EYE MOVEMENT, PERCEPTION, AND LEGIBILITY IN READING

BY MILES A. TINKER
University of Minnesota

This review is a continuation of the series previously published in this journal (69, 70, 71, 73). A few early papers which were missed in the previous reviews are included in the list of references (31, 32, 33, 34, 51, 55, 57, 79, 86). No publications appearing later than October, 1935, are included.

Bibliographies, popular discussions, summaries of part of the field, and textbook surveys will be found in references 12, 28, 35, 36, 37, 45, 46, 49, 50a, 66. New methods and new devices, or modifications of previous apparatus constructed for use in reading investigations are described in references 2, 6, 7, 16, 17, 18, 60, 68, 72, 78, 83.

EYE MOVEMENTS

Recent investigations of eye movements in reading have involved for the most part the measurement of oculomotor behavior in special reading and learning situations, determination of validity and reliability of the measures, and evaluation of the importance of eye movements in diagnostic and remedial reading. A few experiments have been concerned with establishing norms, the use of eye movements as measures of legibility, and the relation of eye movements to other measures of performance.

Studies of eye movements in the reading of oriental and foreign languages continue to appear.

Futch (27) reports on a comparison of the various measures of eye movements in reading Latin and English by beginning and experienced students. Both groups of subjects exhibited less efficiency in reading Latin than in reading English. Although the beginning and the advanced Latin students were affected similarly in going from English to Latin, the performance of the beginners was most affected. The closest approximation to reading English was in the pause duration on the Latin. Number of fixations was in second place while regressions showed greatest differences. The records indicated that marked improvement in reading Latin was possible, especially in regularity of fixations along the line of print and in perception time; nevertheless
the advanced students were still on a word-recognition level. Practically every word was fixated. Some of the very poor students tended to scan the Latin with relatively few fixations and with no attempt to read the material. Increased skill brought a gradual increase in fixations until a certain analytical mastery was achieved. Then with further improvement there followed a decline in pause frequency until there was an approach to the standards of reading in one’s own language. The changed word order in Latin hindered beginning but not experienced students. Comparison of teaching methods revealed that emphasis on grammar and translation without abundant reading of easy material hindered the development of efficient oculomotor behavior. These results suggest important practical applications in teaching Latin.

A new study in the reading of Chinese is reported by Wang (82). Vernacular was read with fewer fixations and regressions, and at a faster rate than literary Chinese (from which the vernacular was translated). Pause duration showed little change from one to the other. Normal reading required more fixations than rapid reading, and careful reading more fixations than normal. Progress of eye fixations along the lines of print was similar to that in reading English. Compactness of the Chinese printing (square shape of words), however, led to smaller spacial distance between fixations than for the English. More words were read per fixation in the Chinese. Since only a few subjects were photographed and because it is practically impossible to equate the amount of practice in English and Chinese for any individual, some aspects of the quantitative comparisons of eye movements in the two languages must necessarily have doubtful validity. In another study of Chinese, Hu (43) found that silent reading was always more efficient than oral reading. More and longer fixations were employed in reading poetry than for a novel or for classical prose.

Oculomotor patterns vary considerably from one situation to another. Ewert (24) discovered that in comparing eye movements during non-activity, reading, and recall after learning, the characteristic eye movements which reading necessitates perseverate in frequency during recall of material previously read and learned. He suggests that these eye movements accompanying the verbal recall are themselves a part of the recall. These interesting results should be checked by photographing the eye movements and plotting the location of the fixations on the original reading copy.

In other special types of studies Antibon (8) made analyses which indicate that eye-movement patterns during writing are somewhat analogous to those in reading, and Ovio (52) points out that there is a certain amount of disharmony of the eyes in reading. In a more complete study, Clark (16a) has compared eye movements of individuals having normal binocular balance with those of an exophoric (muscular imbalance) group. No differences between groups were found in either fixation frequency, regression frequency, pause duration, or time required for complete divergence movements at the beginning
of the lines of print. The lack of a difference here may be partly explained
by the fact that Clark matched his groups on the basis of reading scores,
and eye movements are measures of reading ability. The author did find
that the exophoric group made greater divergent movements at the beginning
of the lines. It was suggested that these divergent movements are a factor in
ocular fatigue. On page 534, the author evidently meant that the size of
the divergent movements was 9.1 minutes larger for the experimental group,
rather than for the control group as stated.

