Use of Reaction Time Methods as Indicators of the Processes Underlying Reading.

Paper presented at the Annual Meeting of the International Reading Association (20th, New York City, May 13-16, 1975)

Decoding (Reading); Higher Education; Information Processing; *Literature Reviews; *Reaction Time; *Reading Processes; *Reading Research; Reading Skills; *Research Methodology; *Task Performance; Word Recognition

This paper describes a methodological approach to the reading processes, involving the use of measures of speed of response or reaction time (RT) obtained from subjects while performing reading related tasks. Following a review of the literature on measures of RT, two experiments currently being done at the University of Minnesota are described and the methodology used is also discussed. The three main objectives of the two experiments were to investigate the relationships between (1) the ease of decoding and the speed of semantic processing, (2) the ease of decoding and the type of perceptual processing (serial or parallel), and (3) word length and speed of semantic processing. It was concluded from the evidence provided in the literature review and the experiments that RT can be a useful index when exploring the processes underlying reading.
Use of Reaction Time Methods as
Indicators of the Processes Underlying Reading

IRA Convention - Symposium 7

1975
Use of Reaction Time Methods as Indicators of the Processes Underlying Reading

Pamela R. Terry

University of Minnesota

What is it that a reader can do that a non-reader cannot do? At the University of Minnesota several researchers have been striving to find answers to this question by exploring factors underlying reading. A promising methodological approach involves the use of measures of speed of response or reaction-time (RT) obtained from subjects while performing reading-related tasks.

In this paper, I will first review briefly some past research involving measures of RT, describe some research currently being done at Minnesota and later discuss some particulars concerning the RT methodology used.

One of the basic disagreements in the field of reading today involves the level of perceptual processing used by readers. Is processing during reading serial or parallel? Opinions range from those who view reading as a process of serially processing words letter-by-letter to those who believe groups of letters, words or even larger segments of the language are processed as wholes in parallel. Reaction time methodology can be fruitfully used to investigate this question.

At one level, this controversy enters the theoretical realm regarding the nature of information processing. On the other hand, there are some very practical and applied issues regarding how reading should be taught. Views on the question "Is processing during reading serial or parallel?" lead to pedagogical decisions involving whether reading is taught as a series of subskills or as a holistic process.
Levels of Perceptual Processing

The dispute over whether words are recognized mainly by serial processing or parallel processing has continued for almost a century. Serial models of word recognition usually assume that the processing of letters is serial or one-by-one. However, even strong supporters of serial processing reach a level where parts of the visual stimuli have to be processed as a unit—that is, in parallel. For example, most serial models assume that the features of letters (e.g. lines, curves, diagonals, etc.) are processed in parallel.

Just as all serial models reach some lower level at which processing is assumed to be in parallel, all parallel models reach a final upper level at which processing is assumed to be serial. This is often at the level of letters, spelling clusters, syllables, morphemes, phrases and even sentences. As Wheeler (1970) has noted, however, "No one proposes that we read a whole paragraph in parallel (p. 79)."

When referring to the reading process, the terms serial and parallel have generally come to mean specific levels of processing. Serial models of reading assume that words are recognized by processing the letters one at a time—serially. Parallel models of reading assume that larger units than the individual letters are recognized as wholes by processing several of the letters simultaneously—in parallel.

Evidence for Parallel Processing of Letters in a Word

The experiments of Cattell (1885) have often been used as evidence of parallel processing and support for the whole word method of teaching reading. Cattell gave fluent readers a number of words to read (connected prose passages from Gulliver's Travels) and an equivalent number of letters to name. When subjects required approximately the same amount of time to
read either the words or the letters, Cattell concluded that subjects
were recognizing words as wholes in much the same manner that they recognized
individual letters as wholes. In a separate experiment he showed that
short words could be identified as quickly as single letters. This data
has often been used to support a model of parallel processing of letters
in a word. There are some common criticisms of Cattell's work that can
be pointed out. He presented words in context. This allowed readers to
predict what words would be likely to occur next, whereas the letters
were not predictable. Cattell also used fluent readers as his subjects.
Less skilled readers might have performed in a very different manner.

