



Universität Potsdam

Reinhold Kliegl, Richard K. Olsen, Brian J. Davidson

Regression analyses as a tool for studying
reading processes: Comment on Just and
Carpenter's eye fixation theory

first published in:

Memory & Cognition. - 10 (1982), 3, p. 287 - 296

ISSN: 0090-502x

Postprint published at the Institutional Repository of the
Potsdam University:

In: Postprints der Universität Potsdam : Humanwissenschaftliche Reihe ; 42

<http://opus.kobv.de/ubp/volltexte/2008/1685/>

<http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-16857>

Postprints der Universität Potsdam

Humanwissenschaftliche Reihe ; 42

ISSN 1866-8364

Regression analyses as a tool for studying reading processes: Comment on Just and Carpenter's eye fixation theory

REINHOLD KLI EGL, RICHARD K. OLSON, and BRIAN J. DAVIDSON
University of Colorado, Boulder, Colorado 80309

Just and Carpenter (1980) presented a theory of reading based on eye fixations wherein their "psycholinguistic" variables accounted for 72% of the variance in word gaze durations. This comment raises some statistical and theoretical problems with their use of simultaneous regression analysis of gaze duration measures and with the resulting theory of reading. A major problem was the confounding of perceptual with psycholinguistic factors. New eye fixation data are presented to support these criticisms. Analysis of fixations within words revealed that most gaze duration variance was contributed by number of fixations rather than by fixation duration.

Just and Carpenter (1980) introduced new approaches to the use of eye fixation data. They presented complete texts of about 150 words and related gaze duration on individual words to a variety of theoretically relevant psycholinguistic variables, including number of syllables, word frequency, beginning of line, novel word, case-role assignment, and end of sentence. It was argued that all possible psycholinguistic processes were completed during the gaze ("immediacy assumption"), and only the fixated word was being processed ("eye-mind assumption"). Gaze duration was defined as the total time the eye was aimed within the boundaries of a word, including intraword saccades and blinks. Psycholinguistic variables accounted for an impressive 72% of the word gaze duration variance based on averages for each word across 14 subjects. Regression coefficients for each of the psycholinguistic variables were interpreted to support a production model of reading. We will raise some questions about the approach Just and Carpenter (1980) have taken in their interpretation of gaze durations. We will show that influences attributed to psycholinguistic processes may be confounded with perceptual limitations and that averaging across subjects may magnify this confound.

Most eye-movement research prior to Just and Carpenter (1976, 1980) has analyzed the duration of fixations. Unlike gaze durations, which measure the time spent on a word, fixation durations are determined according to an outside criterion that sums the time at

a location until there is a substantial shift in the location of the fixation (at least two character positions in our studies). It is assumed that the time necessary to shift the eye can be ignored for purposes of linguistic or perceptual processing. Consequently, the units of analysis are fixations. If analyzed with respect to words, a second dependent measure emerges besides fixation duration, namely, the number of fixations falling on a word. Gaze duration measures do not allow this distinction between the duration and number of fixations. Since the time spent in intraword saccades is probably less than 5% of a gaze duration, gaze duration amounts to calculating the sum of the fixation durations. Finally, fixation-level analysis also allows us to take the location of the fixation within a word into account. We will review research that reveals this to be a critical variable.

A primary focus of fixation-level analyses has been on the influence of acuity limitations in eye guidance and on the amount of information available during a single fixation. It has become clear from this work that the length of an eye movement is strongly related to the limits of letter and word identification in the periphery. Psycholinguistic variables such as word frequency might influence the span of processing during a fixation, but the primary limiting factor is the decline in perceptual acuity with increasing distance from the fovea (McConkie & Rayner, 1975; Rayner, 1975; Rayner, McConkie, & Ehrlich, 1978). In Just and Carpenter's (1980) study, no consideration was given to perceptual factors, and this raises problems for their gaze duration-based model of reading. As we will show, the influence of perceptual factors on processing time was potentially concealed in their model by (1) the use of gaze duration rather than fixation duration analysis, (2) the use of simultaneous regression analysis with highly correlated predictors, and (3) a questionable averaging procedure across subjects. The first part of this report will discuss these issues in some detail. The second part demonstrates their rele-

This research was supported by USPHS program Project Grant HDMH11681-01A1, Richard Olson, coinvestigator. It is based on parts of the first author's thesis submitted in partial fulfillment of the requirements for the MA degree at the University of Colorado. We thank Gene V Glass, Janice M. Keenan, and two reviewers for helpful suggestions. Requests for reprints should be sent to Reinhold Kliegl, Department of Psychology, University of Colorado, Boulder, Colorado 80309.

vance by means of an experiment that was similar to Just and Carpenter's.

Perceptual Factors in Gaze Durations

Just and Carpenter (1980) reported a correlation of .67 between word length (coded in number of syllables) and average gaze duration. Syllables were assumed to be psycholinguistic units of processing for encoding and lexical access in their serial production model. Although other research has demonstrated an independent contribution of number of syllables to gaze duration (Pynte, 1974) (although not in a reading situation), a very sizable amount of variance attributed to number of syllables might be better explained by the spatial properties of a word and acuity limitations in eye guidance.