In a comprehensive study, Gilbert (30) compared the oculomotor
and perceptual habits of good and poor spellers from Grade III
through the senior high school as they learned to spell words. At
every grade level the good spellers employed fewer fixations, a
shorter study time, briefer and less variable pauses, a broader study
unit, and fewer regressions than the poor spellers. In studying
spelling the subject's eyes progressed from left to right with some
regressions. Successive cross-passages over the whole word were
made in this manner until the word was learned. The mature stu-
dent is more systematic than the inefficient student in his eye-move-
ment sequences. Since there is a strong possibility of sampling
errors in the groups of only 3 or 4 subjects used, the growth curves
of eye-movements presented by Gilbert cannot be accepted as valid.
True growth curves can be obtained only when the same group of
individuals are measured on successive years.

A quantitative and qualitative analysis of eye movements in read-
ing prose and objective examination questions differing in type and
by measured amounts of difficulty is reported by Frandsen (25). In
general, perception time, fixation frequency, and regression frequency
increased in the order: true-false, completion, multiple choice, anal-
ogy. In line with the findings in many other investigations, there
was very little variation in pause duration.

Although there was a fair relationship between the eye-movement measures
for the various types of questions, there was little correspondence in per-
formance from prose to the questions, or from one level of difficulty to
another in the questions. Characteristic differences in oculomotor patterns
occurred with the various types of questions. These patterns consisted of
analytical sequences of movements and a typical procedure which was adapted
to the type of question being read. Analysis of these patterns indicated that
the oculomotor processes were subordinated and flexible in their responses
to the central processes of apprehension and assimilation.

The eye movements of good readers have been studied by Walker
(80, 81). The mean pause duration was .255 sec.; there were 2.92
ems per fixation (25 ems per line), and .70 regressions per line.
Increase in difficulty of the reading material brought an increase in pause duration and a decrease in the size of the fixation field, but the extent of the *forward* shift of the eye movements was little affected. Walker also discovered that a relatively large amount (10%) of the reading time was taken by the moves. This amount is larger than any previous investigator has found and needs checking, preferably in a study employing reading material varying in type and difficulty. Lack of vertical movements and marked binocular coordination in the records suggests that monocular horizontal movements are sufficient for studying the performance of good readers.

In comparing the eye-movement records of successful and unsuccessful college students matched for intelligence, sex, and major subjects, Eurich (22) discovered no marked or statistically significant differences. The successful students revealed slightly shorter pause duration, a slightly greater pause frequency and a slightly smaller regression frequency. Successful students read slightly faster and comprehend slightly better than unsuccessful ones of the same intelligence. Litterer (47) found for college students only a moderate relationship between visual apprehension and eye-movement records \((r = .47\) to \(.63\)). Fourth and fifth grade norms for eye-movement measures have been established by Eurich (23). The variabilities presented with the medians enable one to compare individual cases with these standards.

Black print on a white background is read faster than white print on a black background. Taylor's investigation (67) showed that this difference in favor of black printing is due to a smaller fixation frequency. Pause duration in the 2 forms of printing was the same.

Workers in the field of eye movements have only recently become interested in knowing the reliability and validity of their measures. Litterer (47), Futch (26), Eurich (20, 21) and Frandsen (25) have all found high reliabilities for eye-movement measures. Most of these coefficients fall between .75 and .90. When determined by this method (split-half or odd-even) eye-movement scores have high internal consistency. Test-retest reliability coefficients will be published shortly by the reviewer.

