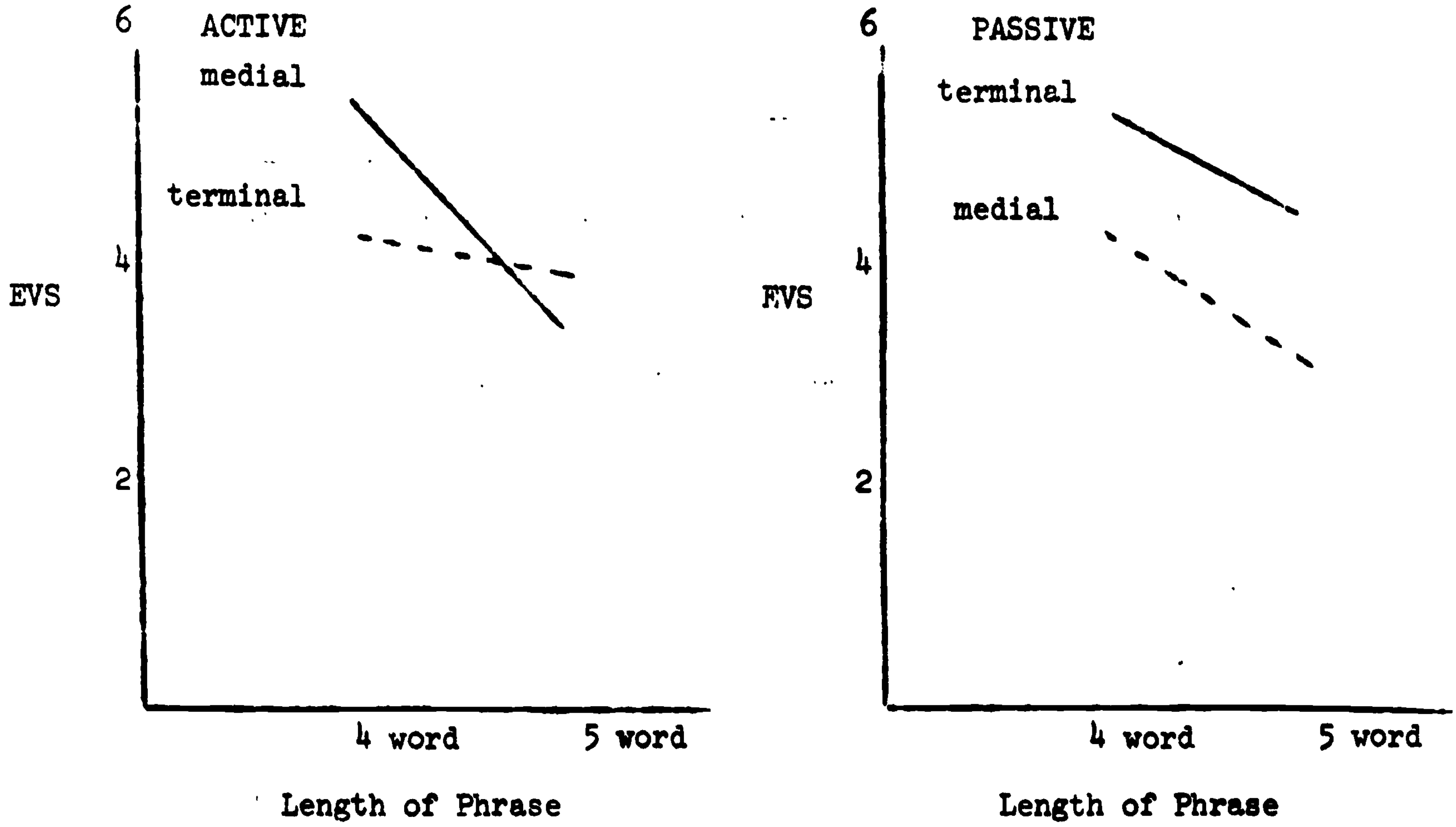
Summary of ANOVA: filled/non-filled spaces: active/passive

Source	d.f.	SS	MS	F
Columns ("X" vs. "non X")	1	433.16	433.16	281.27*
Rows (Passive vs. Active)	1	3.52	3.52	2.28 ns
Cells	3	438.75	(146.25)	
RxC interaction	1	2.07	2.07	1.34 ns
within	252	387.77	1.54	
Total	255	826.52		

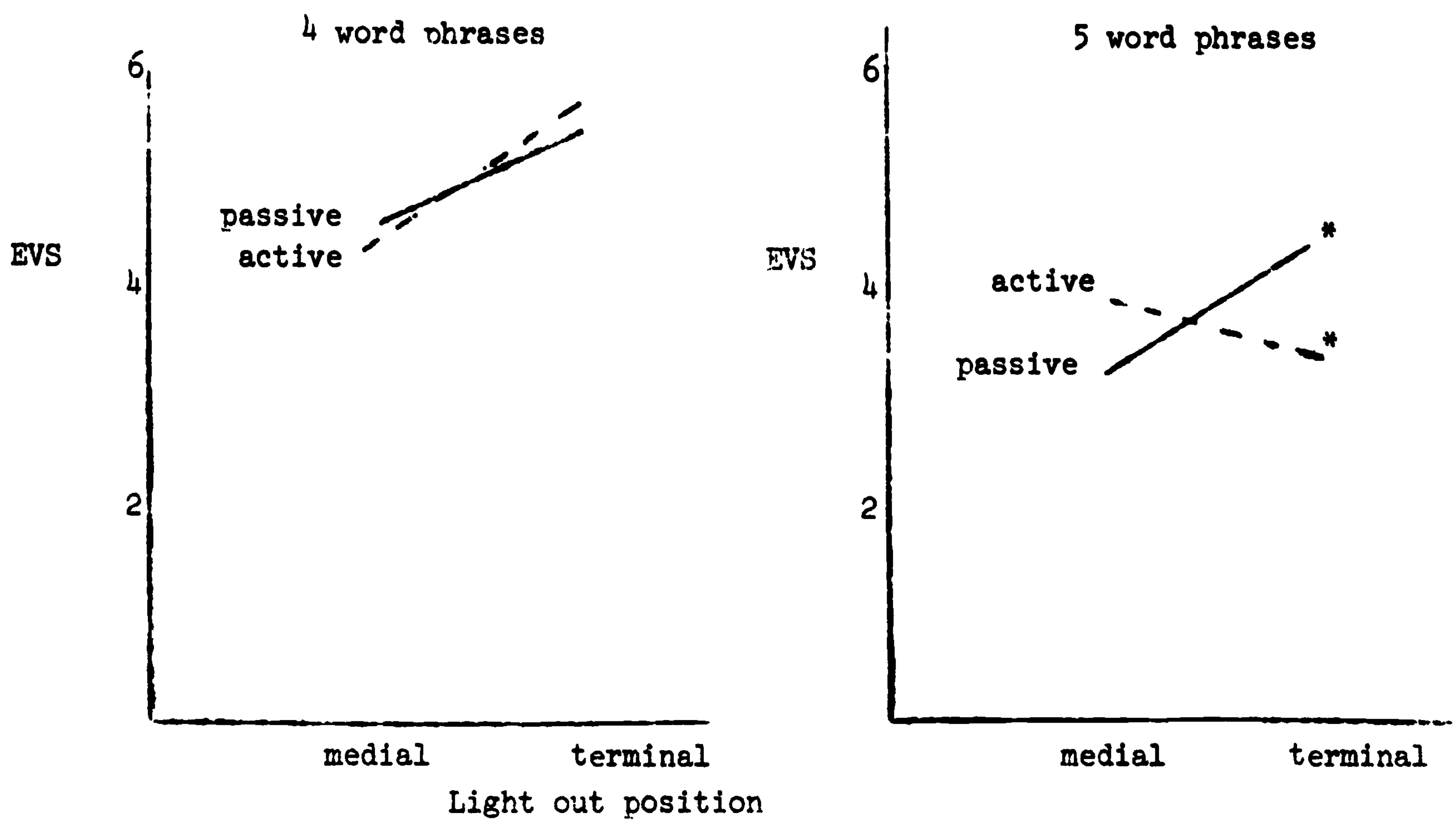
*significant at .001 level
ns = not significant

FIGURE 1

Mean EVS plotted according to light off position for active and passive sentences



Mean EVS for critical Light-off positions



* = significantly different

It should be noted that the EVSs are longest for the terminal light out positions in four word phrases. This could be understood by the combined effect of: (1) subjects stopping at the end of the three word prepositional or passive by" phrase when the light is turned off in the middle of the preceding four word phrase (in this case the phrase boundary restricts sampling ahead). When the light is turned off at the end of the phrase Ss can easily manage the 3-word phrase and most of the next phrase as well. And (2) the restrictions with the end-of-phrase light out position for five word phrases where the phrase following the 3-word prepositional or "by' phrase is too long for most readers to handle. Here again the EVS difference of a full word between active (3.4 words) and passive (4.4 words) sentences illustrates how the more constrained or predictable passive form facilitates sampling the text in reading.

DISCUSSION

The finding that "X"ing in the interword spaces reduced the EVS can be interpreted in two ways. First, that removing the white spaces between words eliminated this useful cue and secondly that interposing the "X" symbol, which is itself a letter (and therefore functions as a relevant cue,) interferes with the identification of the first and last letters of words. A further study was carried out to determine whether it is the absence of space as cues or the nature of the fill-in symbol which limits the sampling and is reported in an addendum to this paper.

Although it is disappointing not to have complete confirmation of a passive-active difference as reported in earlier studies (Levin and Kaplan), the last graph in Figure I shows the trend for an increased EVS in passive forms in the more constrained part (end) of the sentence. Since the same materials were used, it was expected that the results would replicate their findings. It is possible that the failure to find completely compatible data resulted from the distracting effects of shifting from normal paragraphs to stimulus materials with "X" in the white spaces. At any rate, the suggested analysis of the differential distribution of constraints seems to be theoretically sound and a profitable avenue of investigation.

ADDENDUM

The results from the previous study raise several questions:

1. Was the reduction in the EVS the result of removing the white spaces between words or the effect of an increased difficulty in identifying the first and last letters of words because 'X" shares distinctive features with letters? Would another symbol that shares no features with letters produce similar effects?
2. Does the filler symbol have to be placed in every space or can the EVS be reduced equally well by selectively placing the interfering symbol in the target phrase? For example, what is the effect on the EVS if the symbol occurs before, after or surrounding the first word in the phrase compared to the last word?

3. What is the effect on the EVS of turning the light off at the beginning (first phrase), middle (second phrase) or end (third phrase) of the sentence? That is, will the EVS be longer at the beginning or end of the sentence? The Levin & Kaplan data suggest that the EVS tends to extend further toward the end of the sentence because of the higher constraints found in that part of the sentence.

Eight adult subjects were divided into two groups. EVS measures were obtained using the same method reported in the earlier study. For one group "X"s were selectively placed throughout the text. The other group received the same material with asterisks typed in these same positions. Short paragraphs of three to four sentences were used. All sentences were in the active voice and followed the same pattern as illustrated below. Holding light-off position constant for a phrase, the position of the symbol was systematically varied through the first, second, third and fourth word in the phrase, occurring before the word, after it and totally around it. Diagrammatically this differential placement of the interword filler is as follows.

1st phrase	2nd phrase	prepositional phrase		3rd phrase	
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 W2 W3 W4*)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 W2 W3 *W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 W2 W3 *W4*)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 W2 *W3 W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 W2 *W3* W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 *W2 W3 W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(W1 *W2* W3 W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(*W1 W2 W3 W4)	(W1 W2 W3 W4)
(W1 W2 W3 W4)	(W1 W2 W3 W4)	(W1 W2 W3)	↑	(*W1* W2 W3 W4)	(W1 W2 W3 W4)

Parenthesis indicate phrases, W1 W2...stands for the first word in the phrase, the second word. etc., and the arrow points to the place where the light went out.

This scheme was repeated for the first and second phrases in the sentence (a total number of 27 sentences) and the order was randomized for presentation to subjects. Target phrases were matched as closely as possible for both word length, letter similarity and word frequency.

Results

The data for the mean EVS according to whether the sentence contained x or asterisk is presented in Table A. The "X" symbol yielded an average EVS of 4.59 words while the asterisk produced an EVS of 4.01 words. An analysis of variance (Table A) on this data revealed that this difference is not significant (F= .33). This finding suggests that it is the elimination of the spaces between words itself and not the nature of the symbol which is the important feature.

Table A also shows the mean EVS for the phrases within the sentence. The EVS was 3.42 words for the first phrase, 4.29 words for the second phrase, and 5.5 words for the third phrase. An analysis of variance on this data indicate that these means are significantly different ($P < .001$). This finding supports the notion that the EVS varies in accordance with sentence structure and is large for the more highly constrained or predictable part of the sentence, i.e. toward the end of the sentence in this case.

