The Influence of Task and Format on Reading Results with an Online Text

Abstract—The aim of this study was to map the influence of reading task and text format on reading results with an online text. To this purpose, an experiment performed by Gordon et al. (1988) was replicated and enhanced. In four conditions, subjects were given a reading task (summarize or answer specific questions) and an online text (linear or hypertext format). In all conditions, both text and task were administered through the World Wide Web. After the subjects had completed their reading, all were given the same assignment: make a summary and answer specific questions. No significant main effects of the independent variables (format and task) were found on the performance of the subjects. There proved to be a significant interaction effect, however, on the completeness of the summaries. The most thorough summaries were written by subjects who were told before the experiment that they would have to summarize the text, and who were presented with the text in a linear version. As far as reading time was concerned, there was a significant difference between the format conditions: reading the text in linear format took more time than reading the text in hypertext format.

Index Terms—Document design, hyperlinks, hypertext, linear text format, reading task, text format.

Manuscript received July 17, 1998; revised March 11, 1999.

S. Spruijt is with Bertelsmann On Line, 3515 CP Utrecht, The Netherlands (email: SSprijtt@bol.com).
C. Jansen is with Nijmegen University, 6500 HD Nijmegen The Netherlands (email: C.Jansen@iet.kun.nl).
IEEE PII S 0361-1434(99)04915-2.
We can decide on the consequences that the expected reading task should have on the text format to be chosen only once when we have an articulate idea about the combined effects of task and format on reading results.

**AN EARLIER STUDY:**
**GORDON et al. (1988)**

**Although a number of recent publications argue in favor of specific attention for the influence of the task on the reading process** [1-4], Gordon et al. [5] are some of the few researchers who use both task and format as independent variables in a reading experiment.

Gordon et al. [5] started out with the hypothesis that the typical shape of hypertext would facilitate comprehension and memory of expository texts. In addition, they assumed that the benefits of hypertext would be best felt if the subject matter of a text was relatively difficult and unfamiliar, and if the reader was making a serious attempt to learn the material. To test these hypotheses, Gordon et al. measured the main and interaction effects of two independent variables: format and text.

Gordon et al. divided a group of 24 subjects into two subgroups. The first subgroup was given a technical text (about attentional factors in jet aircraft crashes) in a linear format and a second technical text (about speech analysis and recognition) in a hypertext format. The second subgroup also was given a linear text and a hypertext, but in this group, the text type was general interest (the topics were: falling in love and reverse sterilization). All texts were presented on a PC monitor. The subjects working with the technical texts were asked to carefully study the material and be prepared to answer a number of (unspecified) questions about the content of the text. The other subjects were encouraged to read the text as they would normally read material of interest to them. After having read the texts assigned to them, all subjects in this experiment were given both an open task (a free recall test) and a closed task (answering probe questions, 48 on average) for each of the two stories. Reading times were recorded. Furthermore, answers were collected to questions about the subjective perceptions of the linear and hypertext formats.

From the results they obtained, Gordon et al. concluded that their expectations were not confirmed. The reading of none of the texts was improved by the hypertext format. The hypertext format even proved to have a negative influence on the quality of the answers given to the probe questions about the texts. Moreover, in the general interest texts, the hypertext format resulted in poorer achievements in the free recall tests than the linear format did. As far as the total reading time was concerned, there were no significant differences for text format nor significant interaction effects. Finally, the majority of the subjects preferred the linear text, and general interest readers stated that hypertext required more cognitive effort.

These outcomes contradict the general conclusion that Chen and Rada [2] draw from a meta-analysis of experimental studies on hypertext systems. On the basis of the experiments they analyzed, Chen and Rada claim that hypertext systems prove to be appropriate for open tasks, such as browsing and assimilating. According to their conclusion, users complete closed tasks faster with non-hypertext systems, whereas they benefit more from hypertext systems for open tasks. Chen and Rada show no reservations here. They explicitly state that the hypothesis that users perform more effectively with hypertext for open tasks than for closed tasks was supported by all studies taken into consideration. This assertion is rather surprising considering that one of the experiments on which they base their conclusion is the study by Gordon et al. [5]. If only for this reason, it seems worthwhile to exactly replicate the experiment performed by Gordon et al.

Gordon et al.'s setup suffers from a serious flaw, however. In their study, there were not two independent variables, as they suggest, but, in fact, three variables were manipulated: text format on the one hand and an inextricable combination of two other variables (text type and reading task) on the other hand. Due to the introduction of this combined variable, it is virtually impossible to track back any potential effect differences to either the text presented to the subjects (technical or general interest) or the task they were asked to prepare themselves for (study carefully or read as usual). In other words, in this experimental design it could not be decided which of the independent variables (text type or reading task) would be responsible for which part of a possible effect on any dependent variable.