Pause duration, fixation frequency and regression frequency have all been used without discrimination as measures of reading speed or reading proficiency. The implication has been that the measures mean the same or nearly the same thing. Since perception time (sum of pauses) is a good practical criterion of speed in reading, the other photographic measures may be compared with it. Using relatively large groups of subjects who read several kinds of material, Tinker and Frandsen (76) found a high correlation between fixation frequency and perception time, a fair correlation between regression
frequency and perception time, and a highly variable correlation between pause duration and perception time. It was concluded that fixation frequency was a good measure of speed of reading, regression frequency a fair one, and pause duration a poor one. These findings are in substantial agreement with preliminary findings cited by Tinker (71, ref. 94) and with later results reported by Eurich (21) and Futch (26).

Robinson and Murphy (61), in checking the accuracy of eye-movement records obtained by direct observation against photographic records, found that a large number of the movements were missed by the experimenter doing the counting. Group trends are revealed fairly well, however, by the direct counting method.

Attempts have been made to validate eye-movement measures by correlating them with achievement on standardized reading tests. Litterer (48), employing speed of reading and comprehension tests as criteria, obtained raw correlation coefficients for adults between test scores and eye-movement measures ranging from $-.27$ to $-.51$. Comparable results cited by Eurich ranged from $-.02$ to $-.24$ for adults (20), and from $-.11$ to $-.55$ for children (21); by Futch (26), $+.06$ to $-.63$ for children. It is clear from these results that validity with achievement test scores as criteria is low. In these experiments the material read before the camera has either been entirely different from that in the reading tests or has been read with a different reading attitude. Data in the field of reading indicate that there are many reading skills which are somewhat independent. When the reading attitude changes from one test to another, or when the material in one test is not strictly comparable to that in the other, the highest correlation found between test scores is approximately $.60$, but the coefficient may be anywhere between $.00$ and $.60$. Since coefficients between $.25$ and $.45$ are quite representative of the correlation, the validity coefficients cited above may indicate adequate validity for the eye-movement measures. At the best, however, this is an unsatisfactory method of determining validity. Probably the most adequate way of computing validity of eye-movement measures of reading ability is to employ as criteria the scores on material which is strictly comparable to that read before the camera. In a study not yet published the reviewer found validity coefficients computed in this way to be uniformly very high ($-.87$ to $-.99$) except for pause duration ($-.10$ to $-.31$).
Eye-movement records have been employed to study individual cases with reading and speech difficulties.

According to Jasper and Murray (44) the eye movements of stutterers during oral reading show marked irregularities. These readers revealed marked tremor in adjusting the eye to fixations in reading, and relatively marked vertical eye movements. There was some evidence (1) that the dominant eye came to the correct accommodation and adjustment in the fixation pause earlier than the non-dominant eye, and (2) that there were "blocks" in the eye movements similar in nature to those that exist in the movements of speech-muscles for these subjects.

In a comprehensive study of 21 retarded readers in comparison with a control group, Robinson (58, 59) has attempted to evaluate the rôle of eye-movement habits in determining reading ability. Training for comprehension had no effect on the eye movements. However, the training for rate which included eye-movement "pacing" produced more words read per fixation and fewer regressions. Robinson states that the "pacing" improved eye-movement habits, which in turn increased reading efficiency. This conclusion may be questioned. There is much evidence that eye movements, for the most part, are merely symptoms rather than causes of reading proficiency.

Gates and Bennett (29) and Gates (28) have emphasized the apparent rôle of eye movements in reading disability. They recommend that the eye be trained to follow a finger moved along the line of print, or a moving spot of light projected along the line by a motion picture machine. These highly formal techniques for training eye movements are in most instances of questionable value and may frequently become a liability. A careful consideration of the part played by eye movements in reading suggests that Gates and Bennett have made questionable inferences concerning the rôle of eye movements in enhancing tendencies to reversals in perceiving words.

Two critical papers on eye movements are reported by Tinker: (1) an evaluation of the rôle of eye movements in diagnostic and remedial reading (74) in which he concludes that measurement and "pacing" of eye movements may be dispensed with in the reading clinic without lessening the effectiveness of diagnosis or remedial instruction; and (2) an evaluation of the use and limitations of eye-movement measures of reading (75).