Huey (1908) described several experiments of Erdmann & Dodge (1898)
that support a theory of perception in word wholes. In testing
recognition times for words of different lengths (4, 8, 12, 16 letters)
they found that the longer words needed comparatively little additional
time—the longest ones requiring only about 20 percent more time than
the shortest.

Suggestive evidence for parallel processing of a limited number of
separate units at a time comes from experiments such as that of Sir William
Hamilton (1859, p. 176) who in lectures prior to 1859 stated, "If you
throw a handful of marbles on the floor, you will find it difficult to
view at once more than six or seven at most, without confusion." Kaufmann
et al. (1949) updated this experiment by showing slides of dots ranging in
number from 1 to 200 for 250 msecs. exposure time. Subjects made no
errors on slides of six dots or less.

In considering whether we read by letters or word-wholes, Huey (1908)
had subjects read down columns of single letters and 4-letter, 8-letter,
12-letter, and 16-letter words. He found relatively little increase in
the amount of time required to read words up to 8 letters long, but great increases in the time necessary for reading the longer words. He concluded that "recognition of familiar and comparatively short words is little affected by doubling the number of letters; and this seems confirmatory of the view that such words are recognized in one unitary act, as wholes (p. 101)." The longer times required to read the longer words were attributed to unfamiliarity and increased eye movements necessary for taking in the longer words.

Kinsbourne & Warrington (1962), after finding that 8 msecs. was the minimum exposure time necessary to process a single letter on a tachistoscope, discovered that adding 2 or 3 more letters to process required no additional time. However, their subjects were highly practiced in tachistoscope viewing, having had hundreds of presentations. In addition, they knew that no more than 2 or 3 letters would appear in the stimulus field.

Haber (1970) found that although naive subjects seemed to show serial processing in word perception, practiced subjects showed parallel processing when performing the same task.

Evidence for Serial Processing of Letters in a Word

Cattell (1885), in spite of other findings generally supporting parallel processing, found that when subjects were presented with unfamiliar words or long words (over 8 letters) the words took longer to recognize than a single letter.

When Pillsbury (1897) had subjects read words briefly exposed in a tachistoscope, he noted that their ability to recognize typographical errors in the words decreased from first to last letter. Pillsbury concluded, "This seems to indicate that the subject read through the word from left to right (emphasis in original, p. 350)."
When Huey's (1908) subjects read columns of words and letters he found that even though the times required to recognize letters and short words (4 or 8 letters) were close, there was still a systematic increase in times for recognition according to word length.

Gough is perhaps the leading scientific proponent of serial processing. He stated (1971, p. 335), "I see no reason, then, to reject the assumption that we do read letter-by-letter. In fact, the weight of the evidence persuades me that we do so serially, from left to right."

Stewart, James and Gough (1969) had subjects read nouns 3 to 10 letters long as quickly as possible. Whereas many earlier studies had rather gross measurements of latencies involved in word recognition (e.g. Huey, 1908), or used only the visual duration of the stimulus exposure as the dependent variable leaving the response interval between onset of stimulus and onset of response unassessed, the Stewart et al. study measured the latency between stimulus onset and the pronunciation of the word as the dependent variable. They found that recognition times increased steadily with the number of letters per word. However, the function was negatively accelerated with a greater increase in latency with length for short words than for long words.

Gough & Stewart (1970) found that short words (3-letters) were recognized 35 msecs. faster than longer words (6-letters) when deciding whether strings of letters were words or not.

Pearson & Kamil (1974) presented nouns and verbs varying from 4 to 7 letters in length and measured subjects' response time to initiate pronunciation. The verbs also appeared in their inflected forms with the addition of the affixes -s, -ed, or -ing. Results showed a relationship between word length and recognition latency. However, this evidence
In support of serial processing was found on only the first trial block. After repeated exposures, differences in latency as a function of word length disappeared. The addition of inflectional endings varying from 1 to 3 letters in length caused an increment in word latency equivalent to the addition of only a single letter, suggesting that such affixes are either treated as one unit, or, as the authors suggest, are perhaps clues for morphological segmentation.