The research question, however, as introduced by Gordon et al. [5], seems important enough to undertake another effort. Not only from a theoretical point of view, but perhaps even more from a practical standpoint, it is worthwhile to try and gain more insight into the influence that the tasks readers start out with have on their performance. More and more, document design specialists can opt for more than one format for presenting their information. Evidently, the more knowledge available to document designers about the effects of their choice for a specific format in a given situation, the more well-considered and user-friendly their decisions can be.

For this reason, we decided to replicate and enhance the study...
of Gordon et al. [5]. To avoid the problems that affected their experimental design, we devised a number of adaptations. The most important modifications are the following:

1) Just as Gordon et al. [5] did, we used two independent variables. However, instead of using format and a combined variable text/task, we kept the variable text constant, and only manipulated the variables format and reading task.

2) The independent variable format was changed from a within-subjects variable to a between-subjects variable. All 24 subjects in the experiment of Gordon et al. [5] were involved in two conditions: they all had to read two texts and to fulfill the corresponding assignments. In our study, each of the 46 subjects was included in only one condition. With this intervention, we prevented the occurrence of undesirable learning effects.

3) Whereas in the study of Gordon et al. [5] subjects were informed in only global terms about their reading tasks, our subjects were instructed more precisely: one group was presented with two very specific questions, and these subjects were told that they had to answer both questions using the information in the text. The subjects in the other group were told to be prepared to give a summary to another person who would be interested in getting the "big picture" from the information in the text. In this way, we tried to improve on the ecological validity of the experiment, at least where the closed assignment was concerned. Asking users of an online text to look up the answers to a small set of specific questions seems to be more natural than telling them to carefully study the material and to be prepared for a large number of (unspecified) questions.

4) We made use of the facilities created by the World Wide Web, which evidently did not exist when Gordon et al. performed their experiment [5]. All our subjects worked individually on a multimedia PC that was connected to the Internet. At our Web site, the instructions, text, and assignments for all conditions were presented online. The subjects also carried out their assignments via the Web site's interface. During the reading process, the total reading time was recorded.

Design and Implementation of the Experiment
As shown in Table I, our study was based on a 2 x 2 factorial design with two between-subjects variables. One text of about 2000 words was presented both in linear form and as a hypertext to a total of four groups of subjects. There were two linear text conditions (1 and 2) and two hypertext conditions (3 and 4). The subjects in conditions 1 and 3 started reading with a summary task in mind, while the subjects in conditions 2 and 4 were asked to answer two specific questions.

In all conditions, the texts were offered online to ensure that no effect differences occurred as a result of differences between paper and electronic environment [6]. All material was presented to the subjects in Dutch.

After the text had been read, all subjects were faced with an assignment that exceeded the task that they expected to have to carry out. All subjects were invited to make a summary of the text and to answer two specific questions. The total reading time for each subject was recorded. After having read the text and having completed the assignment, all subjects were asked to assess the text and the accompanying assignment in a short questionnaire.

Subjects Originally, 60 subjects were invited to participate. However, due to some technical problems in transporting the data electronically, the achievements of 14 participants turned out to be useless for further analysis, so that eventually the data from 46 subjects could be processed. All subjects were graduates from Dutch universities or first-year students of the Arts Faculty of Utrecht University. The graduates and first-year students were divided equally among the four conditions.

Text Content The subject of the text was a topic that was expected to be equally unfamiliar to all subjects: the *Eigen Bijdrage Regeling van het Ziekenfonds* ("Personal Contribution Regulations of the Dutch National Health Scheme").

<table>
<thead>
<tr>
<th></th>
<th>Linear text</th>
<th>Hypertext</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Condition</strong></td>
<td><strong>Subjects</strong></td>
<td><strong>Subjects</strong></td>
</tr>
<tr>
<td>1 (9 subjects)</td>
<td>2 (11 subjects)</td>
<td>3 (13 subjects)</td>
</tr>
</tbody>
</table>

TABLE I
Experimental Design

<table>
<thead>
<tr>
<th></th>
<th>Expected task: Make a summary</th>
<th>Expected task: Answer specific questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Condition 1 (9 subjects)</td>
<td>Condition 2 (11 subjects)</td>
</tr>
<tr>
<td>Hypertext</td>
<td>Condition 3 (13 subjects)</td>
<td>Condition 4 (13 subjects)</td>
</tr>
</tbody>
</table>
The text was composed of more or less independent information units which did not need to be read in a fixed order.