A related finding is the number of instances Ss read to phrase boundaries. Out of 72 opportunities Ss read to phrase boundaries 26 times in the first phrase, 38/72 in the second phrase and 54/72 in the third phrase. In no case did Ss read beyond the immediate phrase to the end of the next phrase when the light went out at the beginning of the first phrase. In only 4 out of the 38 instances of reading to phrase boundaries did Ss read to the end of the next phrase with the light out position at the beginning of the second phrase (three case were the same subject). But when the light went out at the beginning of the third phrase Ss extended their EVS to the end of the next phrase (8 words and the end of the sentence) 28 times.

Table B presents the mean EVS for different words in a phrase according to whether the word had a filler symbol occurring before, after or totally surrounding it. An ANOVA of these data indicate that these means were significantly different ($p < .005$). Although there appears to be no consistently effective means of reducing the EVS by this procedure of selective placement of filler symbols, a graph of the data shows some interesting trends. First, as one might expect initial filling in before the first word of the first phrase is very effective in shrinking the EVS. What is more interesting is the effect of 'X'ing in or placing an asterisk in the phrase boundary of the first two phrases. This procedure significantly curtails the EVS. It is also important to note that with the last word totally surrounded by filler symbols this reduction disappears as though the first asterisk or "X" acted as a kind of "warning." Although this is the case with the first two phrases the drop in the EVS when the phrase boundary is filled in disappears with the 3rd phrase. Apparently the high predictability (from within sentence constraints) occurring in this part of the sentence is sufficient to override the filled" phrase boundary effect, as well as producing longer EVSs and a greater tendency to read to phrase boundaries, reported above.

SUMMARY

The tentative conclusions of this study may be summarized as follows:

1. Normal EVSs were obtained using randomized filler symbols.
2. Finding no difference between the asterisk and "X"es as filler symbols suggests that the white spaces between words are the relevant cues in sampling the written text.
3. The EVS is longer at the end of the sentence than at the beginning.
4. The tendency to read to phrase boundaries increases as one progresses through the sentence.
5. The EVS is particularly susceptible (greatly reduced) to filling in the space in the phrase boundary, except toward the end of the sentences where constraints are higher.

Table A

Mean EVS for "X" and Asterisk by Phrases

	"X"	"*"	
Phrase 1	3.77	2.94	= 3.42
Phrase 2	4.37	4.21	= 4.29
Phrase 3	5.63	4.89	= 5.50
	4.59	4.01	

ANOVA of EVS for symbols, phrase and word in phrase

SOURCE	SS	d.f.	MS	F
symbols ("x" vs "*")	3.4	1	3.4	.33
Subjects	61.8	6	10.3	
Phrases (1st vs 2nd vs 3rd)	155.8	2	77.9	28.9 ($p < .001$)
word in phrase	45.6	8	5.7	3.6 (.005)
symbols x phrase	10.8	2	5.4	2.0
error	32.4	12	2.7	

Table B

Mean EVS for Word in Phrase

first phrase

before & after	*W1 2.75	W1*W2 3.88	W2*W3 4.13	W3*W4 4.0	W4* 2.63
Total	*W1* 2.75	*W2* 3.63	*W3* 3.5	*W4* 3.5	

second phrase

before & after	*W1 4.38	W1*W2 4.0	W2*W3 4.63	W3*W4 5.0	W4* 3.88
Total	*W1* 4.88	*W2* 4.63	*W3* 3.5	*W4* 4.25	

third phrase

before & after	*W1 6.13	W1*W2 5.5	W2*W3 6.5	W3*W4 5.5	W4* 6.00
Total	*W1* 5.25	*W2* 3.63	*W3* 3.38	*W4* 7.25	

REFFRENCES

Levin, H. and Cohn, J. A. Studies of oral reading: effects of instructions on the eye-voice span., 1967.

Hochberg, J., Levin, H. and Frail, C. The effects of removing interword spaces. In preparation.

Levin, H. and Kaplan, F. Studies of oral reading: the eye voice span for active and passive sentences. 1966.

Levin, H. and Turner, E. A. Studies in oral reading: sentence structure and the eye-voice span. 1966.

Schlesinger, I. M. Sentence Structure and the Reading Process., In press, 1966.

Eye-Voice Span (EVS) Within Active and Passive Sentences

Harry Levin and Eleanor L. Kaplan[1]

Cornell University

Recent studies concerned with decoding processes in language usage have examined the ability of subjects to learn, memorize, shadow or produce various linguistic forms. Although studies using these techniques have yielded results of considerable interest, we believe that the study of reading provides an even more powerful way of examining the processing of language. The study reported here, is concerned with the sensitivity of the decoding process in oral reading to manipulation of the constraints within sentences.

The eye-voice span (EVS) provides a sensitive tool for examining the effects of different types of constraint. In oral reading, the eye is usually ahead of the voice; and the EVS is an index of this distance between the eye and the voice. It can be measured by looking at the amount of additional material reported immediately after the original text has been removed. Until recently this technique has been used as a device for assessing reading ability. However, there is evidence that it can also be used to ask questions about the effects of several kinds of variables which are likely to influence decoding strategies. For example, both Lawson (1961) and Morton (1964) have found that the size of the EVS varies as a function of orders of approximation to English. The more redundant the material, i.e., the higher the order of statistical approximation to English, the larger the EVS. Similarly, Schlesinger (1965) and Levin and Turner (1966) using normal text, found the EVS to expand and contract in accordance with the location of constituent boundaries. These studies, however, were not directly concerned with the extent to which intra-sentence constraints are actually exploited by the reader. On the other hand, since there is considerable evidence that such constraints exist, it is likely that subjects may be using them.

Two classes of sentences within which the constraints were known to be differently distributed were selected for study. Clark (1965) had found that the pattern of contingencies between major sentence parts was quite different for active and passive sentences. His subjects generated sentences from active or passive sentence frames from which two or three of the major sentence parts, the actor, verb or object, had been deleted. An uncertainty analysis of the results yielded a measure of both the diversity of each of the sentence parts and the extent to which the sentence parts covaried. The uncertainties associated with the actors, verbs and objects, and the patterns of constraint between them were found to be different within the two forms. Of particular interest here were his observations concerning the directionality of constraint within Active and Passive forms. This is represented in Figure 1. The arrowheads represent the direction of constraint. The verb constrains the

---

Insert Figure 1 about here

---

object and the object constrains the verb in both active and passive sentences. The important finding here is that the latter part of passive sentences, the verb and the actor, is highly constrained by the first part, the object; this was not true for the corresponding parts of active sentences. The latter part of active sentences, the verb and object, were relatively independent of the first part, the actor.

In addition, Clark (1966) and Roberts (1966) later demonstrated that recall for different sentence parts could be predicted from these uncertainties and contingencies.

It was hypothesized, therefore, that if the EVS is sensitive to within sentence constraints, it should increase toward the middle of the passive form but a corresponding increase was not expected in the active form. More specifically, the EVS should increase in size when the reader reaches information which specifies that the sentence is a passive. Although this information is signaled by the form of the verb phrase, direct confirmation comes only when the "eye" reaches the "by" phrase. There is no reason to believe that there is information for this decision prior to either of these phrases.

The previous data, then, indicate that the EVS is longer for constrained or predictable materials and that in the passive, the latter part of the sentence is more constrained by the preceding part than in the active.

Finally, in keeping with the findings of both Schlesinger (1965) and Levin and Turner (1965), it is expected that the EVS will also vary in accordance with both the location and size of the constituents within the sentence, regardless of sentence type.

Method

Subjects. Eighteen college students attending Cornell University were paid for their services as subjects.

Stimuli. Two different phrase lengths and two different sentence types were selected for study, which in combination comprised the four sentence types used:

Active sentences composed of 4-word phrases.

Passive sentences composed of 4-word phrases.

Active sentences composed of 5-word phrases.

Passive sentences composed of 5-word phrases.

The 4-word sentences contained 19 words, broken up into five phrases or constituents: 4 words, 4 words, 3 words, 4 words, 4 words. The 5-word sentences contained 18 words divided into four phrases or constituents: 5 words, 5 words, 3 words, 5 words. The short 3-word phrase represented the "by" phrase in the 4- and 5-word passive sentences and a short prepositional phrase in the 4- and 5-word active sentences. The sentences were constructed so that the first half of both active and passive sentences were structurally identical. For example:

Passive: The cute chubby boy was slowly being wheeled by the maid along the narrow lane to the country store.

Active: The brash tall man was certainly being loud at the meeting of the new group on the main campus.

Each sentence was embedded in a separate paragraph of either four or five unrelated sentences. Since exploratory data indicated that *Ss* scanned the first line before beginning to read aloud, the target sentence was never first; but, it could occur in any other position in the paragraph.

In addition, ten paragraphs made up entirely of unrelated words were included in order to ascertain the relative contribution of syntactic structure *per se* to the variations in the EVS. Finally, an additional 20 paragraphs were included as fillers. The first sentence in these paragraphs was treated as the target sentence in order to encourage *Ss* to attend equally to all sentences in the paragraphs. These 20 sentences were not included in the data analysis. Thus, there was a total of 142 paragraphs. *Ss* were assigned to 1 of 6 different random presentation orders.

Critical positions. In order to satisfactorily examine differences in processing strategies which occur as a consequence of differential linguistic constraints, the EVS was systematically measured at numerous places (referred to as critical positions) within the set of experimental sentences. EVS scores were obtained at various points starting after the third word and after every succeeding word up to the "by phrase" in the passives and to the corresponding point in the active sentences which was a prepositional phrase.

More specifically, the critical positions were distributed as follows. There were 6 critical positions each for the 4-word active and 4-word passive sentence types, and 8 critical positions each for the 5-word active and 5-word passive sentence types. There were four sentences for each critical position, making a total of 24 passive and 24 active 4-word sentences and 32 passive and 32 active 5-word sentences. In the 4-word sentences it occurred after the third, fourth, fifth, sixth, seventh and eighth words in the target sentence. In the 5-word sentences it it occurred after the third, fourth, fifth, sixth, seventh, eighth, ninth, and tenth words in the target sentence. These positions were selected on the basis of results from an exploratory investigation. In addition, the target sentences were positioned so that there were at least three words preceding the critical word on the same line and at least eleven succeeding it for the 4-word sentences and eight succeeding it for the 5-word sentences. Results from an exploratory investigation indicated that these distances were sufficient. Few Ss extended their EVS to the last word in the sentence.

Procedure. The paragraphs were exposed on a small ground glass rear projection screen directly in front of the S, who was positioned so that he could scan the lines with minimal head movement. The size of the letters when projected on the screen was approximately equivalent to that found in texts. A fixation point indicating where the beginning of each paragraph would appear eliminated the problem of having the S search the screen each time a new paragraph was exposed. The contrast between the letters and the background was sufficiently low as to eliminate any afterimages.

As soon as the paragraph appeared on the screen the subject began to read aloud at his normal reading rate. When the S reached the predetermined place in the target sentence, the projector shutter was closed, removing the material from view. The subject then proceeded to report all material seen but not yet read aloud. The number of correct words reported was taken to be his recall EVS for that sentence. Following this report, as a control for both guessing and for any tendency for a subject to be conservative in his report, a recognition test was administered.

A full record was made of where the S's voice was when the shutter was closed, by simultaneously recording the S's voice on one track of a stereo tape and the shutter click on the other track. The time taken to read to the critical position was recorded on a Hunter KlocKounter which was activated by the opening and closing of the shutter.

Recognition test. A series of recognition lists were constructed, one for each target sentence. All content words occurring after the critical word were included on the list. Function words were omitted since such words occurred more than once in a given target sentence making it difficult to know which function word was being referred to. In order to discourage guessing the lists were constructed so that only 50% of the words actually appeared in the target sentence, while the other 50% were words that were visually and semantically similar to those which had appeared in the sentence. That is, the first three or four letters of these words were identical to those that had actually appeared; and, in addition, they could be readily substituted for the words that had appeared in the text. After reporting all words seen but not yet read aloud (recall EVS), the S searched through the list of words circling any additional words that he recognized as having appeared in that sentence. This second score was taken as a measure of the S's recognition EVS. The S was informed that only half of the words on the list had actually appeared in the sentence.

Results and Discussion

Two sets of EVS scores were obtained for each S: (1) the recall EVS, the number of correct consecutive words reported after the shutter was closed, and (2) the recognition EVS, the recall score plus the number of additional words correctly selected from the recognition lists. In the latter case, only the recognition words selected that immediately followed to the right of those already reported were included in the recognition EVS. However, since only content words were included on the lists a S was given credit for any intervening function words. For example, in the sentence, "The cute little girl with the long hair..." if the critical position occurred after little, and the S reported having seen girl as well as recognizing long and hair, he would be credited with a recognition EVS of 5 words and a recall EVS of one word. The recognition EVS was not corrected for guessing since the ratio of incorrect to correct recognition word choices was approximately 1 to 1,000.