From the evidence supporting parallel processing and that supporting serial processing, it seems clear that in reading, the level of processing can vary depending upon a number of factors. Such factors would include both subject variables (such as knowledge of the language, decoding skill, experience, etc.) and task variables (such as legibility of type, amount of context, word frequency, comprehension load, etc.).

The following experiments were done to further investigate specific instances of either serial or parallel processing under varying conditions of favorability for decoding processes.

Initial Pilot Studies

In order to contrast the effect of proficient and inefficient decoding skills upon comprehension, adult readers were provided with reading material that they could easily process syntactically and semantically, but that was printed in transformed orthography (mirror-image text).

This rather simple orthographic transformation was hypothesized to cause normally fluent readers some of the same decoding difficulties as beginning readers--i.e. difficulty in remembering certain letters, problems remembering the direction of scan, and difficulties confusing similar letters such as b, d, p, q. One of the adults' main problems was, of course, interference from past experience. However, this same interference
happens quite commonly in real life when, for example, one learns to read another language. In such cases one must often learn to recognize new letter shapes composed of familiar features or learn that the same letters must be associated with different sounds, or, in cases such as Hebrew, learn that the direction of scan must be reversed.

Measures used to assess decoding proficiency consisted of reading speed (in words per minute--wpm) and number and type of reading miscues. The comprehension score was the number of comprehension questions answered correctly.

The most striking result of this preliminary study and other variations on it was the extreme drop in reading speed when reading transformed text. These normally fluent readers dropped to reading rates as low as 3.9 wpm. They struggled to decode the words and often resorted to spelling out the letters of the words or sounding them out, as tape recordings show.

Comprehension scores were significantly lower (and in several cases dropped to chance levels) when subjects struggled with the decoding problems of reading transformed text compared to their scores when reading similar passages in normal text.

As an interesting aside, several of the subjects were either teachers-in-training or parents of young children learning to read. They all commented that the experience of struggling to decode the transformed text gave them renewed respect for the frustrations and difficulties their children face in learning to read.

Another intriguing tidbit resulting from these initial studies was the finding that a number of the subjects complained of eyestrain from reading such "tiny" print. Several requested that the large print found
in children's books be used when reading transformed text. Both the normal and mirror-image versions of the stories had been printed by the same elite typewriter. This information seemed to imply that these subjects were reading mirror-image text in a manner qualitatively different from their reading of normal text as well as the more obvious quantitative difference found in reading speed. It seemed probable that these subjects were looking at smaller features of the stimuli such as individual letters or letter-features instead of their more normal reliance on larger units such as words, configuration clues, or phrases. This led to further examination of possible differences in the levels of processing used depending on the ease of decoding. These initial studies seemed to indicate that serial processing is the norm when decoding is difficult as opposed to parallel processing when decoding is easy.

These initial studies also suggested further examination of another factor. Several times subjects claimed they were unable to decode a particular word because there was a slight gap in the printing of the letter or a slight smudge or smearing of the word. Yet, when they had completed the task and questions and out of curiosity turned the page over and held it to the light so that they could read the story in regular orthography, they could read all of the words perfectly, smudges, smears and gaps notwithstanding. This was in spite of the fact that they were reading a very faint image from the wrong side of the paper that was much more distorted in terms of clarity than the passage of transformed text on the right side of the page. The difference was that the passage, when read from the wrong side of the paper was written in familiar orthography. A degradation factor was therefore included later.
in experiments to aid in investigating levels of processing and as an additional decoding problem to be overcome.

Experiments with Fluent Readers

The three main objectives of the following two experiments were to investigate the relationship between:

a) ease of decoding and speed of semantic processing
b) ease of decoding and type of perceptual processing (serial or parallel)
c) word length and speed of semantic processing

Experiment I

Method

Hypotheses

It was hypothesized that a positive relationship would be found between ease of decoding and speed of semantic categorization. In other words, it was predicted that words written in transformed (mirror-image) text would be processed more slowly than words written in normal orthography.