**Independent Variable: Text Format** In the linear conditions (1 and 2), the text presented to the subjects was made as analogous to the original text as possible. To enable the readers to move through the text, on each screen FORWARD and BACK buttons were added, as well as a series of numbers (1-15) the readers could click on to go to a specific page. Furthermore, each page had a link to the Table of Contents.

In the hypertext conditions (3 and 4), the information was presented in layers. Moreover, where it was logically possible, cross-links were added: hyperlinks that allow the reader to jump from one conceptually related text unit to another without the necessity of a hierarchical relationship. Furthermore, the Table of Contents was constantly visible in a separate frame, and it was possible to visit an index and to perform full-text searches. Both the linear and the hypertext versions were presented on the WWW.

**Independent Variable: Task** As indicated above, a distinction has to be made here between the task the subjects thought they were going to carry out and the task that was actually given after they had read the text. Prior to the experiment, the groups with the summary task (1 and 3) were asked to read the text in such a way that they could tell someone else what the General Personal Contribution Regulations mainly are about. The other groups (2 and 4) were asked to search for answers to two specific questions:

(A) Suppose you used a hospital bed for six days. How much would you have to pay?

(B) What is the procedure for paying the General Personal Contribution for the National Health Scheme?

The answer to question (A) could be found on a rather high text level. To detect the answer to question (B), the readers had to go one step deeper and, consequently, had to read the text in somewhat more detail.

In the test after the reading task, for all subjects the summary assignment was combined with answering the specific questions (A) and (B). The moment the assignment was provided was determined by the subject. When he or she clicked on the READY button, which was available on each screen, the display of the text stopped, and the assignment was presented on a new screen.

**Dependent Variables** Product variables measured were:

- the total number of text elements mentioned correctly in the summaries as an indication of the completeness of what readers remember from the text; and
- the scores on questions (A) and (B) as an indication of the quality of the answers readers supply when asked to solve specific problems with the text.

The process variable scored was total reading time as an indication of the efforts readers spend when going through the information.

Evaluation assessments used were subject’s opinions on:

- clarity, difficulty, level of interest, and news value of the text,
- tuning of the text to the task that was presented before reading,
- traceability of the information needed, and
- confidence in the correctness of the results.

**Hypotheses** Earlier, we mentioned a striking contrast between the general claim by Chen and Rada [2] that hypertext systems are especially appropriate for open tasks, and the conclusion from the experiment by Gordon et al. [5] that the hypertext format only leads to poorer achievements in free recall tests than the linear format does.

Because of this discrepancy in the literature, we decided to formulate only two-sided hypotheses (i.e., hypotheses without a specified direction) as to possible differences in the results of our subjects. All in all, we tested nine such hypotheses: six hypotheses pertaining to the dependent product variables “summary completeness” and “quality of answers to specific questions.” The other three hypotheses concerned the dependent process variable “time.” Below we present the three hypotheses concerning summary completeness; the other six hypotheses were stated along the same lines. H1 and H2 below pertain to possible main effects of format and task on summary completeness; H3 pertains to possible interaction effects of the independent variables.

H1: There is a difference in completeness of the summaries written by subjects who are presented with the text in linear format when compared with completeness of the summaries written by subjects who are presented with the same text in hypertext format.

H2: There is a difference in completeness of the summaries written by subjects who are prepared for an open-ended task when compared with completeness of the summaries written by subjects who are prepared for a closed task.

H3: The results obtained are affected by the combination of format and task. There is a difference in com-
pleteness of the summaries written between the four conditions text in linear format/subjects prepared for open-ended task; text in linear format/subjects prepared for closed task; text in hypertext format/subjects prepared for open-ended task; text in hypertext format/subjects prepared for closed task.

**RESULTS**

**Summary Scores** In Table II, the results are presented of an analysis of variance conducted for the total number of text elements mentioned correctly in the summaries.

It turns out that the subjects who read the linear text format, on average, mentioned more text elements \((m = 8.49)\) than did the subjects who worked with the hypertext format \((m = 6.67)\). This difference did not reach statistical significance \((p = 0.073)\), however, and neither did the difference between the average scores in the “open” conditions \((m = 8.06)\) and the “closed” conditions \((m = 6.91)\) \((p = 0.133)\).