The pattern of recall and the recognition EVS scores is equivalent across all sentences. As might be expected, the recognition EVS containing both recall and recognition components, is higher than the recall score alone. In the four-word phrase sentences, the mean recall EVS is 4.5 words and the recognition EVS is 4.9 words; for the five-word phrase sentences, the two mean scores are 4.0 and 4.4. Thus, although the recognition EVS score was consistently higher at all critical positions sampled within both active and passive sentences, the recall and recognition EVS scores were distributed similarly. Since the recognition EVS score yielded no new information, the subsequent analyses treat only the recall EVS scores.

Four-word sentences. In a three-way analysis of variance with repeated measures for the 4-word sentences, the main effects, Subjects, Position, and Type are significant beyond the .001 level. Only the Position X Type interaction was significant ($p < .001$). The Scheffé test for comparison of treatment means reveals that Positions 1 through 4 are significantly different from Positions 5 and 6 ($p < .05$) (see Figure 2). Active and Passive sentences do not differ for Positions 1

Insert Figure 2 about here
----

through 4, but they do for critical Positions 5 and 6. There are no differences among the critical positions in the active sentences. Among the passives, however, Positions 1 through 4 differ significantly from Positions 5 and 6 at the .01 level.

Five-word sentences. A similar three-way analysis of variance for 5-word sentences again yielded three significant main effects, Subject, Position and Type. Again only the Position X Type interaction is significant ($p < .001$). The Scheffé test for differences among treatments means shows that Positions 1 through 6 are significantly different from Positions 7 and 8 ($p < .01$) (see Figure 3). Active and passive sentences

----
Insert Figure 3 about here
----

do not differ for Positions 1 through 6, but they do for Positions 7 and 8 ($p < .01$). No positions in the Active sentences differ from each other. For the Passive Type, Positions 1 through 6 are significantly different from Position 7 and 8 ($p < .01$).

The findings, then, are (1) in sentences composed of four- and five-word phrases, the EVS span is longer for the passive at the two terminal critical positions than for the active sentence; (2) within the active sentences, the position in which the shutter was closed had no effect on the EVS; and (3) in the passive sentences the EVS at the two final positions was larger than at critical positions earlier in the sentence. These findings support the major hypothesis that the EVS varies in accordance with intrasentence contingencies. The results show that the EVS is longer for passive sentences at that point where the active and passive forms begin to be differentially constrained (Clark, 1965). Since (1) the first portions of both active and passive sentences were identical and (2) the short three-word phrase in the active sentence was a prepositional phrase where the intra-phrase contingencies would not be expected to differ from those within the passive "by" phrase (Aborn and Rubenstein, 1958; Treisman, 1965; Fillenbaum, et al., 1964), these differences must be attributed to the structure of the sentence as a whole.

That these data represent differences in decoding which are dependent upon differential linguistic constraints is further supported by three additional findings. First, and indeed an important point, is the low false alarm rate for the recognition test. The ratio of correct to incorrect word choices on the recognition test was well over 1000:1. Considering the way in which these lists were constructed, it seems most improbable that the EVS represents subjects' ability to guess. Rather it is more likely that the reader was utilizing the structure present in the written material. This is supported by a second finding that the EVS was constant at all positions in the strings of unrelated nouns and verbs.

The EVS for these sentences was fixed at 2 words at all positions sampled.

And, finally, in keeping with the findings of both Schlesinger and Levin and Turner, the EVS for both Actives and Passives expanded and contracted in accordance with the location and size of sentence constituents. To evaluate this relationship between the EVS and sentence constituents, the frequency with which Ss read to constituent boundary and non-boundary positions was compared. This was accomplished by weighting the two scores by the probability of S's stopping at either boundary or non-boundary positions. The tendency for the EVS to coincide with constituent boundaries was highly significant ($p < .001$). (Overall, Ss read to boundary positions 361 times and to non-boundary positions 202.7 times.) This lends additional support to our notions concerning decoding processes and linguistic structure especially since there is ample evidence (e.g., see Fodor and Bever, 1965) for the existence of greater constraints within sentence constituents than between them.

Conclusion

The two findings which emerged from this study are that the EVS increases within the passive sentence at that point where the constraints within active and passive sentences begin to differ, and that the EVS tends to terminate at phrase boundaries. Both results lead to the interpretation that the amount of written material which is picked up and processed depends on the type of constraint and the amount of predictability within messages. This interpretation, however, does not explain how the reader is able to make use of available structure. Similar to the suggestion made by Mehler, Bever and Carey (1967) it is possible to argue that overt scanning strategies, or more specifically eye movements, are different for variously structured material. On the other hand, it is also possible that these data are the result not of some overt process, but of some internal decision-making mechanism, which utilizes its knowledge of linguistic structure to determine how the message shall be processed, such as in Analysis by Synthesis. Future research will be directed toward delineating between these two interpretations.

References

Aborn, M., Rubenstein, H., & Sterling, T. Sources of contextual constraint upon words in sentences. J. exp. Psychol., 1959, 57, 171-180.

Clark, H. Some structural properties of simple active and passive sentences. J. verb. Learn. verb. Behav., 1965, 4, 365-370.

Clark, H. The prediction of recall patterns in simple active sentences. J. verb. Learn. verb. Behav., 1966, 5, 99-106.

Fillenbaum, S., Jones, L. V., & Rappoport, A. The predictability of words and their grammatical class as a function of rate of deletion from a speech transcript. J. verb. Learn. verb. Behav., 1964, 2, 186-194.

Fodor, J. A. & Bever, T. G. The psychological reality of linguistic segments. J. verb. Learn. verb. Behav., 1965, 4, 414-420.

Lawson, Everdina. A note on the influence of different orders of approximation to the English language upon the Eye-Voice Span. Quart. J. exp. Psychol., 1961, 13, 53-55.

Levin, H. & Turner, Ann. Sentence structure and the Eye-Voice Span. Mimeograph. 1966.

Mehler, J., Bever, T. G., & Carey, P. What we look at when we read. Percept. & Psychophys., 1967, 2, 213-218.

Morton, J. The effects of context upon speed of reading, eye-movements and Eye-Voice Span. Quart. J. exp. Psychol., 1964, 16, 34-354.

Roberts, K. The interaction of normative associations and grammatical factors in sentence retention. Paper read at MPA, Chicago, 1966.

Schlesinger, I. M. Sentence structure and the reading process. Unpubl. Monograph, 1965.

Treisman, Anne. Verbal responses and contextual constraints in language. *J. verb. Learn. verb. Behav.*, 1965, *4*, 118-128.

Footnotes

1/This study was supported by funds from the U.S. Office of Education. We acknowledge with thanks the contribution of computer facilities by the Bell Telephone Laboratories, Inc., Murray Hill, N.J.

Figure Captions

Figure 1. Directionality of the constraint between sentence parts (After Clark, 1965).

Figure 2. Mean Recall EVS for Various Critical Positions: Four Word Phrase Active and Passive Sentences.

Figure 3. Mean Recall EVS for Various Critical Positions: Five Word Phrase Active and Passive Sentences.

First Graders' Use of Grammatical Context in Reading [1]

Rose-Marie Weber

Cornell University

To demonstrate that early readers use context to facilitate their identification of words in a passage would only confirm the experience of anyone who has observed a child learning to read. But little can be said specifically, for example, about the interplay between the use of graphic cues on one hand and the use of contextual cues on the other, about the developmental stages in the exploitation of cues by the novice, or about the relative import of various linguistic and extralinguistic features that comprise context. In particular, grammatical structure as an aspect of context has hardly been considered in regard to reading, despite its central position in the language as the vehicle for semantic as well as extralinguistic content, and despite the well-known restrictions on the occurrence of words in sentences which grammar entails. It is obvious from their abilities to speak and understand that first graders control the grammar of their language with only minor substantive differences from the way their parents do. The extent to which they bring their grammatical competence to bear on the reading task is another matter, however. In the studies of errors during oral reading reported below, the sensitivity of first graders to grammatical structure is described in an attempt to assess the grammatical dimension of their reading performance.

The contribution of verbal context to adults' reading has been investigated from several points of view, but all in all has received little attention. The facilitating effects of preceding context have been reported by Tulving and Gold (1963) and Morton (1964), who found lower recognition thresholds for tachistoscopically presented words exposed in the context of a meaningful sentence compared to anamolous contexts or in isolation. Salzinger, Portnoy, and Feldman (1962) and Coleman (1963) are among those who have analyzed the improvements in Ss' cloze test scores with increasing statistical approximations to English--that is, decreasing contextual distortions--and have thus complemented the evidence from perception and learning studies for the integrative power of linguistic structure. In the literature on children's reading abilities, the cloze technique has been used as a measure of comprehension, but consideration of the use of context in and of itself is almost entirely absent. One exception is Goodman's report (1965) that young readers were less accurate in reading words presented in a list than when the same words were incorporated into stories.

Context in these studies usually refers not only to the grammatical structure of sentences, but also to their communicative content, including such features as truth value. Although semantic and syntactic features have been separated in perception and memory studies, their differential effects on reading have come to attention only recently. Ford and Levin (1967) compared the recognition of homographs in a syntactic frame and in the context of a word with high associative value and found that the syntactic frame had greater effect on eliminating the ambiguity of the stimulus. Ruddell (1965) constructed passages in accordance with the grammatical structures found in the speech of fourth graders and reported higher comprehension of the passages composed of high frequency patterns than of those made up of infrequent patterns.

Our approach to young readers' use of the constraints imposed by their grammar is through the analysis of oral reading errors. The degree to which an error approximates a correct response can be gauged at several linguistic levels. Consider the sentence He shook the pig and out came some money. The erroneous response 'dimes' for money would suggest that the reader ignored the graphic display and responded in terms of the context established by reference to a situation involving a piggy bank and by the expectation that a word following some at the end of a sentence would complete a noun phrase. On the other hand, the response 'many' for money would indicate that the reader attempted to respond in terms of the relations between letter and sound patterns, but at the expense of attention to the preceding context, including the specific grammatical restrictions within a noun phrase beginning with some. Such an analysis emphasizing the correct features of an erroneous response can reveal a reader's appropriate strategy in the use of graphic and grammatical cues in spite of his imperfect handling of some feature or other. Therefore, as a first step in assessing how early readers bring their grammatical competence to bear on the reading task, we consider the acceptability of oral reading errors to their preceding grammatical context within the sentence. The assumption here is that an acceptable response demonstrates the readers' sensitivity to grammatical constraints. The distinction between such a response and one that upsets the grammatical structure of a sentence, for instance, 'many' for money in our example, is significant simply because the grammatical structure of a sentence is central to its communicative function. Although not all ungrammatical sentences are incomprehensible, their lack of coherency often precludes interpretation. Next, in order to demonstrate young readers' handling of graphic and grammatical cues relative to one another, we report the degree to which responses approximate written words in terms of graphic similarity and with this index we show the interaction of graphic similarity and grammatical acceptability. Finally, from another perspective on children's sensitivity to grammatical constraints, we note how the grammaticality of a set of errors affects the readers' own correction of those errors.

The first corpus of oral reading errors was collected in a first grade class of 21 children in the Ithaca City School District, Ithaca, New York (Class I). From December through May two observers noted the errors as the children read from their texts during daily small-group instruction. Scores on the Lorge-Thorndike IQ test (n=20) averaged 109.2, s.d. = 14.5. The 1950 series of the New Basic Readers (Scott-Foresman) served as the principal reading materials; high achievers went on to supplemental readers (The Reading Caravan, D.C. Heath, 1964). Instruction largely followed the recommendations of the teacher's guides, supplemented by the presentation of all consonant sound-letter correspondences to the class as a whole. During most of the observation period, the class was divided into four instructional groups according to the teacher's judgment of the children's abilities. For purposes of this discussion, the two groups that showed the most progress are combined into the High Group (n=12) and the other two into the Low Group (n=9). By May almost all the children in the High Group could handle unfamiliar materials, but the Low Group did not have the skills to transfer to words that they had never seen before. This difference shows up on the results of the

Metropolitan Achievement Test: the High Group (n=11) scored at mean grade levels 2.6, 2.9, and 2.8 on the word knowledge, word discrimination, and reading subtests, respectively, while the Low Group (n=8) scored 1.8, 1.8, and 2.1 on these subtests.