**Visual Apprehension and Perception**

Study of visual apprehension and perception constitutes only a minor portion of the recent experimental work in reading. A few reports have appeared on span of apprehension which, as shown in previous papers in this series of reviews, appears to be related to perception in reading. A series of investigations in this field is reported by Anderson and Crosland. A preliminary study (2) revealed no
advantage in reporting seen letters in a left-to-right order over a right-to-left, or over a center-alternate order. In considering the effects of eye dominance on apprehension of letter series, they (3) found that left-eyed subjects excel right-eyed subjects at the right side of an exposure-card of 9 letters, and that right-eyed subjects were superior to left-eyed subjects for the leftmost letters of the series. This superiority was relatively greater for the left-eyed subjects. Place of eye-fixation on the series had no effect on the results. In a second study (4), there was a rather prominent tendency for left-handed subjects to apprehend letters near the middle of the letter-series more accurately than right-handed subjects. When the effects of eyedness and handedness were combined (5), definite advantage in apprehension of letters within specific parts of the letter series occurred. Considering all 4 studies, the left-eyed and left-handed subjects were superior to those who were right-eyed and right-handed. It is possible that further analysis and discussion will reveal a significant bearing of these findings upon the reading process.

Chapman and Brown (13) report that the enhancement of clearness of certain letters in a series by variations in size, color or position is accompanied by a reliable decrease in the span of apprehension. This is an agreement with the "law of reciprocity." Another determinant of visual apprehension is reported by Dahmen (19) who found that recresal tended to reduce errors in the reports of his subjects.

Confirmation of previous findings on tendencies toward grouping in visual apprehension of letters, toward transposition of letters, inability to report all symbols seen clearly, and the like, is found in Aveling's report (9). Letterer (48) has attempted to determine the relationship between visual apprehension scores for meaningful material and reading, but practically none between nonsense series of letters and reading. All apprehension scores, however, were significantly associated with intelligence.

Chou (15), continuing analyses reported earlier, emphasizes what he calls the Gestalt qualities involved in the perception of Chinese characters. In contrast to this, Hollingworth (42) states that it was demonstrated long ago that words do not function as "Gestalten." The fact that the shape of a word, or seeing certain letters of a word leads to perception of the word is considered to be only the ordinary law of perception. Partial stimuli function redintegratively; incomplete Gestalten do not exhibit "closure." The term "Gestalt" as applied to words is considered to be only a smoke screen. Most will agree that "Gestalt" in the above sense should not be applied to words. Word-form is a more descriptive and more adequate term
to employ when characterizing the effect of shape of words on their perception.

Further evidence that the beginning of the word plays a dominant rôle in word perception is given by Grund (40). Data collected from the tachistoscopic reading of school children demonstrated that initial letters of German nouns are apprehended much more readily when they form part of a word than when they are alone or in a nonsense series of letters. The effect upon perception of these letters at the beginning of a word is not one of simple dominance; the initial letters are merely important determinants in the perception of the word as a complex unit.

Employing a modified tachistoscope, Preston (56) discovered that familiarity was an important factor in speed of word perception. Familiar words are perceived and named at a much faster rate than unfamiliar words of the same length. Speed for familiar words, however, was highly diagnostic of speed for unfamiliar words. That is, the speed factor is common in both situations. Speed of single word perception was found to correlate fairly well with speed of reading and vocabulary knowledge but only to a slight degree with reading comprehension.

Legibility of Print

The Chinese continue to manifest an interest in the legibility of their own language. Using large groups of Chinese school children as subjects, Ai (1) compared the relative efficiency of quick perception of materials printed in vertical and horizontal arrangements. In general, at Grade VI, at the junior high school level, and in the senior high school, vertical printing was read faster and more accurately than material in the horizontal arrangement. Although these results are consistently in favor of the vertical alignment, one should know how much experience the readers have had with each arrangement before coming to definite conclusions regarding the relative legibility of the 2 ways of printing. Chou (14) presents further analyses of data already reported to show that the total Gestalt-qualities of a Chinese character determine to a large degree its legibility.