The scores for text elements mentioned in the summaries did display a significant interaction effect of format and task \((p = 0.004)\). Post-hoc analyses showed that subjects in the linear/open condition scored significantly better than subjects in the other conditions. The interaction effect was due to the differences in the linear/open condition \((m = 11.44)\) when compared to the linear/closed condition \((m = 6.08) \ (p = 0.025)\) and to the hypertext/open condition \((m = 5.73) \ (p = 0.021)\).

All in all, the summary scores we found certainly do not corroborate Chen and Rada’s [2] assertion about the supremacy of hypertext for open-ended tasks. In this experiment, the linear format proves to be superior, especially when readers are prepared for such a task.

**Answers to Specific Questions**

The total scores for the specific questions display no significant main effects nor any interaction effects. The same applies to the scores for question (B): What is the procedure for paying the General Personal Contribution for the National Health Scheme? However, for question (A): Suppose you used a hospital bed for six days. How much would you have to pay?, the scores did show a statistically significant interaction effect \((p = 0.034)\).

As shown in Table III, the largest differences were found when comparing the hypertext/open condition with the linear/open and the hypertext/closed conditions. The relatively low scores for the hypertext/open condition did not result, however, in statistically significant differences with either of the other conditions. Although the post-hoc analysis did not show significant differences, it is remarkable that the subjects in the linear/open and in the hypertext/closed conditions performed at an optimum: a 100% score.

**Reading Time** As can be seen in Table IV, we found that subjects working with the text in hypertext format spent significantly less reading time \((m = 18.99 \text{ minutes})\) than did subjects who were faced with the linear text \((m = 25.14 \text{ minutes})\). There was no main effect for reading task nor was there a significant interaction effect.

**Opinions** At the end of the online questionnaire the subjects had to fill out, there were seven questions that ascertained the personal opinions of the respondent. The questions are cited in the first

<table>
<thead>
<tr>
<th>TABLE II</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TEXT ELEMENTS MENTIONED IN THE SUMMARIES</strong></td>
</tr>
<tr>
<td>Expected task</td>
</tr>
<tr>
<td>linear</td>
</tr>
<tr>
<td>hypertext</td>
</tr>
<tr>
<td>total</td>
</tr>
</tbody>
</table>

* and y: difference statistically significant \((alpha = 0.05)\).

<table>
<thead>
<tr>
<th>TABLE III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CORRECT ANSWERS TO QUESTION (A): SUPPOSE YOU USED A HOSPITAL BED FOR SIX DAYS. HOW MUCH WOULD YOU HAVE TO PAY?</strong></td>
</tr>
<tr>
<td>Expected task</td>
</tr>
<tr>
<td>linear</td>
</tr>
<tr>
<td>hypertext</td>
</tr>
<tr>
<td>total</td>
</tr>
</tbody>
</table>
column of Table V. The second column indicates which of the alternatives was selected most frequently. Statistically significant differences between the conditions were not found for any of these evaluation questions.

The picture is rather clear. Overall, the subjects were positive about the clarity of the text, about the ease with which they could read and find the relevant information, and about the tuning of the text to the assignment. After reading the text, the subjects did not feel either especially confident nor insecure about their ability to do the assignment. The only clearly negative evaluation score pertained to the text's level of interest: most subjects, in all conditions, found it very boring.

**DISCUSSION**

Two kinds of conclusions can be drawn from this study: conclusions about the effects of medium and task on reading results, and conclusions about the usefulness of the Internet for performing document design experiments.

In this study, we found no main effects of the independent variables on the product variables that were measured (summaries and answers to specific questions). As for the reading process, we did find a significant difference in the time spent by the subjects. The same text was read significantly faster when presented in hypertext format than in linear text format. As for the evaluation of the characteristics of the texts, we found no significant differences between the four conditions.

What we did find, however, were two interesting interaction effects. The first pertained to the number of text elements mentioned in the summaries: subjects in the linear/open condition scored significantly better here than did subjects in other conditions. The other interaction effect was found for one of the specific questions ("How much would you have to pay in a given situation?"). Here the subjects in the linear/open and in the hypertext/closed conditions performed at an optimum, while the subjects in the hypertext/open condition reached the lowest scores.

These results suggest that the display of information in a hypertext format leads to faster reading than a presentation in linear format, without necessarily leading to inferior reading results or less positive opinions about the text. An

<table>
<thead>
<tr>
<th>Expected task</th>
<th>open</th>
<th>closed</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>linear</td>
<td>27.87</td>
<td>22.92</td>
<td>25.14*</td>
</tr>
<tr>
<td>hypertext</td>
<td>20.67</td>
<td>17.31</td>
<td>18.99*</td>
</tr>
<tr>
<td>total</td>
<td>23.61</td>
<td>19.88</td>
<td>21.66</td>
</tr>
</tbody>
</table>

*: difference statistically significant (alpha = 0.05).