It was further hypothesized that degradation would more seriously impair processing of words written in transformed orthography than that of words written in regular orthography. Consequently, a significant interaction between the two factors of degradation and orthography was predicted.

Design

A 2 X 2 factorial design was used. Factors were orthography (regular or transformed) and clarity (degraded or non-degraded letters). Degradation was achieved by randomly deleting 30% of the dots forming each letter on the TV screen with the stipulation that such deletion not change one letter into another, e.g. change 1 into i. Forty college students were randomly assigned to the four treatment groups.
Experimental Task

Individual subjects were seated before a Nova Computer with a TV screen facing them. A Donder's - c response method was used in which subjects were to press the button if an animal word appeared on the screen. If the word was not an animal, they were not to press. The reaction time data gathered by the computer measured the time that elapsed between the first appearance of the word on the TV screen and the time the button was pressed in response.

A neutral cue (a plus sign [+]) acted as a fixation point on the TV screen showing the subject where he should be looking and also warned the subject that the target stimulus was about to appear. The cue remained for 1000 msec (1 second) followed by a 1000 msec blank field. The target (the animal or non-animal word) remained on the screen 3000 msec unless terminated sooner by a button press.

The computer measured accuracy and latency of response. In order to make a correct decision, the subject had to decode the word and semantically process it into the appropriate category of animal or non-animal. Thus, in addition to knowing whether the subject was correct or not, the RT data provided information about how long it took the subject to read the word and semantically process it.

Word List

Seventy-two animal and twenty-four non-animal words were selected from the American Heritage Word Frequency Book (Carroll, 1971), and included the highest frequency animal words found in the book. For ease of presentation, the words were randomly divided into 4 separate lists of equal word frequency, each list containing 18 animal words and 6 non-animal words. Word length ranged from 3 to 8 letters. The non-animal words were
of the same lengths and of similar high frequencies.

The 3 to 1 ratio of animal to non-animal words allowed sufficient positive responses for adequate data collection (only positive responses yielded reaction time data) while including enough 'catch' trials to assure that the subject was reading the words and not just guessing. The non-animal words acted as catch trials to be sure the subjects were actually taking the time to process the information and not responding simply on the basis of time estimations or internal response predictions. If too many errors had occurred on catch trials, the data would not have been used.

Results

As seen in Figure 1, the response latencies for both regular nondegraded and regular degraded orthographies were quite fast, approximately 575 msec. For the transformed text, response latencies were considerably slower, as expected. The ANOVA indicated a significant difference between regular and transformed orthography, $F(1,36) = 142.4, p < .001$; between degraded and non-degraded text, $F(1,36) = 8.8, p < .01$; and a significant interaction between orthography and degradation, $F(1,36) = 4.6, p < .05$. A simple effects analysis showed no difference between degraded and non-degraded text under regular orthographic conditions, $F(1,36) = 1.3, NS$, but significant differences between degraded and non-degraded text under transformed orthography conditions, $F(1,36) = 9.12, p < .005$. Similar results were found for accuracy. Degradation had no effect on accuracy of categorizing words in regular text, but had significant effects on categorization of words in mirror-image text.

Discussion

Effect of Decoding Ease on Speed of Semantic Categorization

In regard to the role of decoding in comprehension, when fluent readers
encountered words in regular orthography which they could decode with ease, semantic categorization occurred with great rapidity. On the other hand, when decoding was made difficult by presenting orthographic transformations, processing was significantly slower. Furthermore, as the decoding task became increasingly difficult by combining mirror-image text and degradation in a single condition, the difficulty in processing was significantly increased. There was thus a positive relationship between ease of decoding and speed of semantic processing.

**Effect of Decoding Difficulty on Type of Perceptual Processing**

It was found that degradation had differential effects upon speed of semantic categorization depending on the type of orthography used. When words were printed in regular orthography, degradation had no effect. However, when words were presented in mirror-image text, degradation resulted not only in words being categorized significantly slower, but also in significantly fewer words being categorized to begin with.