Another corpus of oral reading errors was collected in a second Ithaca class of 24 children (Robinson, et al., 1966). The mean Lorge-Thorndike score of 110.5 (n=23, s.d. = 9.2) was not different from that of Class I, but the significantly smaller variance (F = 2.48, 19, 22 d.f.; $p < .025$) reflects the greater homogeneity of these children. Here the data were collected under conditions that allowed rechecking. The children's reading was tape-recorded monthly in the presence of a familiar adult. The selections, drawn from the various texts that comprised the instructional materials, showed a greater range of vocabulary and style than the basal readers used in Class I. The reading program in this class was deliberately eclectic, including a rather thorough phonics program involving auditory and visual discrimination, attention to consonants in various positions and early introduction of vowels. For most of the data collection period, November through June, the class was divided into three main instructional groups. In June they numbered 11, 7, and 6 children, in order of decreasing achievement. The Metropolitan Achievement Test scores indicate that the overall progress in this class was greater than in Class I. On the word knowledge, word discrimination, and reading subtests, the High Group (n=9) scored 3.0, 3.1, and 3.4; the Mid Group (n=7) scored 2.5, 2.8, and 2.6; and the Low Group (n=6) scored 1.8, 1.9, and 1.8.

The first grade reading materials were not especially designed for the investigation of errors, but were selected as representative of the materials that the children faced every day. This means that few children read identical passages, and the selections were not necessarily well-matched to the children's abilities. This lack of control may obscure many of the variables that deserve attention. It is defensible, however, on the grounds that these data reflect the behavior of children learning to read under typical circumstances.

Several categories of erroneous responses were identified and errors were classified according to these categories. Only responses that differed overtly from the text were considered. That is, no hesitations or refusals to respond were taken into account. The substitution of an erroneous single word response for the expected response was taken as the basic error type. Several other types were also identified: a word omitted impulsively from the response, a word spoken although no corresponding written word appeared on the page, and a reversal in word order, possibly in combination with other types of errors. Note that while these categories could be applied to letter-sound discrepancies, they were applied only to whole words here. Following is an example of each type of error:

Type	Printed word/Response
(1) Substitution	funny/family
(2) Omission	the <u>black</u> umbrella/the umbrella
(3) Insertion	down the creek/down <u>to</u> the creek
(4) Scramble	In went the animals/In they all went

A total of 1072 errors were recorded in Class I and a total of 871 in Class II. As will be seen, the achievement groups within each class are more or less evenly represented in these totals. However, it should be noted that the higher groups covered much more reading material than did the low groups in making comparable numbers of errors.

From the distribution of the types of errors presented in Table I, it can be seen that substitutions are by far the most frequent.

Insert Table I about here

Judgments on the grammatical acceptability of an error were to be made in terms of the words in a sentence that preceded the error. Therefore, since not all errors are amenable to this analysis, certain errors had to be eliminated from consideration. First of all, errors of omission, although they can be judged in terms of their effect on the entire sentence, are not subject to judgment within preceding context only. Secondly, words that stand at the beginning of the sentence are hardly subject to any grammatical restrictions, so that including them in the sample of errors to be judged would inflate the total of grammatically acceptable responses. In Class I 23% (241) of the total number of errors occurred at the beginnings of sentences. When these were eliminated along with the omission errors (some omissions were initial), a total of 753 usable errors remained. In Class II, 15% (125) of the errors were made on the first word. With the elimination of these and the omissions, 718 errors remained.

In order to assess the use of grammatical context by the children in these classes, the errors were judged with respect to grammatical acceptability within the preceding context of the stimulus sentence.

All the stimulus sentences were accepted a priori as grammatical, although some sequences in the atrophied language of primers may be questioned. The analysis was undertaken on the assumption that any error which maintained the sentence as a grammatical sequence would share a significant grammatical property of the stimulus word that was misread: privilege of occurrence in a syntactic construction. An error was judged acceptable if the stimulus sentence could be completed after the point of the error in any way, not necessarily by the remainder of the stimulus sentence. Consider the following examples of errors:

	Printed sentence	Erroneous response
(1)	Spot can help Dick.	Spot and ...
(2)	Puff did not say what she wanted.	Puff did not say that ....
(3)	I will see what it is.	*I will see what is it.
(4)	She looked and looked, but she could find no food at all.	*She looked and looked, but she cold....

Of these four examples, the first two errors were judged as grammatically acceptable with respect to the preceding context of the sentence; the sequence with the error could be completed into a grammatical sentence. The second two, on the other hand, were judged as grammatically unacceptable. By their very occurrence, they upset the grammatical structure of the sentence.[2]

The analysis for grammaticality in the performance of the children in Class I, who read from basal readers, shows that an overwhelmingly large proportion of the errors (91%, n=753) were grammatically acceptable to the preceding context. The differences in achievement between the High and the Low Groups in reading, as reflected not only by progress through textbooks, scores on tests, but also overall classroom performance, led us to expect that the groups would differ in the use of preceding verbal context. However, the difference in grammatical acceptability between them is negligible: 92.3% (n=465) for the High Group as compared to 88.9% (n=256) for the Low Group.

In Class II, we might expect that the more varied style of writing and the greater range of vocabulary that the children faced would affect their use of grammatical context. Again, however, the proportion of grammatical responses approaches 90%. For the class as a whole 87.7% (n=718) did not violate preceding grammatical constraints. Among the three ability groups within the class, the proportions show that the strong readers are no more successful in this respect than their weaker classmates: High 87.5% (n=192); Mid 87.0% (n=299); Low 89.4% (n=227).

Given the slots in the sentences where the errors occurred, we do not know the proportion of random words that would be grammatical in these contexts. It is clear from the 90% figure, however, that words are not identified by these children without reference to preceding syntactic constraints, and that this aspect of context is exploited, perhaps sometimes overused, by them. Moreover, the high proportion of grammatically appropriate errors reflects the strong expectations by first graders that written sentences will conform to the restrictions that the grammar of their language imposes. These findings, then, do not support the characterization of the relatively low achiever as a word-by-word reader. Rather, they suggest that children--no matter what their potential for acquiring literacy skills--bring to the task a fundamental linguistic ability, which in its rigidity shapes their reading responses into familiar language structure.

Many of the substitution errors are grammatically acceptable to preceding context as well as graphically and phonologically similar to the stimulus words, e.g., the substitution of 'that' for what in the sentence Puff did not say what she wanted. However, there are instances of errors that are grammatically acceptable--even semantically appropriate--but graphically dissimilar, such as the substitution of 'tell' for say in the same sentence. Such errors would indicate a heavy dependence on contextual cues and disregard of the graphic display and its relationship to the sound system of English. On the other hand, there are errors, such as 'cold' for could in the sentence, She looked and looked, but she could not find any food at all, which suggest close attention to the spelling and its correspondence in sound, but at the expense of the information provided by the preceding context. The interplay in the use of information from the stimulus display itself as opposed to the information from context has been investigated by Tulving and Gold (1963) and Morton (1964). In both studies evidence was found to confirm the hypothesis that the amount of information needed about the stimulus for its identification varies inversely with the amount of available contextual information. From our analysis of the grammatical acceptability of reading errors, we have seen that first graders exploit contextual information to a high degree. How might this aspect of their performance relate to their use of graphic information? With the hypothesis about the inverse relation between context and stimulus in mind, we might suppose that in the relatively rare cases when the readers disregarded the grammatical constraints of preceding context, their attention was directed to analyzing the details of the graphic display, or even to working out the relationships between the letters of the stimulus and its pronunciation. Evidence for this attention to graphic information could be demonstrated if the errors which did not conform to preceding context approximated the stimulus words more closely in terms of graphic features than did the contextually appropriate errors.

In order to describe the degree to which an erroneous response approximates the stimulus in terms of graphic patterns and therefore phonological patterns, an <u>index of graphic similarity</u> was devised. By this measure, various graphic characteristics of the stimulus are compared with characteristics of the erroneous response word, transcribed in traditional orthography. The features taken into account are the number of letters shared by the stimulus-response pair, the position of shared letters within the words, the position of shared letters relative to each other, the average length of the words, and the difference in length between the stimulus and response words. Two features have not been taken into account: the similarity in the shapes of letters, such as <u>o</u>, <u>c</u>, and <u>e</u>, and the distinction between upper and lower case letters. It should be noted that these features have no phonological correlates.

The graphic similarity scores were computed according to the following formula:

$$GS = 10 \left[\left(\frac{50F + 30P + 10C}{A} \right) + 5F_1 + 27B + 18L \right]$$

F = the number of pairs of adjacent letters in the same order shared by S and R:

S HOUSE / R HORSE F = 2

S EVERY / R VERY F = 3

R = the number of pairs of adjacent letters in reverse order shared by S and R:

S WAS / R SAW R = 2

C = the number of single letters shared by S and R:

S SPOT / R PUFF C = 1

S FAMILY / R FUNNY C = 2

A = average number of letters in S and R:

S EVERY / R VERY A = 4.5

R_1 = ratio of number of letters in the shorter word to the number in the longer:

S EVERY / R VERY $R_1 = 4/5$

B = 1 if the first letter in the response is the same as the first letter in the stimulus; otherwise B = 0:

S FAMILY / R FUNNY

E = 1 if the first letter in the response is the same as the last letter in the stimulus; otherwise E = 0:

S FAMILY / R FUNNY

Examples include:

SPOT/PUFF	75	FAMILY/FUNNY	528
ARE/EAT	117	BUMPED/BANGED	683
WILL/LIKE	200	DUCK/DUCKS	754
THE/TO	343	BEGAN/BEGUN	820
WANTS/BOATS	450	THREE/THERE	960

The weights assigned to the selected features reflect our intuitions about the significance of various cues for the identification of words. For example, the greater weight given to shared beginning letters over end letters, and in turn the weight given to shared end letters over common letters elsewhere in the word reflect the importance of the positions of letters for word recognition. The literature has shown (e.g., Marchbanks and Levin, 1965) that readers, as they master left-to-right orientation, exploit the letters in the end positions as salient cues yielding high information. Because common adjacent letter patterns reflect the formation of units of a higher order than common single letters, special value is assigned to adjacent pairs, especially if they are in the same order. Since the number of adjacent pairs is a function of word length, the average number of letters is included in the formula.

Some validation for the usefulness of the index was provided by adults' rankings of word pairs. Two lists of ten words and their misreadings were selected from the errors in our study. The pairs on one list were chosen arbitrarily, while those on the second differed by roughly 100 points according to the graphic similarity index. Fifteen college students were asked to rank the word pairs on a list in terms of the similarity of their appearance. The rankings within the arbitrary list correlated .93 with rankings based on the graphic similarity index; within the selected list, the rankings correlated .89.

In and of themselves, the graphic similarity scores for the classes under study are of interest, for they reveal the expected differences between the high and low achieving subgroups. The mean score for all the substitution errors in Class I (n = 856, including 146 with no shared letters and therefore scores of 0) was 350.79. The Low Group mean of 267.47 (n = 353) is well over a hundred points below the High Group mean of 407.87 (n = 503). The mean graphic similarity score for Class II was similar to the mean for Class I: 356.44 (n = 818, including 79 with no shared letters). The means for the various ability groups were: High 396.11 (n = 215); Mid 363.36 (n = 340); Low 315.05 (N = 263).

However, we are not concerned only with the evidence for the more or less successful use of graphic cues, but rather with the evidence for the interplay between the use of such cues and the carryover of information from the preceding context. The hypothesis on the inverse relation in the use of contextual and stimulus information suggested that grammatically unacceptable responses would on the whole share more graphic features with the stimulus words than would the responses that conformed to preceding grammatical context. Table 2 shows the mean graphic similarity of the substitution errors of Classes I and II according to their grammatical acceptability.

Insert Table 2 about here

Support for the hypothesis is indicated by the fact that the mean for ungrammatical errors exceeds the corresponding grammatical mean in every comparison.

From these results we can infer that when the readers neglected the constraints of the preceding grammatical context they were attending to the task of identifying and perhaps decoding the features of the graphic display. Thus the inverse relationship found in adults' handling of the two types of information is demonstrated by our findings with children also. It should be noted, however, that Morton's and Tulving and Gold's adult subjects compensated for an experimentally distorted context by exploiting the stimulus information; the experimental materials were manipulated so that an imbalance between the contextual and stimulus information was created. The performance of the children in this study, however, indicates an emphasis on one of the two sorts of information in a situation in which both are available. The errors, of course, demonstrate the misuse of available graphic cues. But the increase in graphic similarity scores when preceding grammatical context is disregarded suggests that a relatively intensive analysis of the graphic display sometimes results in the neglect of contextual information. For these beginning readers, the ability to use information from both sources efficiently is not entirely in hand. But it should be recalled that disregard of preceding grammatical constraints shows up in only about 10% of the errors; the children seldom neglected grammar for letters.

Up to this point we have considered only the verbal context that precedes an error. But, although an error may be appropriate to what precedes, it may not fit into the subsequent context of a sentence as it is written. The effect that an error may have on the entire sentence is also interesting, for the coherency of the sentence as it continues after an error is certainly significant to a reader's comprehension of the sentence. Thus, in order to assess the children's sensitivity to the grammatical effects of the errors they have made, we noted their own corrections of their errors in light of grammatical acceptability of the error with respect to the entire sentence.