<table>
<thead>
<tr>
<th>Evaluation question</th>
<th>Most frequent alternative</th>
</tr>
</thead>
<tbody>
<tr>
<td>How clear was the document? (5-point scale; 1=very unclear; 5=very clear)</td>
<td>4: clear</td>
</tr>
<tr>
<td>How difficult was the text? (5-point scale; 1=very difficult; 5=very easy)</td>
<td>4: easy</td>
</tr>
<tr>
<td>How interesting was the text? (5-point scale; 1=very boring; 5=very interesting)</td>
<td>1: very boring</td>
</tr>
<tr>
<td>How do you evaluate the news value of the text? (5-point scale; 1=very low; 5=very high)</td>
<td>4: high</td>
</tr>
<tr>
<td>How well was the text tuned to your assignment? (5-point scale; 1=very bad; 5=very good)</td>
<td>4: good</td>
</tr>
<tr>
<td>How easy was it to trace the information needed? (5-point scale; 1=very difficult; 5=very easy)</td>
<td>4: easy</td>
</tr>
<tr>
<td>How confident were you after reading the text that you could do the assignment successfully? (3-point scale; 1=confident; 2=neutral; 3=unconfident)</td>
<td>2: neutral</td>
</tr>
</tbody>
</table>
exception, however, arises when readers wish to grasp the main lines of the text, possibly to make a summary. Then the most productive solution still seems to be offering the information in linear format. This finding is contrary to what Chen and Rada claim [2], but it is in line with what Gordon et al. [5] found presuming that in their study it was not only the general interest text, but also the open-ended reading task that led to poor performance in a hypertext environment.

Document design specialists can bear this outcome in mind, when in creating an online document, they have to choose between a typical hypertext or a typical linear format. Only when most readers are expected to use the document to get the big picture is the linear format apparently to be preferred. When the primary tasks that the readers are expected to set for themselves are to find specific information, the hypertext format turns out to be superior.

As to using the Internet for performing document design experiments, we feel that this can indeed be effective and efficient. So far, the literature on using the Internet for experiments in other fields is limited. The results that are reported, however, for instance from a comparison of the same psychological experiments performed in the laboratory and via the WWW, do not give rise to serious doubts on the validity of the findings [7]. It seems clear that using the Internet can have some advantages. One is that subjects are free to choose where they want to participate, and when they wish to do so. That, of course, may make it somewhat easier to find subjects who are willing to cooperate. It should be noted, however, that this freedom of the subjects to participate where and when they wish also reduces to some extent the control that researchers have on the conditions of their experiments.

Another advantage has to do with collecting and processing the data.

Free software, now available on the Web, may be of considerable help in handling incoming data, especially for surveys. Here we refer to Common Gateway Interface programs (CGI) [http://or.psychology.dal.ca/~wcs/hidden/home.html].

Finally, research material which is distributed through the Internet does not have to be removed when the experiment is finished. As long as the material stays on the server, it is accessible. That may be appreciated, for instance, by reviewers of articles that refer to the experiment, by readers of the reports that are published eventually, and by other researchers who wish to replicate the experiment. In our case, all material from the experiment discussed in this article can be found at a Dutch Web site [http://www.kun.nl/bclett/jansenspruijt/welcome.html].

References

Stella Spruijt graduated in Communication Studies at Utrecht University in 1997. She specialized in electronic communication, particularly on the Internet. From 1995 to 1998, she worked as an account manager and consultant for an Internet service provider, advising companies about marketing communications on the Internet. Since 1998, she has been working as a marketing communications manager at BOL Netherlands, the Dutch branch of the international e-commerce initiative by Bertelsmann AG.

Carel Jansen graduated in Dutch Literature and Linguistics at Utrecht University. His Ph.D. dissertation dealt with communication problems between government and citizens. From 1976 to 1989, he worked as an assistant professor at the University of Twente. From 1989 until 1998, Jansen was an associate professor at Utrecht University; from 1993 to 1998, he was also full professor in the field of Technical Communication at Eindhoven University of Technology. He currently holds the chair of Business Communication at Nijmegen University. Most of his research is on the subject of instructional texts. He is author and coauthor of articles and books on topics such as communication problems, government forms, software manuals, and technical writing.