In exploring reasons for this result, the most likely explanation seems to be that different hierarchical levels of visual processing are used depending on the ease of decoding the material or the decoding proficiency of the individual. If one assumes that a fluent reader reading regular text can process words in a holistic, parallel, unitary manner by perhaps drawing on configuration cues, redundancy or other means, then degradation at the level of individual letters would not be expected to interfere greatly with word recognition. One would already be processing higher order units above the letter level and not be disturbed by marring of individual letter features.

Alternatively, if one is struggling with unfamiliar orthography, as is the case when reading mirror-image text or when one is a beginning
reader, then processing may be occurring serially at the level of individual letters. The fact that subjects reading transformed text often resorted to spelling out the letters of the words when reading them seems to give this explanation added validity. When degradation occurred together with mirror-image text it disturbed the input information at the very level at which processing was occurring, leading to the severe disruption found in speed and accuracy of semantic processing.

Experiment 2

Method

Hypotheses

A positive relationship between ease of decoding and speed of semantic processing was again expected. In addition it was hypothesized that words written in regular orthography would be so well-learned that there would be no relationship between word length and speed of semantic processing. With words written in transformed text, however, it was probable that serial processing of letters would occur resulting in a positive relationship between word length and speed of semantic processing. The spacing factor was an attempt to disrupt word features as opposed to the disruption of letter features introduced by degradation in Experiment 1.

Design

A 2 X 2 factorial design was used. Factors were orthography (regular or transformed) and spacing (even or uneven spacing between letters in a word). Forty college students were randomly assigned to the four treatment groups.

Experimental Task

The task was the same as in Experiment 1.

Word List

The word list included 64 animal words from Experiment 1 and 20 non-animal
words. Words ranged in length from 3 to 6 letters. Non-animal words (catch trials) were also selected to be 3 to 6 letters in length and of similar frequency and visual appearance. The ratio of animal to non-animal words was 3:1.

Results

As seen in Figure 2, there does not appear to be any relationship between number of letters in a word and response latency for words presented in regular orthography. On the other hand, one notes curvilinear negatively increasing relationships between word length and response latency for transformed text. The analysis of variance indicated the following: There was a significant difference in response latency between regular and transformed text, \( F (1,36) = 435, p < .001 \); a significant difference in response latency on number of letters in the word, \( F (1,36) = 39.44, p < .001 \); and a significant interaction between the orthography and the number of letters per word, \( F (1,36) = 39.35, p < .001 \). Looking at the simple effects found in the interaction, there was no significant difference in latency for different word lengths when they appeared in regular orthography, \( F (3,108) < 1, NS \). However, for words printed in transformed text, there was a significant effect on latency related to word length, \( F (3,108) = 78.63, p < .001 \). The hypothesis concerning spacing was not confirmed.

Discussion

This study found support for the two competing models of word recognition. Both serial and parallel processing occur, but under different conditions. When fluent readers encounter familiar words presented in regular orthography, they seem to be able to chunk that information holistically, at least within the upper limits of 3- to 6-letter words as used in this study. On the other hand, when fluent readers
encounter words which tend to pose a decoding problem, as was done in this study by presenting words in mirror-image text, we find evidence for serial letter-by-letter processing. Thus, we may say that fluent readers adopt different strategies of word recognition, depending upon factors having to do with the ease of decoding the visual input.

**Implications**

When decoding is difficult, as is often the case for beginning readers, we should find serial processing. With greater familiarity with text, or as the reader becomes more skilled, the decoding task should become simplified and we should find evidence for parallel processing.
Reaction Time Methodology

As shown in the previous studies, RT can be a useful index to use when exploring the processes underlying reading. Most of the early studies measuring speed of reading-related processes (e.g. speed of word recognition studies) can be criticized for two reasons: 1) The data were obtained from tachistoscopes where only the brief stimulus exposure duration was controlled and the amount of time until a response occurred was unassessed. 2) The measurements were too crude and imprecise.

With the development of sophisticated computer hardware, we can today measure latencies from stimulus onset to response onset with an accuracy and precision unavailable in the past.

When measuring RT with a button press, there are three standard categories of response--Donders-a, Donders-b, and Donders-c.