Paterson and Tinker have continued their series of studies on typographical factors influencing speed of reading. A comparison (53) of material in 10 point, 19 pica line length, Scotch Roman, set with 1, 2, and 4 point leading or space between lines, with material in 10 point type, and a 19 pica line length, set solid, revealed that: (1) 1 point leading does not facilitate speed of reading as compared with text set solid; (2) 2 point leading increased speed of reading by
7.5%, while 4 point leading increased speed of reading by only 5.0% in comparison with the material set solid. The advantage of 2 point leading over both smaller and larger amounts of space between lines may hold only for the particular size of type and length of line used in this investigation. Further studies should undertake to determine, therefore, the effects of leading with larger and smaller type sizes and for various lengths of line. Such studies are under way.

In study number IX of their series, Tinker and Paterson (77) determined the effect that reduction in size of newspaper print had upon speed of reading. The problem was raised by the need of preserving newspaper accounts of historical events for historical scholars and libraries. Such reductions may be made by photographic, by the planographic offset-printing process, or by some similar technique. When newsprint is reduced to 80% of normal only a slight and insignificant reduction in speed of reading occurs. If the reduction is 50% of normal, reading is slowed by about 10%. When the print is reduced to 30% of normal size, however, speed of reading is slowed up by almost 75%. From a practical viewpoint, such as the need for preserving historical documents, one may conclude that newsprint may be reduced to about half-size without undue loss of legibility.

In the final study, Paterson and Tinker (54) investigated the relative legibility of 10 different styles of type face. No important differences in rate of reading text material were found for the 6 following type faces: Scotch Roman, Garamond, Antique, Bodoni, Old Style, Caslon O. S., Kabel lite, and Cheltenham. Two type faces retarded speed of reading: American Typewriter type by 5.1%, and Cloister Black type (Old English) by 16.5%. They concluded that type faces in common use are equally legible.

The relative legibility of 10 type faces (the same as those used by Paterson and Tinker) was obtained by Webster and Tinker (84). The legibility was determined by measuring the distance from the eyes at which single words could be read correctly. Legibility measured in this manner differed markedly from legibility measured in terms of speed of reading. While word-form is apparently an important determinant of speed of reading, size, heaviness, and width of printing stroke are more important when perceiving words at a distance. Some factors which increase perceptibility of words at a distance reduce speed of reading. By the distance method, the order from most to least legible was: American Typewriter, Cheltenham,
Antique, Old Style, Caslon O. S., Garamond, Bodoni, Scotch Roman, Kabel lite, Cloister Black. Many of the obtained differences in perceptibility were highly stable. In another report (64), a way to recognize type faces is described by Stanley.

Greene has conducted 2 experiments on the legibility of typewritten material. In the first study (38) which involved lithoprinting reproduction of typewriting, the results indicated: (1) there was no difference in speed and accuracy of reading 7, 10, 12, and 14 point type; (2) the material with leading, especially the longer lines, was read somewhat faster (3.5%) than text set solid; and (3) text in the very long line was read slightly slower than in the short line.

In a second study (39) Greene discovered no reliable differences in speed of reading text in Ionic linotype and in typewritten samples. This is not in line with the results cited above from Paterson and Tinker's investigation. Examination of the printed samples in each study reveals that: (1) in Paterson and Tinker's experiment the typewritten material occupied about 33% more lines than the Scotch Roman (ordinary print); (2) in Greene's study, however, the typed material covered approximately the same number of lines as the Ionic. This difference in the size of the typewriter type undoubtedly explains why Paterson and Tinker found that such material was read slower than ordinary type, and Greene found no difference between it and ordinary type (Ionic). In an earlier report (71, ref. 89), the reviewer has demonstrated that compactness of printing is an important determinant of reading speed.