Errors made in the presence of a prompting teacher are not amenable to the analysis of children's own spontaneous corrections of their errors. For this aspect of our study, five weekly readings by 20 members of Class I during the last two months of instruction were tape-recorded. Each child read a selection into a microphone. No one was present to interrupt, look critical, or supply corrections, and so the children could ignore or correct whatever errors they made. Each passage was a familiar page from the regular instructional materials (Scott-Foresman, New Basic Readers, 1950's; D.C. Heath, The Reading Caravan, 1964) that had been covered in small-group instruction during the previous week. As before, the class is divided into the High (here n = 11) and Low (n = 9).

Of the total 200 errors under consideration, 54% (107) were made by the members of the High Group. Their rate of errors per 100 words was 3.9 in contrast to the Low Group's 6.7, indicating that the better readers read much longer texts to accumulate a comparable number of errors. The proportion of substitution errors for the class was 84%; omissions, insertions, and scrambles comprised the rest.

In order to describe the grammatical acceptability of an error in the context of the entire sentence, as error was first scored for its acceptability within the preceding context and then for its effect on the remainder of the sentence. Consider the following:

Printed sentence	Response
(1) Spot can <u>help</u> me.	Spot can <u>hear</u>/ me./
(2) He said, "<u>Can</u> I help you?"	*He said, "<u>Come</u>/ I help you?/
(3) Down comes the <u>car</u>.	*Down comes the <u>cars</u>.//

Judgments of grammaticality were made in the response sentence immediately after the occurrence of the error and at the end of the sentence, that is, at the points indicated by the slashes, although the children did not necessarily read to the end of the sentence after making the error. All in all, the effect on the grammatical structure of each sentence in which the errors occurred was scored as

1. Grammatical up to and including the error and grammatical to its end.
2. Grammatical up to and including the error, but ungrammatical to its end.
3. Ungrammatical up to and including the error, and therefore ungrammatical to its end.

The 200 errors occurred in 167 sentences. In cases of more than one error in the sentence, the written stimulus was, as usual, taken as the right-hand context. However, for second or third errors in cases of more than one, the partial sentence as it was read, i.e., including earlier errors, was taken as preceding context. For instance, if the stimulus sentence was <u>Where did Bunny go?</u> and the child responded 'Where is Bunny going?' the response 'going' was scored within the context 'Where is Bunny....' It should be noted that because this analysis was concerned with both the preceding and following contexts of the errors as well as children's sensitivity to the errors, misreadings at the beginning of the sentences and omission errors were not discarded. Rather, for these errors the preceding context was scored as grammatical.

The distribution of the errors according to their effect on the grammaticality of the entire sentence is given in Table 3 for the High and Low Groups as well as for the Class I as a whole.

Insert Table 3 about here

The proportion of errors that were judged grammatically acceptable with regard to the preceding context, as shown in columns 1 and 2, is similar to the corresponding proportion presented earlier: 94.5% (91.0% of 134 if 46 beginning errors and 20 omission errors are discarded). The High and Low Groups show no significant differences. Although only about 6% of the errors in this corpus immediately upset the grammatical structure by their occurrence, another 32% render their sentences ungrammatical when the context following the error is taken into consideration.

The evidence for the children's sensitivity to preceding grammatical constraints suggests that they would bring their knowledge of structure to bear on the correction of these errors. Here the problem is not simply to identify a word, but to notice that a response is incorrect, to locate the error, and usually to re-identify a word. We would expect that the errors which violated the grammatical structures of the sentence would be more frequently noted than those that did not. Those errors that conformed to the grammatical structure, on the other hand, would be passed over because, on the grammatical level at least, they would not be noticeable.

Table 4 shows the frequency with which the children disregarded or corrected the errors that maintained the sentence as a grammatical string in contrast to those that rendered the sentence ungrammatical.

Insert Table 4 about here

A greater proportion of all errors was disregarded rather than corrected. But whereas the grammatical errors were disregarded more than twice as often as they were corrected, the ungrammatical errors were corrected nearly twice as often as they were ignored. The effect of the error on the grammatical structure of the sentence, then, is indeed recognized by young readers. Again they demonstrate their sensitivity to the structure of the language, but in this case they are evaluating their responses rather than anticipating them.

We noted that there was no significant difference in the use of preceding context between the High and the Low Groups. However, a breakdown of the data as shown in Table 5 indicates that the above effect of ungrammaticality on corrections is created almost entirely by the performance of the High Group.

Insert Table 5 about here

The High Group passed over 73% of the grammatical errors, but ignored only 15% of those that upset the grammaticality of the sentence. In contrast, the Low Group passed over 68% of those that maintained the sentence as a grammatical string, a figure comparable to the High Group's, but passed over an almost equally high proportion of errors that upset the syntactic structure, 58%.

Having made a response that was incorrect, the better readers showed by their corrections that they were sensitive to the grammatical context of the entire sentence. The fact that the poor readers did not correct their ungrammatical errors to the same extent does not necessarily indicate that they were insensitive to deviant grammatical structures. Perhaps such readers do not have efficient strategies for finding errors that upset syntax, or perhaps they simply have a different standard of what is acceptable as oral reading.

The basically linguistic nature of the reading task has not received adequate formulation that would point up its fundamental similarity to the perception of spoken language. These analyses of oral reading errors have provided substantial evidence that beginning readers use their knowledge of grammar to narrow down the words which compete for a given sentence slot, just as they surely do in understanding speech. The materials that the children read here were not adjusted to the skills of the individuals; their handling of easy materials in contrast to difficult ones was overlooked. The shifts in strategy that might have taken place with maturity were also ignored. But from a broad perspective, the notable finding was that weaker readers do not differ from their more skilled classmates in respect to the use of grammatical constraints for the identification of words in a string. It is as though the children resisted uttering a sequence that did not conform to an acceptable sentence. Having made errors that did not fit into the grammatical context of the written sentences, however, only better readers consistently demonstrated their rejection of ungrammatical sentences by correcting themselves.

References

Chomsky, N. Aspects of the Theory of Syntax (Massachusetts Institute of Technology Press, Cambridge, Mass., 1965).

Coleman, E.B. Approximations to English: some comments on the method. American Journal of Psychology, 1963, 76, 239-247.

Ford, B. & Levin, H. Studies in oral reading VII: Homographs in a semantic context. 1967, Mimeo.

Goodman, K. A linguistic study of cues and miscues in reading. Elementary English, 1965, 42, 639-643.

Marchbanks, G. & Levin, H. Cues by which children recognize words. Journal of Educational Psychology, 1965, 56, 57-61.

Morton, J. The effects of context on the visual duration thresholds for words. British Journal of Psychology, 1964, 55, 165-180.

Robinson, J., et al. The first grade study, 1965-66. Project Literacy, 1966, Mimeo.

Ruddell, R.B. The effect of oral and written patterns of language structure on reading comprehension. Reading Teacher, 1965, 18, 270-275.

Salzinger, K., Portnoy, S., & Feldman, R.S. The effect of order of approximation to the statistical structure of English on the emission of verbal responses. Journal of Experimental Psychology, 1962, 64, 52-57.

Tulving, E. & Gold, C. Stimulus information and contextual information as determinants of tachistoscopic recognition of words. Journal of Experimental Psychology, 1963, 66, 319-327.

NOTES

1 These studies were supported by funds from the United States Office of Education of the Department of Health, Education, and Welfare. Assistance from the staff of the Ithaca City School District, Project Literacy, and the Laboratory for Research on Language Skills at Cornell is gratefully acknowledged.

2 Scorers were familiar with the notion of grammaticality as it has been elaborated in grammatical theory. As acceptable strings, they included only those that they judged grammatical. It should be noted, however, that violations of selectional rules, as Chomsky has formulated their characteristics (1965, p. 149, passim), were allowed as grammatical. Ten per cent of the classroom data from Class I and the taped data from Class II were double-scored for reliability; the entire corpus from the tape-recordings in Class I (see section 3) was double-scored. In all cases agreement among scorers on the judgment of grammatical acceptability was over 90%.

Table 1

Frequency of Error Types

Class	Sub.	Omit.	Insert.	Scramble	Total
I	79.9 (856)	8.5 (91)	9.2 (99)	2.4 (26)	100.0 (1072)
II	93.9 (818)	3.2 (28)	2.6 (23)	.2 (2)	99.9 (871)

Table 2

Mean Graphic Similarity Scores

According to Grammatical Acceptability

Group	Grammatical	Ungrammatical
I. High	379.81 (365)	550.47 (32)
Low	255.26 (218)	467.33 (30)
Total	333.24 (583)	507.02 (62)
II. High	392.17 (159)	414.39 (23)
Mid	347.13 (254)	452.85 (39)
Low	294.09 (198)	433.76 (25)
Total	341.66 (611)	437.20 (87)

Table 3

Frequency of Various Grammaticality Judgments

Group	(1) Gram to err; gram to end.	(2) Gram to err; ungram to end.	(3) Ungram to err and to end.	Total
High	68.2 (73)	28.0 (30)	3.7 (4)	99.9 (107)
Low	55.9 (52)	36.6 (34)	7.6 (7)	100.1 (93)
Total	62.5 (125)	32.0 (64)	5.5 (11)	100.0 (200)

Table 4

Corrections According to the Grammatical

Effect of the Error

Grammaticality	Disregarded	Corrected	Total
Gram. to end	70.4 (88)	29.6 (37)	100.0 (125)
Ungram. to end	38.8 (29)	61.3 (46)	100.1 (75)
Total	58.5 (117)	41.5 (83)	100.0 (200)

Table 5

Corrections According to the Grammatical Effect of the Error - High and Low Groups

Group	Grammaticality	Disregarded	Corrected	Total
High	Gram. to end	72.6 (53)	27.4 (20)	100.0 (73)
	Ungram. to end	14.7 (5)	85.3 (29)	100.0 (34)
	Total	54.2 (58)	45.8 (49)	100.0 (107)
Low	Gram. to end	67.3 (35)	32.7 (17)	100.0 (52)
	Ungram. to end	58.3 (24)	41.5 (17)	100.1 (41)
	Total	63.4 (59)	36.6 (34)	100.0 (93)

A STUDY OF THE STICK-IN WATER ILLUSION WITH CHILDREN

James J. Gibson, John M. Kennedy, and Thomas L. Toleno

Cornell University

As part of a research project on the comparison of mediated perception with direct perception, we have been testing the hypothesis that the ordinary array of light coming to the eye contains information about the properties of the objects reflecting the light (Gibson, 1966, Ch. 10). The array of light coming to the eye from a motion picture, a drawing, a diagram, or a page of writing may also contain information about objects but this is perception at second hand, and the information in these displays is of an increasingly different sort from that in ordinary light. Instead of depending on universal laws of projective geometry and perspective transformations, the information increasingly depends on the conventions of picture-making, of graphic symbols, and of alphabets (op cit., Ch. 11).

We assume that the child begins to perceive the world at the outset with no awareness of the perspective appearances of things, but with a crude awareness of their main distinctive features. He detects invariants, not colored forms in a visual field. He is wholly unconscious of his retinal images. Later, as he learns to scribble, to draw, and finally to write, he also learns to perceive by means of tracings, pictures, and writing. In this way his visual perception, even direct perception, comes to be more and more under the influence of social conventions and arbitrary symbols. He becomes susceptible to the pictorial mode of visual experience, and aware of the temporary perspectives in his field of view. Only then is the child struck by the contrast between visual appearances and visual reality. The naive mode of direct visual experience, however, is never wholly superseded. This fact is realized, if dimly, by proponents of visual education and art education.

On this theory, form sensations are the incidental symptoms of invariant-detection, not the basis of perception. It is therefore not the phenomenal constancy of the shape and size of objects that needs to be explained by learning, but the ability to notice sensations--to "see" one's retinal images. In this respect, the theory of information-based perception turns the classical problem of space perception upside down.

A difficulty for this theory, it can be argued, lies in the fact that the light to the eye of an observer does <u>not</u> always contain ecologically valid information about objects, but often contains misinformation. Perception, in short, is inescapably subject to illusions. In such cases, the only way to achieve correct perception is said to be by an intellectual process that overcomes, or corrects, or neglects the false impressions of sense. The true properties of the object can only be apprehended by an <u>indirect</u> process of inference, judgment, or measurements of it, or on a sense that is more trustworthy than vision.

Perhaps the clearest case of an optical illusion to which this argument applies is the stick-in-water. The edge of a straight stick partly submerged in a pool of water appears bent at the air-water interface because the perspective of the edge in the array of light is in fact a bent line, not a straight line. This is true at all angles of the edge to the surface of the water except the perpendicular. How, then, does an observer ever come to apprehend that the stick is really straight? This problem has been a focus of discussion among philosophers and psychologists for a long time (eg., Austin, 1962).

This study asks three questions. First, is there any way in which children can be "shown," not merely "taught," that an apparently bent stick in water is really straight? Can they be given information leading them to realize that it is straight without telling them so, without rewarding them for saying so (Braine and Shanks, 1965), without ever letting them see the stick perpendicular to the surface of the water, without ever taking the stick out of water, without letting the child touch the stick, and, of course, without giving formal instruction in the laws of the refraction of light?