A Donders-a response, or "simple" reaction time, consists of a response (button-press) to the presence as opposed to the absence of a stimulus when only one type of stimulus is presented. In other words, the subject using a Donders-a response is performing a detection task. Instructions for such a task might be as follows: "If a green stimulus appears, press the button."

A Donders-b response occurs when two separate responses are made to two different kinds of stimuli. For example, task instructions might be: "If the stimulus is green, press the left button; if the stimulus is red, push the right button."

A Donders-c response consists of only one response, like a Donders-a, but is more complex. The Donders-c response requires the subject to respond to one type of stimulus, but not to respond to a second type. For example, instructions to a subject might be: "If the stimulus is green,
press the button; if the stimulus is red, do not press the button."
It is like a "Donders-b response with one hand in your pocket."

One must beware of artifacts in RT data. Standard RT experiments
use repeated trials consisting of 1) warning signal, 2) (blank) interval,
3) stimulus, 4) feedback (optional), 5) interval. With simple
reaction time from a Donders-a response, there is the possibility that a
subject is responding to the time interval instead of the stimulus.
For example, a subject may be responding in rhythmic intervals, doing time
estimations rather than scrutinizing the stimulus. This possibility
can be controlled by varying the interval before the stimulus. However,
this can lead to a response bias on longer intervals, because as you
approach the end of the foreperiod interval and no stimulus has appeared,
you build up an expectancy as the probability of the stimulus occurring
increases. A better way to control for time estimation is to intersperse
blanks among the stimuli to act as catch trials, e.g. to "catch" a
response when there is no stimulus there.

With a Donders-b response there may be artifacts due to successive
repetitions of the same stimulus and thus the same response. Subjects
are usually faster on repetitions than on new responses. Moving to use
of a Donders-c response eliminates most response bias as there is only
one response. Sequential effects can be eliminated by preparing the
subject for a particular response by cueing.

The probability of a response can be influenced by the ratio of
correct to incorrect stimuli. At the University of Minnesota several
researchers have used a Donders-c response method with 25-33% catch trials
or a ratio of $\frac{3}{4}$ or $\frac{7}{8}$ to give subjects an expectancy to respond
while still providing enough errors to monitor their accuracy. If accuracy
is too low, reaction time data is not used.

An important benefit of RT data is that they give more information than would be available from measures of accuracy alone. For example, although several students may answer a question correctly, a large variance in the time it takes them to answer may mean that some students know the answer better. Very slow responders may require more study or practice to reach an equivalent state of learning.

Reaction time measures can also be used to infer qualitative as well as quantitative differences in information processing.

For example, both Experiments 1 and 2 found that for each additional letter in a word, there was an increase in reaction time for words written in unfamiliar mirror-image orthography. However, for words written in regular orthography, word length had no effect on speed of semantic categorization. This points to a different manner of processing the two kinds of stimuli. If the processing were qualitatively the same but only differed quantitatively, one would expect to see speed of categorization slowed uniformly under unfamiliar orthographic conditions, regardless of word length. But the finding of both a change in Y-intercept and a change in slope points to different activities occurring when decoding is difficult compared to when it is easy.

**Experiments with Beginning Readers**

Experimentation is currently in progress using RT methodology to study the type of processing beginning readers use. First and second graders and also mentally retarded children are being asked to read orally high and low frequency words under degraded or non-degraded letter conditions. Accuracy and latency data on semantic categorizations of
words varying in length are being obtained from children through the use of a portable minicomputer that can be used right in the schoolroom.

Reaction time data have already shown that fluent adult readers can use either serial or parallel processing depending on decoding difficulty. It will be used to explore the processing abilities and skills of beginning readers. The use of reaction time methodology can provide reading researchers with a very useful tool for discovering exactly what is involved when we read.
References


Figure 1: Differential Effect of Degradation Under Regular and Transformed Text Conditions

- Transformed Text Degraded
- Transformed Text Non-degraded
- Regular Text Degraded
- Regular Text Non-degraded

Average score difference in test.
Figure 2

"Differential Effect of Word Length Under Regular and Transformed Text Conditions"