Gaze duration was introduced in an experiment on sentence-picture comparisons (Just & Carpenter, 1976). Gaze duration was an appropriate measure for that experiment, but it is problematic as a basis for a theory of reading. When a word receives one fixation, gaze duration is equivalent to fixation duration. However, some short words are not fixated at all, and long words may be fixated more than once. Moreover, studies by Kliegl, Olson, and Davidson (in press) and O'Regan (1980, 1981) demonstrated that if a long word receives a fixation close to either end, the chances increase for a second fixation near the opposite end. There is little change in the duration of a single fixation between the one- and two-fixation cases. Thus, time spent on the word is nearly a linear function of the number of fixations. O'Regan's experiments and those of Kliegl et al. indicate that the number of fixations a long word receives can be due simply to the "inconvenient" positioning of the first fixation. The second fixation is employed because of acuity limitations in processing the more peripheral letters. Of course, not every two-fixation case has to be due to incorrect positioning. Rare words might receive more than one fixation because of difficulties in lexical access.

The skipping of words raises another problem for the gaze duration measure. Short words (e.g., "the," "in," "a") are often not fixated directly. Just and Carpenter's (1980) gaze duration approach allocates a 0-msec processing time when a word is not directly fixated. However, other research has indicated that short words may be processed in the periphery without direct fixation. The fact that three-letter verbs tend to get skipped less often than the word "the" would even indicate some form of peripheral processing of content (O'Regan, 1979; Rayner, 1979). Whenever a fixation falls on a very short word, the fixation may be shorter than the average reading fixation, but it is nowhere near zero, even if the next word receives its own fixation.

The two special cases discussed above raise problems for the use of gaze duration as a dependent measure in the analysis of word processing time. The gaze duration measure is too coarse to allow discrimination of these special cases, and it may have led to an inflated estimate

of the relation between psycholinguistic processing time and number of syllables in Just and Carpenter's (1980) model. We will explain how we tested this hypothesis on our own data by deleting words from the data base, both when they received no fixation and when they received more than one fixation. Prior to this, however, it is necessary to point out some limitations of regression analysis in the presence of correlated predictors and to discuss the consequences of averaging across subjects.

Simultaneous and Hierarchical Multiple Regression

Just and Carpenter (1980) based their theory of reading on the regression weights obtained in a simultaneous multiple regression. For example, they estimated that gaze duration increased by 52 msec for every syllable in the fixated word. Similarly, gaze duration increased by 53 msec for every log-unit change in word frequency. Just and Carpenter interpreted these estimates within the framework of their "immediacy assumption," which states that gaze duration reflects the processing time of the word currently fixated. Thus, with more syllables in a word, more processing time is required. However, in a simultaneous multiple regression with substantially correlated predictors, it is not possible to isolate variance exclusively associated with a certain predictor. For example, in our own eye fixation data, we find that number of syllables and word frequency correlate at .62. As predictors in a simultaneous regression analysis (reported in the second section), number of syllables and word frequency correlated at .57 and .63, respectively, with gaze duration. However, because of the predictor intercorrelation, it is possible that most of the variance attributed to number of syllables is due to word frequency, or vice versa.

The question of how to proceed with correlated predictors is still under dispute (Darlington, 1968; Kerlinger & Pedhazur, 1973). Agreement, however, seems to emerge that what is important is not only significance of a predictor, but also its strength as measured in percent variance explained (Hays, 1973, p. 668). The unique amount of variance a predictor explains is the variance that it does not share with any of the other predictors. In hierarchical regression, it is the amount of variance a predictor adds in the presence of all the other predictors. In our data, when number of syllables was entered first, it accounted for 32% of the variance in gaze duration. The remaining unique variance associated with word frequency was only 12%. Similarly, when word frequency was entered first, it accounted for 33%, whereas number of syllables fell to 5%. Both number of syllables and word frequency added a significant amount. All these results allow us to conclude is that number of syllables and word frequency are significantly related to gaze duration. The size of the regression coefficient (i.e., the value that is interpreted theoretically in Just and Carpenter's 1980 model), however, is very dependent on the presence or absence of other predictors in the model.

Hierarchical regression analysis allows one to develop a feeling for the stability of regression coefficients under a variety of models, but it does not offer a solution to the problem of correlated predictors. Its value is limited in the case of explanatory regression models (i.e., models that allocate theoretical significance to a predictor) (Kerlinger & Pedhazur, 1973). The power of hierarchical regression lies in the development of predictive regression models, since it allows the identification of variables that are at least potentially useful for a later development of an explanatory model.¹

Averaging Across Subjects

A second statistical issue deals with averaging across subjects. Just and Carpenter (1980) generated the data for their analysis by averaging gaze durations on each word across 14 subjects. As we noted earlier, gaze duration reflects the number of fixations falling on a word, as well as the duration of the fixations. Since there is little change in fixation duration between one- and two-fixation cases (Kliegl et al., in press; O'Regan, 1981; Rayner, 1979), the average gaze duration to a large extent reflects the probability that a given word will be fixated. Kliegl (1981b) developed a simulation model in which fixations were semirandomly superimposed on a text (the distribution of fixations was based on saccade lengths observed during reading). The simulation produced significant word length (and, consequently, word frequency) effects, simply because long words have a higher chance of receiving two fixations and short words have a decreased chance of receiving any fixation. Moreover, averaging across several simulations influenced the word length effect appreciably. The correlation between word length and gaze duration increased from .47 for two to .77 for eight simulations. Averaging across subjects is justified if it reduces error variance and thus allows for more powerful tests of a hypothesis. However, averaging must not affect the size of the parameter to be estimated, which it clearly does in the case of gaze durations. Therefore, if one wants to use average gaze durations, the significance of a regression coefficient would be better tested against a value delivered by a simulation that does not contain any of the theoretical properties. Testing the assumption that the coefficient is significantly different from zero is testing the wrong null hypothesis, since even a random distribution of fixations on a text will produce such effects.