Second, will the children given this demonstration, or at least some of them, detect the straightness of the stick and assert that it is straight under questioning by the experimenter? The demonstration consists of rotating the stick-in-water around itself as an axis, after seeing a stick-in-air turned around itself as an axis. The child is told only to "watch carefully and see what happens."

Third, what is the invariant over time, the "distinctive feature" of the changing array, that specifies the physical straightness of the stick-in-water? Note that it cannot be simply the optical straightness of the perspective, although this does indeed specify the physical straightness of the stick-in-air. The perspective appearance of the stick-in-water is never straight. The information for its real shape, if it exists, must be independent of its appearances.

The Experiment

Apparatus. The apparatus consisted of a pair of 18-inch wooden sticks attached to ball-and-socket joints at the bottom of a tub filled with water (see illustration). They were square in cross-section. One was straight; the other was bent at an obtuse angle of 160°. The bent stick showed a hairline junction where two parts were joined, and the straight stick showed a similar junction at the same point. When the two sticks were arranged in a vee so that the junctions coincided with the water surface, the physically straight stick had a bent-line perspective and the physically bent stick had a straight-line perspective, as shown in the picture. In effect, the bottom of each stick (and the bottom of the tub) appeared higher than it really was, because of refraction at the water-to-air surface. The lengths of these perspective lines varied, of course, as the observer moved toward or away from the tub, or around it, but the rectilinearity of one and the angularity of the other did not change. The straight stick looked bent and the bent stick looked straight from any visual standpoint.

The physically straight stick was painted black at the top, and the physically bent stick white, so that they could be given arbitrary designations in talking to an observer about them. An identical pair of sticks also attached to ball-and-socket joints was placed for comparison near the tub of water. Since these were wholly in an air medium, the perspective of the straight stick was, of course, straight and that of the bent stick was bent, from any standpoint. (There are unique viewing points, to be sure, from which an angular edge looks straight and a straight edge looks like a point instead of a straight line, but these did not enter into the present experiment.)

Each of the four sticks could be held near the bottom by the experimenter and rotated on its ball-and-socket joint. The physically straight sticks do not show any optical change or "wobble" when thus moved. The physically bent sticks do.

Subjects. A group of 55 children ranging from five to eleven years of age were put through a standard procedure that involved the matching of each stick-in-water with a stick-in-air. Each child was tested individually at school, during school hours. This group includes five children who were eliminated from the experiment either because they could not consistently use (or learn to use) the terms straight and bent (or equivalents) as applied to the sticks-in-air, or because they could not be unequivocally shown to have the expected illusions. Five more had to be eliminated because their later answers were irrelevant or their interest could not be maintained. This left a total of 45 useful subjects.

Procedure. The plan of the experiment was first to establish that a child would match the optically bent stick-in-water with the physically bent stick-in-air and the optically straight stick-in-water with the physically straight stick-in-air when they were motionless; then to rotate the sticks without suggestions or instructions except to pay attention; and finally to determine whether or not he had learned from this demonstration to match the physically straight sticks with one another and the physically bent sticks with one another, despite the illusory optical appearances. The test for learning was therefore stringent, and the learning, if it occurred, had to be strictly perceptual.

A child was shown the setup and then questioned about the sticks on the table. "Are these two sticks the same or different?" and "In what way are they different?" If necessary, he was asked "Would you pick out the straight stick?" and "What do you call the other stick?" and if necessary, "Would you pick out the bent stick?"

Having established that the child could distinguish the shape-quality of straight and bent (or crooked) his attention was drawn to the tub of water. He was told, "This stick we call 'black' because it has a black top, and this stick we call 'white' (pointing to the sticks in the tub)." He was then asked to match a stick on the table with a stick in the tub, "Is this stick (pointing) more like the black stick or more like the white stick?" The same question was repeated for the other stick. Fifty out of 55 children matched in accordance with the false optical information, that is, they showed the illusion.

Next, the child was told, "Now I'm going to turn some of the sticks and I want you to watch carefully and see what happens. And I want you to tell me what you see happening." E rotated the straight stick-in-air (which caused no change in the straight optical line of its edge) and then the bent stick-in-air (which caused a "wobble" of the bent optical line of its edge) and finally the straight stick-in-water (which caused no change in the bent optical line of its edge). Upon seeing this fact, some children seemed to apprehend at once that the apparently bent stick was really straight. If so, the child was immediately given an opportunity to reverse his former match, "Is this stick (pointing) more like the black stick or more like the white stick?" If not, a second and a third demonstration of the consequences of rotation was presented and finally the <u>bent</u> stick-in-water was turned (which caused the straight optical line of its edge to change from straight to angular and back to straight again). In all, six opportunities were given the child to reverse his former matches.

At the end of the experiment, each child who had changed his mind was asked "How do you know that the black stick (white stick) is like this one?" and then "What is it that tells <u>you</u>?" And, finally, "If we were to take this stick (pointing) <u>out</u> of the water which one of these (pointing) would it be more like?" The purpose of this inquiry was to see whether the child would remain firm in his new insight as to real straightness, or would revert to his original perception of apparent straightness.

<u>Results</u>. Of 45 children who showed the stick-in-water illusion by the criterion of matches based on optical appearance, 17, or more than a third, reversed their matches to ones based on an invariant over time when they were presented with this information. Seven of these 17 persisted in maintaining that the apparently bent stick was "more like" the straight stick than the bent stick when they were again faced with the original motionless situation and were required by <u>E</u> to justify their choice. Four of these seven (all being 10 or 11 year-olds) were able to make some reply to the question "How do you know? What is it that tells you?" The others answered simply "I saw it," or merely asserted "It's straight." The latter children could not even begin to describe

the information that led to their conviction, but were nevertheless convinced.

The children who reversed their matches were scattered through all age-groups from five to eleven. The number who learned did not increase significantly with age, nor did the promptness with which children learned in the course of the six demonstrations increase significantly with age. There is no evidence for stages of development in these results.

Discussion

The array of light to the eye of an observer coming from a stick partly in air and partly in water contains misinformation about the shape of the stick. In air, rectilinear edges project as rectilinear perspectives and we can trust this information about the edges of objects. The same thing is true in water; fish can also trust their eyes with respect to the straightness of underwater edges. But in a space of air <u>and</u> water this rule of optical projection does not hold (the speed of light being slower in water than in air) and hence the edge that crosses an air-water boundary may be physically straight but optically bent, or *vice versa*. The appearance is compelling but it does not correspond to the reality; it is an illusion. Considering this and other illusions, philosophers and educators have urged us not to trust the senses but only to trust the intellect.

The results of this experiment seem to show, however, that the array of light coming from a *rotating* stick, an event in time as distinguished from a fixed thing in space, contains information about the shape of the stick, correct information. Some children seem to be able to register or detect this information. Their perception was not a process of reasoning, or of correcting impressions of sense. When asked how he knew that the stick was straight a child might say "I saw it."

What is this information? It is something that becomes evident when the edge is rotated on itself as an axis and ceases to be evident when the motion ceases. The perspective of the straight edge is never a straight line during the motion. But the perspective _is_ invariant with rotation, and in this respect it behaves like an axis, that is, like a mathematically straight line. Perhaps what some children detected was this special criterion of rectilinearity.

There are many ways of defining straightness in mathematics--the ray of light in a homogeneous medium, the stretched string, the shortest distance, the arc of a circle of infinite radius, the line transposable along itself--but the criterion of the absence of "wobble" during rotation is an especially visible one. This distinctive feature of the moving object, not any intellectual construct, may well have been what the children detected.

References

J. J. Gibson. _The Senses Considered as Perceptual Systems_. Houghton Mifflin Co., Boston: 1966.

J. L. Austin. _Sense and Sensibilia_. Oxford, 1962.

M. D. S. Braine and B. L. Shanks. The conservation of a shape property and a proposal about the origin of the conservations. _Canad. J. Psychol._, 1965, 19, 197-207.

A New Theory of Scribbling and Drawing in Children

James J. Gibson and Patricia M. Yonas

Cornell University

"The studies that have been made of the development of scribbling in young children are not very revealing, except to show that children seem to enjoy it. But scribbling is not simply play, or an opportunity for the child to 'express himself;' it is an opportunity for the educating of visual attention and for learning to perceive in new ways" (Gibson, 1966, p. 230). It was predicted from this formula that young children who would scribble with an ordinary crayon or pencil would refuse to continue scribbling when given a special crayon or pencil that left no visible trace. The making of traces on a surface, the controlling of the displayed trace, and the seeing of these new display-variables were assumed to motivate the act of scribbling, not the transient feedback from the activity itself. A test of this prediction is made in the first experiment to be reported.

A second prediction was that young children at the scribbling age would be relatively unwilling to "draw a picture in the air" with a pencil when asked to do so, although older children are known to comply with such a request. The common hypothesis that scribbling yields satisfaction as a motor activity implies that younger children would be at least as willing to do so as older ones. The following tabulation lists the response-produced stimulation arising from each of the three different acts being studied.

	Normal scribbling	Traceless scribbling	Drawing in air
Kinesthesis from joints and muscles	present	present	present
Visual motion of hand and tool	present	present	present
Pressure of tool on skin of hand	present	present	present
Resistance and friction of tool on surface	present	present	absent
Trace of moving tool on surface	present	absent	absent

The first four kinds of input ("reafferent" input or "feedback") are transient inasmuch as they cease when the act ends. The last, however, involves a source of visual stimulation that outlasts the act, and this is assumed to be critical. (For a more elaborate analysis of proprioception and haptic sensitivity, see Gibson, 1966, ch. 6-7).

The background of these experiments is a theory of the development of graphic activity and of pictorially mediated perception in the child and the human species (Gibson, 1966, ch. 11). The theory postulates a "fundamental graphic act." Examples of it are scribbling or finger painting which leave deposits on a surface, and scratching or grooving which leave indentations on a surface. Any surface thus altered provides a new source of visual stimulation, that is, a display in the general meaning of the term.

Experiment I

The hypothesis to be tested is that children will be unwilling to move a stylus against a surface when they discover that it does not leave a trace, as compared with doing so when it does leave a trace, despite the equivalence of the two acts in all other respects. The relative amount of time spent scribbling with each of the tools, then, is the principal index of motivation in this experiment. With such an index, it is possible to observe both verbal and preverbal Ss. A range of different ages and different amounts of scribbling experience was sampled to bring out possible developmental differences.

Method. Two identical manual tools, only one of which produced a trace, were compared. After exploring various possibilities, the non-tracing tool was made from a wooden dowel, painted and shaped to look exactly like the tracing tool, which was a large, No. 2 lead pencil. Care was taken to make the tools equally sharp. (The experimenter could not distinguish between them on the basis of tactual feedback alone.) Double sheets of white paper, $11\frac{1}{2}$" x 17", were taped to masonite boards of the same size. Ink embedded in the second sheet was released by pressure applied to the first; thus a record of the movement of the non-tracing tool was obtained, although it was not visible to the child when he used the stylus.

Fourteen children, ranging in age from 15 to 38 months (mean age 28 months) were observed in their homes in a free play situation, with the mother and occasionally an older sibling present. Two Es were necessary to run the experiment, one to keep time during the sessions, the other to direct the child's activity. To avoid creating a test-like atmosphere, instructions were minimized; the Es simply explained that they had brought some toys along because "they liked to watch children play." When rapport had been established, the child was seated at a table, on the floor, or in his mother's lap. The active E placed a paper-and-board before him and then handed him one of the tools, remarking that it was a "very nice pencil." Most Ss proceeded to scribble without further instruction; a few of the younger Ss responded only after a short demonstration of scribbling by E or by an older sibling. Each child was given a session with both tools, the order of presentation being alternated from S to S.

A stopwatch was started when S began to scribble and was stopped during those intervals when he was not scribbling. It was not stopped when he momentarily paused to point out aspects of the scribble or talk about it. These latter pauses were very short, and since they occurred consistently and seemed to indicate interest in the task, they were considered as part of the time that the S engaged in scribbling. The session was terminated when S said he was finished, or when he asked for another piece of paper, or stopped scribbling. However, if he wished to end the session before 10 seconds had elapsed, he was encouraged to "play a little longer." If S had scribbled for 90 seconds, he was told to tell E when he was finished so he might be given some new material. (Pretests revealed that younger Ss were inattentive during the second session if permitted to scribble for more than 90 seconds during the first.) In the case of younger Ss who were either unable or unwilling to verbalize their wish to stop, repeated rejection of the tool or inattentiveness were taken as the criterion for ending the session; this procedure resulted in a slight overestimation of very short sessions. In most cases, S was not aware that the session was being timed, since the E who operated the stopwatch sat at some distance from him.