Working with first-grade children Buckingham (10) found that 12 point type was read consistently faster than 14, 18 or 24 point type. These latter 3 sizes were all read at approximately the same rate. This advantage was found to hold for 12 point type with 4 point leading but not with 3 point leading. Leading did not seem to help the readability of the larger type sizes except there was a slight advantage for 6 point leading with 24 point type. Since the advantage for the 12 point type with 4 point leading is only 9 to 13 seconds on the average, and the ratio of the difference to the standard error of the difference is only 2.0, the stability of the difference is not high nor is the difference large. Actually one might have concluded that any one of the printing arrangements was as good as any other. Furthermore, if smaller sizes of types had been used they might have been as efficient as the larger sizes. A hint from other lines of evidence tends to show that variation in typography within rather large limits may have little or no effect upon the reading of children in the lower grades. Further research is needed to settle the question.
An investigation by Webster and Tinker (85) of the influence of egg-shell, artisan enamel, and flint enamel paper surfaces upon the perceptibility of words failed to disclose any differential effect during short working periods. Legibility was measured in terms of distance from the eye at which words printed on these papers could be read accurately. The alleged eye-strain from reading material printed on glazed paper must be due to continuous reading in light not uniformly dispersed rather than to inability to see the print satisfactorily in good illumination.

Facts involved in legibility of posters and signs have received some notice. Sumner (65), examining 42 combinations of De Voe's showcard colors, substantiated earlier findings that legibility depends upon the brightness difference between print and its background rather than upon the colors as such. Dark letters upon a light background were more legible than light letters upon a dark background. Severny (63) found that the rectangle and the triangle were the most readily perceived signs employed as road signals. Next came the circle and the square. Rocard (62) found that the visibility of objects lighted by a projector depends mainly upon contrast effect and size.

According to Burkersrode and Burkhardt (11) capitalization of German words reduces legibility.

In an extensive investigation on the legibility of black and white print, Taylor (67) found that in every comparison made, black print on a white background was more legible than white on black.

This was true whether legibility was measured in terms of span of visual apprehension, speed of reading, recognizability in peripheral vision, or the greatest distance at which the print could be read. In very small type these differences in legibility are increased. A serifless style of type was equally legible except in small sizes where the black print is more legible. For nonsense material the white print is relatively more inferior than in sense material. Analyses indicate that the inferiority of the white print is due to effects of irradiation. When black print is used the detrimental effects of irradiation are counteracted by the addition of serifs which tend to preserve the characteristic forms of the letters by emphasizing their corners. With white letters, irradiation will increase the apparent size of the letters, but it will also tend to blur their outlines, close their open spaces, and fuse their parts. This will decrease their legibility. These results will be most marked when the letters are small, and when there are no auxiliary cues to perception as in nonsense material. They should be least effective in the larger sizes of serifless type. The application of these findings to printing practice is obvious.

The examination of large samples of handwriting for legibility by Newland (50) revealed that the illegibilities of only 4 letters (a, e, r, t) contributed 45%, 46% and 47% to the totals for the elementary, high school, and adult groups respectively. Forms of illegibility tended to be characteristic of each age group rather than common to all. For the most part illegibilities were due to (1) the failure to
close letters, (2) closing looped strokes, (3) looping non-looped strokes, and (4) using straight up-strokes rather than rounded strokes. It seems possible, therefore, that preventive and corrective work in handwriting can be both definite and concentrated to the few places where the illegibilities occur.

**Summary**

The experimental study of eye movements, perception, and legibility in reading during the past 3 years is characterized by a marked expansion in eye-movement studies. Much new apparatus has been devised for measuring eye movements, workers have become interested in establishing the reliability and validity of these measures, and most of the investigation has been confined to special reading situations or to diagnosing and remedying disabilities. Few advances have been made in studying perception in reading. Most of the legibility experiments have been concerned with discovering the relative legibility of typographical arrangements encountered in the consecutive reading of children and adults.

**BIBLIOGRAPHY**


18. CROSLAND, H. R., A Technique to Control, and to Measure the Effects of, Fixation in the 'Range of Attention' Experiment. *J. Exper. Psychol.*, 1933, 16, 446-450.


41. Holland, B. F., "Special Apparatus in the Laboratory of Educational Psychology, University of Texas." *Amer. J. Psychol.*, 1933, 45, 139–145.


54. **PATERSON, D. G., and TINKER, M. A.** Studies of Typographical Factors Influencing Speed of Reading. X. Style of Type Face. *J. Appl. Psychol.*, 1932, 16, 605-613.


64. **STANLEY, T. B.,** How to Recognize Type Faces. *Printers' Ink.*, 1932, 25, 70-72.