One averaging approach that would not artifactually inflate psycholinguistic effects is based on the calculation of processing time over character positions. McConkie, Hogaboam, Wolverton, Zola, and Lucas (1979) calculated an idealized processing time profile (IPTP) by averaging across subjects. Assuming a rightward movement of the eye, the IPTP for a single subject is created by spreading the time spent in the left fixation evenly across the letters between the two fixations. This approach assumes that it is the region between two

fixations that is attended to during a fixation. Averaging across subjects will create positive and negative deviations from a baseline. These deviations then reflect increased and decreased processing time of the text properties. Since the IPTP gives every letter or space in the text an equal chance, averaging across several "random" readers would yield an IPTP that approximates a straight line with average processing time equal to the average fixation duration, whereas averaging as suggested by Just and Carpenter (1980) would result in a word profile that deviates from the baseline even if readers were to distribute their fixations randomly across the text. In summary, there might be a partly stochastic component to eye guidance that leads to more fixations on longer words. This suggests a potentially nonlinguistic source of Just and Carpenter's word length effect, and this effect is inflated by averaging across subjects. Just and Carpenter's results showed that the multiple correlation is much lower if individual subjects are analyzed. Our own data showed the same result. To reiterate, the difference is due not only to a reduction in error variance, but also to the fact that averaging increases the chance for parameters to be significantly different from zero even if fixations are randomly distributed across the text.

Despite the problems discussed, eye fixation data can provide a valuable resource for experimental investigation of variables relevant in reading. Just and Carpenter's (1980) analysis should be expanded in a direction that allows incorporation of perceptual and eye guidance aspects of the reading process as considered by McConkie et al. (1979), O'Regan (1979, 1980, 1981), and Rayner (1979). The evaluation of eye guidance, however, may require the analysis of individual fixations and the saccades associated with them rather than gaze duration, which is too coarse a measure to capture some of the important dynamic aspects of the reading process. The following section reports a series of analyses on our own data, including simultaneous and hierarchical multiple regression that demonstrates the relevance of the issues raised above.

EXPERIMENTAL EVALUATION

An experiment was conducted that was comparable to that reported by Just and Carpenter (1980). It provides additional data on the relation between psycholinguistic variables and eye fixations, and it demonstrates the limitations and problems in the interpretation of regression analyses of gaze durations considered in the previous comments. The data were fit by a model that was similar to Just and Carpenter's. In these analyses, the sum of the fixations was used as a measure of gaze duration; this measure is almost identical to Just and Carpenter's "gaze," but it excludes time spent in saccades. The parameters of this model were word length in number of syllables, word frequency, beginning and end of line, function or content word, and end

of sentence. In Just and Carpenter's model, there were also parameters for novel word, end of paragraph, and, instead of a single dichotomous function-content parameter, words were classified according to 1 of 11 categories depending on their linguistic role in the sentence. The differences between the studies and their implications do not bear on our major statistical and theoretical points. Therefore, they will be discussed after presentation of the results.

Method

Subjects. Six volunteer subjects, all of whom had a college degree, participated in the experiment.

Materials. Subjects read the beginning of Camus's (1947/1948) novel *The Plague*. The passage described the city of Oran in Algeria, the people living there, their activities, the climate, and the surroundings of the area. The story was nontechnical prose and presented no problems for comprehension. The passage included 1,214 words and was broken up into 11 TV screens of text. Each screen contained 11 lines of text, with approximately 60 characters/line. The first and the last words of every screen were excluded from the analysis. Taking these reductions into account, there were 1,192 words in the data base for the regression analyses.

Apparatus. A PDP-11/03 microprocessor presented the text on a Ball Brothers television monitor. The display consisted of black letters (7 by 9 dot matrix) on a white background. Subjects could direct the onset and offset of a screen of text with two push buttons.

During reading, eye fixations and movements were recorded by an Applied Sciences eye view monitor (Model 1996) similar to the one used by Just and Carpenter (1980). This system uses a TV camera to record the moving eye, with a sampling rate of 60 Hz. The measurement error of the system under our viewing conditions is within .5 deg of visual angle at least 90% of the time and is rarely greater than 1 deg (see Kliegl & Olson, 1981, for details). In this experiment, 1 deg of visual angle corresponded to three characters on the monitor screen. Thus, the typical accuracy of .5 deg translates into plus or minus one character of fixation output position. The unreduced output of the eye tracker (x-y coordinates and pupil diameter) was stored on-line on a RL01 disk of the PDP-11/03 (for technical details about the hardware and software interface between eye tracker and PDP, consult Davidson, 1981).

Procedure. Subjects were comfortably seated in an adjustable dentist's chair. During the experiment, they rested their heads in a goggles frame that was fixed to the viewing aperture. This minimized head movements and provided enough stability to secure accurate measurements of subjects' eye movements.

After reading two practice passages, subjects fixated a 9-point grid, and the eye-tracker output was calibrated to screen-character position. As this calibration was used for the analysis of eye fixations on the whole experimental text, questions with respect to stability over the course of the experiment arise. McConkie (1981) has persuasively argued that because many systems drift over time, the calibration stability should be checked for all calibration points at the end of each session to evaluate the amount of drift. Unfortunately, this was not done for these subjects. Therefore, Kliegl and Olson (1981) tested new subjects' stability of calibration for the time-span and viewing conditions required in this experiment. In these tests, the calibration was consistently as good at the end of the session as it was at the beginning. Additional tests have confirmed these results. Therefore, we are confident of our accuracy estimates over the course of the experiment.

Following the calibration, subjects were told that they were going to read several pages from Camus's novel *The Plague*.