Results. Table 1 shows the length of time in seconds during which each child engaged in scribbling with the tracing tool and the nontracing tool. Children are listed in order of increasing age. For all Ss, elimination of the trace significantly reduced scribbling activity; the means were 71.7 seconds with the tracing tool and 20.6 seconds with the nontracing tool ($t = 5.35$, d.f. = 13, $p < .001$ 2-tailed).

Insert Table 1 about here

The following observations also support the hypothesis that a lasting trace must occur if scribbling is to be motivated. When using the tracing tool, Ss often called attention to their scribbles by pointing or naming, but this typical behavior did not occur when the tool left no trace. This is not surprising, but it shows the hypothesized importance to the child of the external display. The common reactions to the non-tracing tool included (1) frequent examination of the tool or the paper, (2) increased pressure as judged by the heavier impressions left by the carbon sheet, (3) puzzled looks at the E, and (4) distractableness. Furthermore, eight Ss made the source of their confusion explicit with such remarks as "This one can't work," "It's broken," or "This doesn't got ink!" It was also noted that, whereas scribbling without accompanying visual attention to the paper was rare, it occurred more often when the nontracing tool was used. Two Ss, for example, having discovered that the tool did not produce a trace, continued to move it very slowly across the paper but watched the E instead of the paper in a disconcerted manner. It was as if they expected some further trick to be played on them. Finally, there was a fairly consistent tendency for Ss to produce vertical or horizontal back and forth strokes with the nontracing tool, however complicated or advanced were their scribbles with the tracing tool. A possible interpretation of this finding would be that Ss revert to a more primitive form of scribbling when using the nontracing tool (Lowenfeld and Brittain, 1964) but a simpler explanation is that the back and forth motion is simply the common procedure for "making a pencil write."

The data can be expressed as the ratio of the time employed with the nontracing tool to the time employed with the tracing tool. These ratios are also presented in Table 1. The average amount of time spent with the nontracing tool was only one-third that spent with the tracing tool. We might have predicted that, if children scribble in order to achieve traces, they would not use the nontracing tool at all. This hypothesis was verified in four cases, numbers 6, 7, 12, and 13. These children stopped immediately as soon as they discovered that the tool did not "work." But the children had been implicitly instructed to scribble by being asked to play with paper and with what appeared to be a pencil. Subject 11, for example, was very acquiescent, acting only at E's suggestion during the entire session. Although she scribbled with the nontracing tool longer than any other child, when asked at the end of the session whether she liked that "pencil," she said she didn't like it because "it didn't write." Moreover, any experience with pencils should create expectations of being able to make the pencil produce traces, and some time might be required to discover that the nontracing tool could <u>not</u> be made to do so.

Our hypothesis asserts that scribbling is motivated from the outset by the immediate satisfaction of seeing a trace or display. An alternative hypothesis is that scribbling has only an "activity motive" at the outset; that the child has to learn by association to expect a trace following on the manipulation, after which the trace might contribute to the motivation. On this latter hypothesis, tolerance of the nontracing tool should decrease with age and experience. However, the correlation between age and such tolerance is not significant ($r = -.40$), although a slight trend in the appropriate direction is evident. This is not enough evidence to suggest that the satisfaction of seeing a trace depends on a learned expectation. There is other evidence, on the contrary, suggesting that the satisfaction is immediate and automatic. The behavior of the 16-month-old S (number 2) is regarded as particularly significant here, since she had had no experience with tracing tools prior to our observations (although she may have watched her older brother scribbling). The child was first given the nontracing tool but could not be induced to scribble, even in imitation of her brother. She was next given the tracing tool. She responded as before--waving the tool and occasionally striking the paper with it--until an apparently fortuitous look at the paper as she pounded it with the stylus. From that moment the child scribbled, with great interest and increasing control. Although she had previously gripped the tool in her fist, she came to hold it overhand style. E presented the nontracing tool a second time at the first pause in the child's activity, since it was feared that her attention would wander before a comparison of the two tools could be made. This session was shorter and the child reverted to pounding the paper. It seems, then, that although she had not been taught the use of pencils nor the process of creating "pictures," her interest lay in the production of traces. When they were not forthcoming, scribbling stopped. It is likely that we observed in this child the first manifestation of scribbling, and this seemed to be a discovery of the "fundamental graphic act."

Experiment II

A test of the hypothesis that children scribble so as to carry out motor activity for its own sake would be to ask them to draw in the air with a tracing tool. They should be willing to do so, even after scribbling has developed, if the act originated in this way. Motor kinesthesis from joints and muscles is the same as in trace-making and the "visual kinesthesis" of seeing the hand-and-tool move is also the same. The grasp of the tool is the same. The visual contact of the tool with the surface is absent and the haptic feeling of pressure on the surface and friction over the surface is absent. The gesture as such, however, remains, although the recording of this gesture on the surface has been eliminated.

In order to test this prediction, four three-year-old nursery-school children were asked the following questions:

Do you ever draw pictures in the air?

Can you make a picture of a (ball, or other appropriate object) in the air with this pencil?

Show me how you do it.

If you pretend that there is a big piece of paper here, can you draw a picture? (Why not?)

Do you think this is a good way to make pictures? (E gesturing with pencil.)

All of the children refused to draw in the air, even when told to "make believe" that paper was present. They did not seem to approve of it. These results cannot be attributed to a general unwillingness to perform; all of the children asked for paper on which to draw a "real" picture, one which they could "see." This confirms our own prediction, and casts further doubt on the hypothesis that scribbling is a purely motor activity or that it begins as one.

Discussion

The theory that perception is based on the pickup of information, not on the organizing or interpreting of sensations (Gibson, 1966) distinguishes between direct or immediate perception of the environment and indirect or mediated apprehension based on human artifacts or "surrogates." The former affords perception at first hand; the latter provides for a kind of perception at second hand. Graphic art in general, pictures and drawings in particular, and the special case of writing are all types of man-made sources of stimulus information that permit mediated perception (Gibson, 1966, ch. 11. See also Gibson, 1954, for an earlier version of the theory). Art, picturing, and writing have all developed in man during the last twenty or thirty thousand years. Presumably they all have their root in trace-making and this is why the "fundamental graphic act" is psychologically important. It is probably also important in the development of the child. The foregoing experiments tell us something of its motivation.

The development of display-making in the child. The act of scribbling, daubing, finger-painting, scratching, or altering a plastic surface is not at the outset an act of communication or a social act. It seems to be an act with the sole purpose of producing a new source of optical stimulation that can be looked at by its producer and that continues to be visible. It <u>displays</u> his handiwork. It continues to be visible, of course, not only to him but to others, and the child soon wants others to look at his scribbles, but the trace making begins as a controlled sequential changing of the reflecting capacity of a surface. The trace converts a movement in time into a frozen form in space, and the form is even more interesting to see than the movements of the hand in the air. The latter is a transient feedback that occupies the attention of younger infants.

The graphic act continues to be interesting in later life. The "doodling" of adults has at least this much in common with the work of non-representative painters: it is an exercise in producing and discriminating optical structures. It is good practice in perceiving but it is not communication.

<u>The development of depicting in the child</u>. The fundamental graphic act soon begins to differentiate. Parents encourage the child to "draw things." They would like to think that he can "draw from memory" and represent what he knows. The adult perceives outlines on paper as he would the discontinuities in an optic array that specify the physical edges of objects in the world (the figure-ground phenomenon). Hence any slight resemblance between outlines on paper and the edges of an object meets with parental approval. Eventually the child himself will begin to detect that the edge-information in light can be partly reconstituted

by a linear trace, and this helps him to discriminate straightness, curvature, bentness, tilt, and the openness or closedness of lines. These discriminations are necessary if he is later to perceive writing and printing.

The child has been naively registering the physical edges of objects all along, but now he may begin to notice the perspectives of these edges, the perspectives of balls and boxes, of houses, faces, men, and animals. They are not, of course, frozen in time like his tracings on paper but ever-varying. Nevertheless if he holds still he can freeze them. If he notices this he will begin to be able to take the pictorial attitude. This is a special kind of attention, quite different from his ordinary attention only to the formless invariants of things.

The origin of depicting in our prehistoric ancestors was not, of course, helped by encouragement from the elders. The first cave-painter had to discover for himself the equivalence of lines to edges. But when he did, and when he found that he had made a mammoth, say, appear on the wall of a cave, he must have been astonished, and it must have seemed magical (Gibson, 1966, p. 228 ff). He had been scribbling and finger-painting, one can be sure, long before he made this discovery.

The development of writing in the child. When scribbling has sufficiently elaborated, and when the seeing and producing of line qualities has progressed, the stage is set for the child to learn the skill of alphabetic reading and writing. Reading is perceptual while writing is motor, we say, but the two aspects of literacy cannot be separated, except arbitrarily. The ability is much more demanding than that required for pictorial recognition and pictorial communication since it requires an extra stage of mediation. Pictures freeze some of the direct information about the environment in light, whereas writing freezes speech, which is already indirect information. The child learns to read and write much later than he develops the ability to see and make pictures. Similarly, our human ancestors invented alphabets much later than they did pictures.

The development of mediated cognition in the child. Pictorially mediated perception and verbally mediated cognition enable the child to acquire knowledge about the environment as well as direct knowledge of acquaintance with the environment. The child can then be shown or be told, or be taught. But the visual mediators of second hand knowledge are only superficially understood by psychologists and educators. What we have to understand is the information about the world conveyed by pictures, motion pictures, sculptures, models, toys, exhibits, graphs, writing, print, and books.

Human artifacts can be classified as graphic or plastic, but there is no sharp division between them. Scribbling, drawing, painting, diagramming, mapping, handwriting, and printing are said to be graphic. All these involve traces on a surface. Until very recently in history, the traces had to be made by hand. The manual act of trace-making helps the child to distinguish the variables of graphic information. Some of these variables are straightness, curvature, bentness, tilt, closedness, intersection, and symmetry but there are many more of them not yet analyzed. (An attempt to discover the variables used by children to distinguish capital letters from one another has been made by E. J. Gibson, 1965. A beginning at the psychophysics of pictorial perception has been made by Hochberg, 1962.) The variables of graphic information seem to combine into higher order variables, perhaps without limit. They are endlessly interesting even as such, whether or not they make representations or ideographs, or numerals, or alphabetic letters. The graphic artist is fascinated by them (Kepes, 1944). Even the child at 16 months of age begins to be interested in graphic information by the evidence of the experiments here reported.

References

Gibson, J. J. A theory of pictorial perception. Aud.-Vis. Communica. Rev., 1954, 1, 3-23.

Gibson, J. J. The senses considered as perceptual systems. Boston: Houghton Mifflin, 1966.

Gibson, Eleanor J. Learning to read. Science, 1965, 148, 1066-1072.

Hochberg, J. The psychophysics of pictorial perception. Audio Vis. Communica. Rev., 1962, 10, 22-54.

Kepes, G. The language of vision. Geo. Theobald, 1944.

Lowenfeld, V. & Brittain, W. Creative and mental growth. New York: Macmillan, 1964.

Footnote

1. This research was supported in part by Grant OE 6-10-156 to Cornell University from the U. S. Office of Education, for the analysis of reading skill. One of the subprojects under this grant is for the comparison of mediated with direct perception. The assistance of Elizabeth Goertz in carrying out the experiment is acknowledged.

Table 1

Time Employed in Scribbling with Tracing and Nontracing Tools

Subject	Age (mo.)	Duration of activity (in seconds): Tracing tool	Duration of activity (in seconds): Nontracing tool	Ratio of time with nontracing to time with tracing tool
1	15	72	30	.417
2	16	22	16	.727
3	18	75	34	.453
4	22	35	15	.428
5	23	30	10	.333
6	25	72	3	.042
7	30	123	4	.033
8	30	55	29	.527
9	33	145	12	.083
10	35	90	41	.455
11	37	115	65	.565
12	37	53	5	.094
13	38	53	2	.038
14	38	64	23	.359
Mean	28	71.7	20.6	.325

Effect of Instructions on the Abstraction of Spelling Patterns

Arlene Amidon
(Abstract of MA Thesis Research)

The present experiment was an attempt to examine conditions under which pre-readers could learn to detect regularity in letter sequences. Previous research had indicated that abstraction of spelling patterns was a difficult if not impossible task for first-grade children. In the present study, it was hypothesized that first-grade children could learn to detect regularity in spelling patterns if they were provided with knowledge of what kind of structure to look for and practice in perceptual search across items.

Three groups of ten S's were given practice on a card-sorting task with four-letter words, half of which contained a common two-letter cluster. Instructions were varied according to group. The Rule group was told to sort the items on the basis of any common two-letter cluster. The Letter group was told which specific letters to look for on each problem. The Control group was given no information.

Results indicated that S's provided with a rule and practice in searching for a common spelling pattern performed significantly better on a post-test with new problems than did S's provided with practice in identifying specific solutions on each problem. The Rule group also showed a significant decrease in response latency between the first day of practice and the post-test, suggesting that they had become more efficient in processing information. The Letter S's, however, did show some transfer to new problems, for their error scores on the post-test were significantly lower than those of the Control group.