They were told that they would have to answer a difficult questionnaire about the text afterward and, so, should remember as much detail as possible. Further, they were told not to reread the text. Subjects then read the complete text, changing screens with the push buttons. The memory test results indicated that recall of gist and main ideas was complete, but the recall of minute facts and details was poor.

Data reduction and initial analysis. After the experimental session, eye movements were reduced to a fixation format and rescaled to screen dimensions (i.e., character position and line number). Fixations were plotted under corresponding parts of the text. A small number of fixation data were lost for each subject because of loss of eye-tracker delimiters and blinks.

Each word in the text was coded as first or last word on a line, last word in a sentence, number of letters, number of syllables, and word frequency values according to the Kučera and Francis (1967) word norms. Word frequency values were transformed with the method used by Just and Carpenter (1980) and Mitchell and Green (1978). Namely, every value was incremented by one (to avoid logs of zero) and reexpressed as the log ratio of the value for the most frequent word to the present one. Thus, the value for the most frequent word, "the," is zero, and values for other words become larger as frequency declines. Function words (i.e., determiners, prepositions, auxiliary verbs, and connectives) were coded separately from content words (i.e., nouns, verbs, adjectives, adverbs, and personal pronouns).

Eye fixations were matched with the above information about the text. The dependent measures calculated on every word were gaze duration and number of fixations. Gaze duration sums the fixations during the initial reading of the word only. Fixations were summed as long as the eye did not leave the word boundaries during the first encounter between word and eye. Like Just and Carpenter's (1980) study, regressions to previous words were ignored, and the space before a word was allocated to this word. A more detailed description of the programs used for these initial analyses can be found in Kliegl (1981a).

Results

Simultaneous regression on all data. The variables used in the regression analyses were gaze duration (GD), number of syllables (SL), word frequency (FRQ), beginning and end of line (BEL), function-content word (FC), and end of sentence (ES). The correlations between these variables for the complete set of averaged data are displayed below the diagonal in Table 1. The table

Table 1
Correlations Between Variables Used
in the Regression Analyses

	GD	SL	FRQ	BEL	FC	ES
GD	1.00	.34	.34	-.07	.04	.16
SL	.57	1.00	.62	-.05	.08	.00
FRQ	.63	.63	1.00	.01	.19	.16
BEL	-.05	-.05	.01	1.00	-.01	-.06
FC	.10	.09	.20	.04	1.00	.04
ES	.18	.10	.16	-.06	.04	1.00

Note—Correlations below the diagonal are from a complete set of data; those above the diagonal are from a set of data without words that did not receive a fixation. GD = gaze duration, SL = word length in number of syllables, FRQ = transformed word frequency norms, BEL = beginning and end of line, FC = function-content word, ES = end of sentence.

Table 2
Results of Simultaneous and Hierarchical Multiple Regressions

Dependent Variable	Mean	S1	S2	S3	S4	S5	S6
WPM	219	238	278	169	220	218	191
R	.67	.52	.46	.45	.29	.53	.43
R ²	.45	.27	.22	.21	.08	.28	.19
SE	88	147	156	208	177	164	174
N	1192	1165	1163	1124	1169	1102	1153
Percent Incremental Variance Due to:							
Model 1							
SL	32	18	12	15	7	19	11
FRQ	12	6	9	6	1	7	3
FC	0	0	0	0	0	0	0
ES	1	2	0	0	0	0	1
Model 2							
ES	3	5	0	1	0	1	3
FC	4	2	2	1	1	3	1
FRQ	33	17	20	17	5	19	9
SL	5	3	1	2	2	3	2
Model 3							
LL	45	24	20	20	9	28	14
FRQ	3	2	3	2	0	1	1
FC	0	0	0	0	0	0	0
ES	1	2	0	0	0	0	0
Model 4							
ES	3	5	1	1	0	1	3
FC	4	2	2	1	1	3	1
FRQ	33	17	18	17	5	19	9
LL	9	5	2	3	3	7	3

Note—WPM = reading speed in words per minute, R = multiple correlation for Models 1 and 2, R² = percent experimental variance accounted for by Models 1 and 2, SE = standard error of estimate, and N = number of data points (for all Rs, $p < .01$). See text for explanation of models and parameters.

shows substantial correlations between gaze duration, function-content word, number of syllables, and word frequency.

The complete data set when averaged across subjects yielded a multiple correlation (R) of .67 ($R^2 = .45$) and a standard error of 91 msec. (The corresponding values in Just and Carpenter's, 1980, study, with 17 independent variables and averaged across 14 subjects, were $R = .85$, $R^2 = .72$, and $SE = 88$ msec.) The regression coefficients for beginning and end of line and the intercept were not significantly different from zero ($p > .01$).

The upper part of Table 2 (left column) summarizes these results. Table 2 also displays the corresponding values for the individual subjects. The multiple correlations range from .29 to .53, with a mean of .45. The corresponding value of Just and Carpenter's (1980) study was .60. The different numbers of observations (N) subtracted from 1,192 indicate the amount of text material excluded due to loss of eye-tracker delimiters and blinks.

The following hierarchical analyses will provide an estimate of the unique variance accounted for by some of the predictors. Possible reasons for the somewhat

higher multiple correlation in Just and Carpenter's (1980) study will be considered in the discussion.

Hierarchical regressions. Statistical and eye guidance questions raised in the introduction were addressed by analyzing the eye fixation data in several different ways. In addition to the simultaneous regression analysis employed by Just and Carpenter (1980), parameters were estimated in four hierarchical analyses. Two of the analyses first partialled out word length effects and, therefore, tested for the unique variance in eye fixations for higher level linguistic and grammatical aspects of the text. The other two analyses assessed the unique amount of variance word length properties accounted for when the higher level variables had been covaried out.

In the first analysis, the independent variables were entered in the sequence number of syllables, word frequency, beginning and end of line, function-content word, and end of sentence, allowing a test of cognitive variables (word frequency, beginning and end of line, function-content word, end of sentence) independent of number of syllables, which might be considered more perceptual. The second model reversed this order and tested for the effects of number of syllables after controlling for psycholinguistic variables.

The middle part of Table 2 (Model 1) summarizes the results. Entering number of syllables alone accounts for 32% of the variance, and word frequency entered second still produces an R^2 change of 12%. None of the predictors entered after these first two accounts for more than an additional 1%. For the reverse order (Model 2), end of sentence alone accounts for 3%, function-content word for 4%, and beginning and end of line for 1% incremental variance. Word frequency entered fourth causes an R^2 change of 33%, and number of syllables entered last still accounts for an additional 5%.

The results indicate that syllable length of words and word frequency largely overlap in the experimental variance they account for. Whatever variable is entered first will explain most of it. However, we know that frequency explains at least 12% of the variance after number of syllables is partialled out, and number of syllables explains at least 5% after all the others are partialled out. The importance of function-content word, beginning and end of line, and end of sentence is rather low; even if entered first, not much of the variance is accounted for by these variables.

Table 2 also shows the results of the analyses for the individual subjects. Although the values were lower than those for the averaged data, the pattern across the individual subjects was essentially the same. End of sentence was significant under both models only for Subjects 1 and 6. This could be interpreted as an indication of individual differences, and it replicates a finding of Just and Carpenter's (1980) study.

Up to this point, word length was coded in number of syllables to make our analyses comparable to Just and Carpenter's (1980) and to reflect their assumption that reading occurs in syllable-like chunks. From an eye guidance and peripheral acuity perspective, however, number of letters might be a more appropriate basis for word length effects. Note that "a" and "straight" both have one syllable, but differ dramatically in number of letters. Therefore, the analyses were repeated using number of letters in place of number of syllables.

The results of these analyses are given in the lower part of Table 2 (Models 3 and 4). The substitution of number of letters for number of syllables improved the fit of the model somewhat ($R^2 = .49$ vs. $.45$ for Models 1 and 2). The first hierarchical analysis (Model 3) showed quite a different pattern under this condition. Word length now accounted for 45% of the variance, limiting word frequency to only an additional 3% instead of the 12% it accounted for when number of syllables was used. The only difference in the second analysis (Model 4) from Model 3 was the increase from a 5% change if number of syllables was entered last to a 9% change if number of letters was entered last. The analyses are also reported for the individual subjects and again show the same pattern. No improvement of the fit was obtained by entering both number of syllables and number of letters as parameters.

Regressions on subsets of data. A second series of analyses was performed on various subsets of the data to determine the effects of number of fixations on gaze duration. First, words that were fixated more than once were not included (Criterion A). Second, words not fixated were not included (Criterion B). Third, Criterion A and B were applied simultaneously, leaving only words that received one fixation (Criterion C). Finally, the data were analyzed with number of fixations partialled out. Consequently, the averaged data were based only on the words that complied with the respective criteria.

The first analysis under Criterion A rejected all words with more than one fixation. As was pointed out earlier, these words may tend to bias the gaze duration beyond a value that should be considered representative for the psycholinguistic processing of the word. Two-fixation cases may sometimes occur on longer words due to misplacement of the first fixation and acuity limitations. This nearly doubles the gaze duration.

Table 3 displays results for this analysis carried out for every subject and for the average. R , R^2 , number of observations, and what percentage these represent with respect to the analyses reported above are given under Criterion A. The tabled values indicate that, although the data base was only minimally reduced (i.e., about 10%), the drop in R^2 is substantial from .49 to .39).

Table 3
Results of Hierarchical Regression Analyses Under
Different Selection Criteria

	Mean	S1	S2	S3	S4	S5	S6
Criterion A							
R	.63	.45	.39	.33	.23	.48	.26
R^2	.39	.20	.15	.11	.05	.23	.07
N	1111	1103	1046	971	1069	976	1003
%	93	95	90	86	91	89	87
Criterion B							
R	.40	.37	.28	.37	.21	.41	.27
R^2	.16	.14	.08	.13	.04	.16	.08
N	1172	688	793	888	823	745	849
%	98	59	68	79	70	68	74
Criterion C							
R	.23	.19	.13	.15	.12	.26	.09
R^2	.05	.04	.02	.02	.01	.07	.00
N	1039	626	670	735	723	619	709
%	87	54	58	65	62	56	61
Criterion D							
R	.33	.20	.12	.12	.08	.19	.07
R^2	.11	.04	.01	.01	.00	.04	.00
N	1192	1165	1163	1124	1169	1102	1153
Correlation Between FT and NF							
	.73	.70	.59	.72	.69	.61	.66

Note—Criterion A = number of fixations/word ≤ 1 , Criterion B = of fixations/word ≥ 1 , Criterion C = number of fixations/word = 1, and Criterion D = complete set of data with number of fixations partialled out (for all Rs, $p < .01$).

The second analysis under Criterion B rejected all words that did not receive a fixation. Recall from the earlier comments that certain function words and short content words may be processed peripherally without direct fixation. Thus, a processing time of zero might not represent the true processing time for that word. The second block in Table 3 gives the obtained values. An average of 30% of the words were excluded for each subject. The fit of the model was much worse than the one for the complete data set. R^2 dropped from .49 to .16. Thus, words that are not fixated contribute most of the fixation time variance due to either psycholinguistic or eye guidance factors. Words not fixated may also have been a major source of gaze duration variance in Just and Carpenter's (1980) study.