These results imply that detection of invariance is largely a matter of learning how to approach a problem and developing an appropriate perceptual strategy. Such learning can best be facilitated by directing the S's attention to invariant features present in the stimulus material.

The Acquisition of Information-Processing Strategies in a Time-Dependent Task

Albert Yonas
(Abstract of Ph.D Thesis Research)

The purpose of this study was to demonstrate that perceptual learning is in part a process which reduced the stimulus information that must be detected, economically selecting a small set of features for testing when the task and the alternatives make them informative. The experimental situation allowed subjects to differentiate the displays on the basis of a single distinctive feature test, as contrasted with a control condition in which equally economical functioning was not possible. Experiment 1 attempted to show that the perceptual process can narrow the set of tests to a single one, when full identification of the display is unnecessary, and also to determine whether such learning is equally rapid among 7-, 11- and 18-year olds.

A disjunctive reaction time procedure was used in which subjects categorized the single letter presented as belonging to the positive set or not by pressing one of two buttons as rapidly as possible. Each subject received 135 trials in each of three conditions: the E condition, with a single letter in the positive set; the ANV condition, with three letters in the positive set all sharing a common feature (diagonality) not found in the negative set; and the AOF condition, with three letters sharing no common informative feature. Initially, response times were nearly equal for the ANV and AOF conditions. However, the ANV condition showed a greater decrease in latency with practice than did the AOF condition, whereas the E condition was fastest throughout. The effects of conditions and the interaction of practice with conditions were significant ($p < .001$) and approximately equal for all age groups.

The difference between the two three-target conditions was quite small, perhaps because subjects did not reach asymptote. Therefore, a subject was given 34 days of practice on all three conditions. The data showed an asymptotic trend and, by the 2€th day, response time on the ANV condition had dropped to the level of the E condition, while the AOF condition continued to require 30 msec. more.

To determine whether verbal awareness aids economical processing, 74 college sophomores were given a letter cancellation task with the same letters. One group cancelled A, O, F, and the other two groups cancelled A, N, V. One of the latter groups was told that A, N, V contained a common aspect found only in the letters to be cancelled. These special instructions did not improve performance. Again, both ANV groups performed more rapidly than did the AOF group ($p < .05$).

A final experiment using a transfer design was performed to rule out the possibility that differences in response time were due to differences in the difficulty of particular feature tests and not to the number of tests carried out. Three groups of 33 college students were given identical pre- and post-tests on a classification task. Following pre-test, 200 training trials allowed one group to process a single feature. This same feature, present in different letters, was also informative for all groups during the pre- and post-tests, so that this group would test for this feature throughout the experiment. A second group had training trials in which a different single feature was informative, while, for the third group, no single feature was useful. The first group improved to a greater degree from pre- to post-test than did the other two, nearly identical, groups ($p < .01$). These findings demonstrate that perceptual processing is malleable; it changes as a function of the task and moves toward the use of the most economical feature list useful for the task. The relevance of these results to the theories of perception and perceptual learning proposed by Gibson, Neisser and Hochberg was discussed.

Abstract of

THE USE OF REDUNDANCY AND OF DISTINCTIVE FEATURES IN THE IDENTIFICATION OF VISUALLY-PRESENTED WORDS

by Frank Smith

Part of the lore of visual perception is that words may be read more easily and accurately than letters in isolation. For 70 years evidence has been cited that more letters can be reported from a brief tachistoscopic exposure if they are organized into words, and that words may be read at a distance from which individual letters cannot be discriminated.

These results have never been satisfactorily explained. "Whole word" theories that propose that words are unitary perceptual elements cannot account for the sensitivity of readers to detail within words and for the fact that some nonword sequences of letters may be easier to identify than some words. "Letter by letter" theories are inadequate because readers typically do not identify every letter in a sequence. Both points of view are weak because they fail to account for the recognition of the basic form (whether the whole word or a single letter).

The present study attempts to account for the visual recognition of a word by proposing that two kinds of "partial information" available within the stimulus may be utilized concurrently. Partial information is defined as information that reduces the uncertainty of a letter or sequence without determining the letter or sequence uniquely. Two types of partial information are considered, involving (1) statistical and (2) structural properties of letter sequences.

The first type, redundancy, refers to sequential constraints among letters; the second type, distinctive features, to discriminable elements of letters common to more than one but not to all letters in the alphabet. The experimental model proposes that the two forms of partial information can be utilized at all stages in the process of word recognition. It implies that it is not necessary for any of the letters in a sequence to be identified uniquely before statistical properties of the sequence as a whole can be exploited.

In Experiment 1, 10 observers reported what they saw as three-letter sequences, projected at below contrast threshold, were brought gradually to a level of relatively easy identifiability. There was no tendency for words to be recognized as "wholes" (all three letters identified simultaneously) more often than nonwords, nor was there any indication that words were identified "letter by letter". Letters were identified at lower contrast levels in words that comprised high sequential constraints and least easily in nonwords with minimal sequential constraints. The first two letters reported correctly from high redundancy nonwords tended to be identified more easily than the first two letters reported from low redundancy words, indicating a responsiveness to sequential constraints among letters at the earliest stages of recognition. There was a marked improvement in performance with words, however, when the final letter came to be identified, suggesting that sequential redundancy was not the sole determinant of perceptual choice. Given some information about the structure of a sequence, knowledge of spelling patterns of words could be employed rather than knowledge of the sequential probabilities of letters. The majority of errors occurred when the third letter to be reported from a sequence was being identified.

The 998 errors of letter identification made in the process of identifying all sequences were analyzed. An average of fewer than four confusion types per stimulus letter accounted for two thirds of the letter errors. The conclusion is drawn that observers did not make errors haphazardly, but that most errors could be placed at the intersection of those letters that the sequential constraints or spelling patterns of English would predict and the set of alternatives left open after discrimination of a limited number of distinctive features.

There is evidence that both distinctive features and redundancy were employed at the beginning of the word recognition process as well as at the end. In particular, the level at which observers made their first correct identification of a letter in a sequence was correlated with the difficulty of the sequence as a whole. This result suggests that a parallel process of analysis and integration began as soon as features from all parts of the sequence were distinguished, taking into account probable constraints among the alternative letter sequences specified by the discriminated features. As a consequence, the first letter to be reported may be identified earlier (i.e. on less featural information) and more accurately if the sequence as a whole is high in redundancy.

The experimental model implies that incorrect responses should share distinctive features with the letters actually presented, and that this shared set would in fact be the only features the observer is able to discriminate. It follows that observers might not be able to distinguish between their incorrect responses and the original stimulus if the two are presented side by side at the contrast level at which the error was made. This possibility was examined in Experiment 2, and supported. One observer's judgments were essentially random, indicating minimal ability in discriminating between stimulus and incorrect response; the other observer was successful on no more than 50 per cent of her judgments.

A conclusion from Experiment 1 is that the more difficult a sequence as a whole to identify, the more difficult it might be to identify every individual letter in that sequence. The applicability of this conclusion to words of up to eight letters in length was considered in Experiment 3. Six observers provided data in conditions analogous to those of Experiment 1 except that a tachistoscopic exposure was also employed. Most of the results show a clear separation between comparable letter identification levels for words of similar length. It also appears that the identification of the first letters to be reported from longer words is not necessarily made later than the first identifications in short words. This result suggests that higher thresholds for longer words may in part be attributable to a protraction of the total recognition process rather than to an absolute difference in difficulty among words of different lengths.

Eye-fixation Patterning and Grammatical Structure

Stanley Wanat
(Abstract of Ph.D Thesis Research)

This study attempts to describe the effect of different grammatical structures on the language-user's processing strategies in reading. The behavior we have chosen is eye-fixation patterning in reading, as a reflection of the reader's search behaviors for information.

A recent study of eye-movements in reading by Mehler, Bever, and Carey (1967) showed that "surface phrase structure differences are reflected by different eye-fixation patterns." This study uses stimulus materials in which the "surface phrase structure" is not varied, but in which underlying linguistic constraints are manipulated.

Two sets of stimulus materials were used. Group I contained active sentences and two types of passive sentences. With these sentence types, we wished to see if the reader's visual search for information would parallel the findings in Levin and Kaplan's eye-voice span study of active and passive sentences, and Wanat and Levin's eye-voice span study of the two types of passive sentences. The second set of sentences, Group II, was composed of right-embedded and left-embedded sentences. This part of the study was based on the eye-voice span study of Levin, Kaplan, Grossman and Yang. The two types of passive sentences were generated by using the same sentence and changing only one lexical item to reflect a change in the underlying constraints. Also, any particular right (or left) embedded sentence was generated by permuting the phrases in its left (or right) counterpart.

Method

<u>Subjects</u> Twelve staff members of the Department of Neuro-psychology at the Stanford Medical Center served as subjects.

<u>Apparatus</u> The eye-movement recorder employed in this study was developed by Dr. Norman Mackworth of the Stanford Medical Center. Records of eye-movements were made by a motion-picture camera, operating at five frames per second. The apparatus incorporated a luminous scale grid around the stimulus material, and an optical system that photographically superimposed the image of the iris and pupil on the image of the stimulus material. Each frame of the film was examined to locate the center of the pupil image, and from that, the area of fixation was determined by reference to the image of the scale grid surrounding the stimulus materials.

Procedure Each subject read all of the stimulus sentences. The sentences were presented one at a time. There were two sessions, one week apart, for each subject. Each subject read 8 sentences from each of the sentence types, except that he read 16 active sentences. The sentences were equally divided between the two sessions, and each subject read half of the sentences orally, and half silently. Sentences from the various types were randomly ordered.

Measures Analysis of the data yielded three classes of measures:

A. 1. number of fixations (not counting regress ions)
   2. number of regressions

B. 1. time spent in fixating (not counting regressions)
   2. time spent in fixating during regressions

C. 1. mean time per fixation (not counting regressions)
   2. mean time per regression

These three classes of measures were determined for each sentence type. A more detailed analysis, by positions within each sentence type, is now being done.

Results

For oral readings of Group I sentences (active and the different types of passives), differences in the number of regressions was highly significant ($P < .001$). The number of regressions, for 12 subjects, 8 sentences per subject, were:

PASSIVE, with BY, with LOCATIVE	42
PASSIVE, without BY, with LOCATIVE	23
PASSIVE, with BY, with AGENT	21
ACTIVE	14

For oral readings of these same sentence types, differences in the time spent in regressions was highly significant ($P < .001$). The times, for 12 subjects, 8 sentences per subject, were:

PASSIVE, with BY, with LOCATIVE	75
PASSIVE, without BY, with LOCATIVE	36
PASSIVE, with BY, with AGENT	31
ACTIVE,	18

These results indicate that the reader's search for information, as indicated by his eye-fixation patterning, is related to the linguistic constraints holding between the constituent parts of the sentence.

Some Effects of Intonation and Pause on Sentence Processing

Boyce L. Ford
(Abstract of Ph.D. Thesis Research)

Two experiments were completed in the investigation of pause and intonation as an aid to sentence processing. One experiment included both intonation and pause as possible facilitating cues and the other included only pause.

The task was a simple repetition of an orally presented sentence. The time to begin the repetition was the dependent variable. Only monosyllabic words were used. The structure of the nine word sentences was determiner, noun, preposition, determiner, noun, verb, preposition, determiner, and noun. The six word sentences had the same structure but with the first prepositional phrase deleted. Subjects were aged 5, 7, 9 and 11.

Intonation was varied by presenting a sentence in either a monotone or with normal intonation. Pause was varied by placing a one second pause at either a phrase boundary or within a phrase. It was expected that both intonation and pauses at phrase boundary would facilitate sentence processing and hence decrease the amount of time necessary to initiate a response.

This prediction resulted from the assumption that a grammatical phrase would function as a unit. Fractionation resulting from either removing the cues possibly inherent in intonation, or by placing an artificial pause within a phrase would upset the integrity of this unit and interfere with efficient sentence processing.

The experimental results only partially support this prediction. Intonation did facilitate sentence processing, but a pause at a phrase boundary interfered with processing, i.e. resulted in longer latencies. This last result was not expected and was at first thought to be spurious, but was replicated in four different groups in two different experiments using different subjects and sentences. These replications would seem to validate the finding.

The apparent facilitating effect that breaking up grammatical units has on sentence processing appears to be a function of the information contained in the input unit. If the input is a series of discrete grammatical units, each of these units must become integrated both grammatically and semantically. But if chunks can be presented that contain information from two or more of the discrete grammatical units, part of the integration necessary with discrete grammatical unit input is no longer necessary; i.e., the input stimulus has already been integratd when it is presented. If one component of sentence processing is integration of discrete units decoded over time, then it would not be surprising to find faster processing of material when more than one unit is presented in a unit of time.