The third analysis (Criterion C) was performed for words that received exactly one fixation. The corresponding values are displayed in the third block of Table 3. Although these estimates are still based on more than half of the original data, the fit of the model is now very poor ($R^2 = .05$). Note, too, that only for individual subjects is the data base substantially reduced; for the group analysis, it is almost as large (1,039) as the averaged data base (1,192). Nevertheless, the fit of the model for the average data was not much better than the ones for the individual subjects.

Instead of analyzing subsets of the data, one could treat number of fixations on a word as a covariate. The bottom block of Table 3 (Criterion D) shows the results of this analysis. Only 11% of the variance remained, compared with 49% when number of fixations is not partialled out. Thus, even a simple linear adjustment of gaze duration demonstrates the strong linkage between gaze duration and number of fixations. The correlations between number of fixations and gaze duration are given in the last row of Table 3.

The above analyses suggest that most of the variance in gaze duration is contributed by the number of fixations rather than by the duration of single fixations. We will argue in the discussion that number of fixations is as likely a result of perceptual acuity as a result of psycholinguistic processing demands. Predictors that do not allow a discrimination between these two alternative explanations are of limited value for the development of a theory of reading based on eye movements.

Discussion

Regression analyses on gaze duration. Regression analyses on the same data base can yield very different results under different theoretical assumptions about the priority of parameters. In natural discourse, indicators of different perceptual and psycholinguistic processes are highly correlated. This problem severely limits the utility of regression analyses for identifying the separate contributions of these variables. The results of simultaneous multiple regression are completely ambiguous in this regard, but hierarchical multiple regression allowed

the conclusion that word frequency and word length explain some variance independent of each other. This independent contribution, however, is very small compared with the amount of variance they share.

A similar argument can be made with respect to the two possible indicators of word length: number of letters and number of syllables. The amount of variance accounted for by number of syllables was shown to be completely contained within the amount of variance accounted for by number of letters. Therefore, it is hard to know from multiple-regression analyses whether syllable length variance actually represents the processing of syllables or whether it simply is not as precise a coding of word length as number of letters.

Length effects on fixation duration and lexical access.

Separate hierarchical multiple-regression analyses of zero, one, and two fixation words indicated that most of the variance in gaze duration was due to the number of fixations on words rather than to the duration of single fixations. Furthermore, in subsequent analyses of the present data (Kliegl et al., in press), there were no significant effects of either number of letters or number of syllables (between 5 and 11 letters) for words that were fixated only once, and these words comprised the majority of this sample. This result is inconsistent with Just and Carpenter's (1980) notion of serial processing of syllable units or Gough's (1972) serial processing of letter units. It argues for the unitary word code initially proposed by Cattell (1885) and most recently refined by McClelland and Rumelhart (1981).

An explanation for the lack of a length effect on the duration of single fixations may be that when familiar words are being processed primarily for their semantic content within a constrained semantic environment, as in typical reading situations, lexical access may be attained on the basis of parallel letter processing or partial information, such as beginning and end letters and word length. Empirical support for this explanation comes from studies by Green and Shallice (1976) and Terry, Samuels, and LaBerge (1976). They found no effect of letter or syllable length on the time required to perform semantic classification tasks. Other researchers found that even vocalization latency for high-frequency words was not affected by word length between one and two syllables or four to eight letters when words were presented in a story context to skilled readers (Perfetti, Goldman, & Hogaboam, 1979).

We should emphasize that the range in which we found no significant word length effects on fixation durations of words fixated only once was 5-11 letters. Similar ranges were used in the semantic classification tasks. We would not be surprised if word length effects emerged with wider ranges of length, say between 1 and 15 letters. Also, the interpretation of an absence of word length effect against Just and Carpenter's (1980) serial processing model is valid only if we accept their assumption that the fixated word is the only one being

processed. The fact that "the" is skipped more often than three-letter verbs indicates that this assumption is sometimes wrong. This complicates the interpretation of fixation duration data. In any case, it should be clear that there are grounds for questioning Just and Carpenter's serial syllable processing model. New approaches to the analysis of eye fixation data may be required to resolve the issue.

Separation of perceptual and psycholinguistic factors.

A major focus of this paper was on the separation of perceptual and psycholinguistic influences on gaze duration. Hierarchical analyses contrasting letter and syllable length effects argued that the spatial properties of the words might be more critical than the language properties. Support for perceptual factors was also obtained in further analyses of this data set (Kliegl et al., in press). Analysis of twice-fixated words showed that double fixations in reading of continuous text are in agreement with assumptions of inconvenient placements of the first fixation. Because twice-fixated and skipped words accounted for most of the variance in gaze duration, it seems likely that much of what Just and Carpenter (1980) counted as psycholinguistically relevant variance was actually due to peripheral acuity limitations. However, psycholinguistically relevant variables also influenced our subjects' eye fixations. After partialing out letter length, hierarchical multiple-regression analyses revealed a remaining 3% of the variance due to frequency. Word frequency also accounted for 1% of the variance in single-fixation duration, and it had a significant effect on the number of fixations on a word when letter length was controlled by selective sampling in the text.

While the relevance of both perceptual and psycholinguistic factors has been clearly demonstrated, their exact relative contribution still remains in question. Our data suggest the influence of perceptual factors, but new analytic approaches may be required to definitely answer the question. Perceptual factors may also be present in Just and Carpenter's (1980) data, considering that word length associated with number of syllables accounted for 46% of the variance when entered first, although there are several reasons why psycholinguistic variables may have had more independent influence in that study. Their material was more difficult, with novel words and more variance in word frequency. Also, there were differences in memory task demand, number of subjects, and model design. We turn now to examine the effects of each of these differences.

Comparison of the present study with Just and Carpenter's. The difference in number of subjects between the present study and Just and Carpenter's (1980) raises an important statistical point. The demonstration in the first section showed that pooled data lead to inflated estimates of variance for word length. Thus, part of Just and Carpenter's higher correlation may be due to their pooling of 14 subjects, compared with

6 in the present study. However, the fact that the average multiple correlation for individual subjects was .60 in Just and Carpenter's study and .45 in this one indicates that this is not the sole source of the difference.

The advisability of pooling subjects should be given serious consideration in future research. Averaging across subjects not only reduced error variance but also generated a dependence between gaze duration and word length that could be obtained even if fixations were randomly distributed across the text, that is, without any assumptions of local control of eye movements. If one wants to average, the procedure used by McConkie et al. (1979) is clearly superior.

There were several differences between the studies in the respective regression models. Just and Carpenter (1980) used 17 predictors; this study used 5. Two of the predictors excluded from the present study were the most powerful ones in Just and Carpenter's study: novel word and last word in paragraph. The size of the regression weights obtained by Just and Carpenter for () and last word in paragraph (157 msec) by far exceeded any of the other weights, which ranged from 25 msec to 71 msec. The reason for not including the novel word predictor in the present study was that there were few novel words in the Camus passage. The inclusion of last word in paragraph as a predictor is debatable even within the context of Just and Carpenter's study. This predictor coded the last word on the screen. They comment that this measure might be confounded with a motor response the subject gave to signal that he was finished reading. In the study reported here, the last word on the screen was not included in the analysis.

Perhaps the most theoretically interesting differences between the studies were in memory task demands and difficulty of the material. Just and Carpenter (1980) used technical text paragraphs on flywheels in combustion engines, thermoluminescence, staphylococci, and so on. Fifteen separate paragraphs were presented, and subjects were tested after each paragraph. This combination of technicality of reading material and frequent memory tests is very likely to have introduced different reading behavior compared with our study. After a few tests, their subjects may have realized that it was important to memorize the facts presented in the text. This might have caused them to very carefully examine words that carry the main ideas. Usually, these words are less frequent and longer.

CONCLUSION AND NEW DIRECTIONS

Statistical Perspectives

It has been the contention of this paper that multiple-regression analyses of eye movements are limited in their contribution to our understanding of the reading process as long as predictor variables associated with different postulated component processes are highly correlated.

For example, a measure such as word length is highly correlated with word frequency, and these measures are indicators of completely different hypothetical constructs. Hierarchical regression analyses were used to demonstrate this problem both in our data and those of Just and Carpenter (1980).

We focused on the percent of variance accounted for by different variables, whereas Just and Carpenter (1980) emphasized the regression weights. Both measures are of limited value with correlated predictors, and this was most easily demonstrated by noting changes in the percent of predicted variance with hierarchical regression analyses. In an attempt to bypass the problems of regression analyses, we selectively sampled words of specified length to separate the contributions of psycholinguistic variables from potentially confounding word length effects (Kliegl et al., in press). We found that with word length controlled (between five and nine letters), word frequency accounted for a significant 1% of the fixation duration variance for words that were fixated only once. Of course, the percentage is higher than this if word length is allowed to vary, but then frequency is confounded with length and we are unable to theoretically interpret the effect. Except for creating very artificial texts, post hoc sampling of words within natural texts seems to be the only available method for unconfounding many theoretically interesting variables.

By sampling words from a text to conform to an orthogonal design, one loses the ability to generalize. In this context, percent variance and the regression weights do not necessarily represent the model for the total population of words. However, as Hays (1973) points out, it still is a valid statistic for comparisons between experiments that employ similar constraints. For example, if one were interested in the effect of word length and word frequency on fixation duration, one is restricted to words of medium length for which a sufficient range of frequency values can be found. This design would still allow a test for effects of, for example, task demand, reading ability, and knowledge on fixation durations within the range of word length and frequency that is amenable to an orthogonal design.

Eye Fixation Measures

The use of gaze as a dependent measure is likely to bury important dynamic aspects of the reading process. Since the number of fixations is critical to the gaze measure, it will be important to determine the conditions under which two or more fixations occur. Kliegl et al. (in press) reported the two-fixation cases of this study to be in agreement with assumptions of inconvenient viewing positions and perceptual acuity limitations. An inconvenient viewing position, however, might be the result of processing constraints at the previous word. These constraints could be of a cognitive nature. Given high cognitive processing demands, the accuracy of the saccade generating system (i.e., its peripheral preprocessing ability) might suffer and lead to inadequate positioning of the next fixation.

Inconvenient positions might also reflect a property of the saccade generating system itself, if there is a lack of resilience to make very short or very long saccades in a normal reading situation. A fixation that lands at the end of a word may be followed by a saccade that will overshoot the convenient viewing position on the next word. Likewise, a fixation at the beginning of a long word may be followed by a saccade that will undershoot the convenient position on the next word. Kliegl et al. (in press) were able to predict the fixation location on a word with the length of the previous word and the position of the last fixation. The results were in agreement with the idea of a lack of saccadic resilience during normal reading. These results show that some double fixations might be a result of the distribution of previous word lengths in the text and the location of fixations on these words; they might have nothing to do with the word fixated presently. An extensive analysis of these sequential dependencies, however, may lay the groundwork for an understanding of eye guidance and may subsequently lead to an understanding of how higher cognitive processes modulate the dynamics of the reading process. With respect to this goal, an expansion of Rayner's (1979) analyses of conditional probabilities might be the most promising approach.

REFERENCES

- CAMUS, A. [*The plague*] (S. Gilbert, trans.). New York: Random House, 1948. (Originally published, 1947.)
- CATTELL, J. M. Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philosophische Studien*, 1885, 2, 635-650.
- COHEN, J., & COHEN, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum, 1975.
- CONGER, A. J. A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement*, 1974, 34, 35-46.
- DARLINGTON, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, 69, 161-182.
- DAVIDSON, B. J. On-line collection of eye movement data. *Behavior Research Methods & Instrumentation*, 1981, 13, 112-114.
- GOUGH, P. B. One second of reading. In J. F. Kavanaugh & I. G. Mattingly (Eds.), *Language by ear and by eye*. Cambridge, Mass: M.I.T. Press, 1972.
- GREEN, D. W., & SHALLICE, T. Direct visual access in reading for meaning. *Memory & Cognition*, 1976, 4, 753-758.
- JUST, M. A., & CARPENTER, P. A. Eye fixations and cognitive processes. *Cognitive Psychology*, 1976, 8, 441-480.
- JUST, M. A., & CARPENTER, P. A. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 1980, 87, 329-354.
- HAYS, W. L. *Statistics for the social sciences* (2nd ed.). New York: Holt, Rinehart & Winston, 1973.
- KERLINGER, F. N., & PEDHAZUR, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston, 1973.
- KLIEGL, R. Automated and interactive analysis of eye fixation data in reading. *Behavior Research Methods & Instrumentation*, 1981, 13, 115-120. (a)
- KLIEGL, R. *Eye movements in reading: An attempt to separate perceptual and psycholinguistic factors*. Unpublished master's thesis, University of Colorado, 1981. (b)
- KLIEGL, R., & OLSON, R. K. Reduction and calibration of eye

- monitor data. *Behavior Research Methods & Instrumentation*, 1981, **13**, 107-111.
- KLIEGL, R., OLSON, R. K., & DAVIDSON, B. J. On problems of unconfounding perceptual and language processes. In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes*. New York: Academic Press, in press.
- KUČERA, H., & FRANCIS, W. N. *Computational analysis in present-day American English*. Providence, R.I.: Brown University Press, 1967.
- MCCLELLAND, J. L., & RUMELHART, D. E. An interactive activation model of context effects in letter perception. Part 1. An account of basic findings. *Psychological Review*, 1981, **88**, 375-407.
- MCCONKIE, G. W. Evaluating and reporting data quality in eye movement research. *Behavior Research Methods & Instrumentation*, 1981, **13**, 97-106.
- MCCONKIE, G. W., HOGABOAM, T. W., WOLVERTON, G. S., ZOLA, D., & LUCAS, P. A. Toward the use of eye movements in the study of language processing. *Discourse Processes*, 1979, **2**, 157-177.
- MCCONKIE, G. W., & RAYNER, K. The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 1975, **17**, 578-586.
- MITCHELL, D. C., & GREEN, D. W. The effects of context and content on immediate processing in reading. *Quarterly Journal of Experimental Psychology*, 1978, **30**, 609-636.
- O'REGAN, J. K. Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics*, 1979, **17**, 578-586.
- O'REGAN, J. K. The control of saccade size and fixation duration in reading: The limits of linguistic control. *Perception & Psychophysics*, 1980, **28**, 112-117.
- O'REGAN, J. K. The "convenient viewing position" hypothesis. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception*. Hillsdale, N.J.: Erlbaum, 1981.
- PERFETTI, C. A., GOLDMAN, S. R., & HOGABOAM, T. W. Reading skill and the identification of words in discourse context. *Memory & Cognition*, 1979, **7**, 273-282.
- PYNTE, J. Readiness for pronunciation during the reading process. *Perception & Psychophysics*, 1974, **16**, 487-504.
- RAYNER, K. The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 1975, **7**, 65-81.
- RAYNER, K. Eye guidance in reading: Fixation locations within words. *Perception*, 1979, **8**, 21-30.
- RAYNER, K., MCCONKIE, G. W., & EHRLICH, S. Eye movements and integrating information across fixations. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 529-544.
- TERRY, P., SAMUELS, S. J., & LABERGE, D. The effect of letter degradation and letter spacing on word recognition. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 577-585.

NOTE

1. The presence of suppressor variables in the model can lead to increments in the variance accounted for that are not due to the variable added in the last step but are the result of the simultaneous presence of two variables. Fortunately, there are procedures to determine whether or not suppressor effects are present (Cohen & Cohen, 1975; Conger, 1974).

(Received for publication October 6, 1981;
revision accepted February 10, 1982.)

ANNOUNCEMENT

Twenty-Third Annual Meeting of the Psychonomic Society

The 23rd annual meeting of the Psychonomic Society will be held at the Hyatt Regency Hotel in Minneapolis, Minnesota, November 11-13, 1982. The Call for Papers will be mailed to Society members during May 1982. *Papers dealing with the application of psychological science to real-world problems are encouraged.* The program and hotel reservation forms will be mailed to members in the early fall. The program will also be printed in the September issue of the *Bulletin of the Psychonomic Society*. The Secretary-Treasurer, J. Bruce Overmier, may be contacted for further information at the Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455 [(612) 373-3430].