Measuring Reading Behavior in Policy Documents: A Comparison of Two Instruments

Abstract—Techniques for observing selection and reading behavior in professional documents, such as the thinking-aloud and the click-and-read methods, may affect the reading process to be observed. Such so-called reactivity problems complicate the use of these instruments in experimental research and usability testing. If their influence is unknown, any experimental results obtained with these instruments may be caused by the testing method. One way to detect reactivity effects is to compare different instruments in a series of experimental studies. In this initial study, we compared the thinking-aloud method, the click-and-read method, a combination of these two methods, and a silent reading condition. Subjects read and judged a 53-page policy document in one of these conditions. We investigated whether or not different observation instruments caused specific differences in information selection, judgment, and knowledge. Thinking aloud did not cause any differences in the selection of information. Both the thinking-aloud and the click-and-read methods affected the judgment task outcome. Thinking aloud led to many positive and few negative judgments, whereas silent reading led to many negative and few positive judgments. The results for the click-and-read method showed a tendency toward the same effect. Neither method affected the knowledge test results.

Index Terms—Click-and-read method, information selection, policy documents, thinking-aloud method.

Instruments for tracking selection patterns and other reading data in professional documents are indispensable for both practitioners and researchers. An example illustrates this claim.

Suppose you are a technical writer who is designing an online help system for a company that is about to introduce new software to its employees. You intend to include stepwise procedures in your document, as these are generally assumed to enhance task performance (e.g., [1, pp. 227–241]). An article by David Farkas [2] makes you doubt whether or not a standard stepwise presentation is indeed the best format. Farkas describes various alternatives for streamlined steps, such as paragraph style procedures, and he argues that different models for procedural discourse enable different rhetorical strategies. Farkas’ analysis shows that the need to motivate employees to use new software might be a reason for choosing a paragraph format with an implied author who gently encourages the audience by adding extra information to the procedures. On the one hand, Farkas’ analysis is very attractive and convincing. On the other hand, you may fear that people who are used to streamlined steps may be
annoyed by a paragraph format because it may be patronizing or lengthy in their eyes. An experimental study examining how readers process streamlined steps and how they process paragraph style procedures could help show how writers make decisions: What information will be selected or skipped and which problems will be encountered?

In another example: a publisher who wants to convert a paper textbook into a well-adjusted online or CD-ROM version could profit from the same type of observation study. The study would ask what exactly are students’ needs; does the way they use the paper textbook require a certain presentation in the online version? Is it possible to compare students who use the paper version with students who use the new online version of the document? How can new designs of this document be tested?

Many researchers and practitioners in the communication field are interested in similar topics. They want to understand how people select information to support various tasks and how they actually use this information. They test whether new designs or redesigns of documents work, asking questions such as: Will people read the whole document? What parts do they skip for what reasons? This type of research requires advanced instruments to collect qualitative and quantitative data about the way readers select and read information, preferably without affecting the selection and reading process.

It is not at all easy to select instruments that meet these requirements. Two possible instruments are the thinking-aloud method and the click-and-read technique. A third well-known technique is eye movement registration that use modern technology (e.g., Eye Gaze[13]), and are more valid and feasible than the older techniques, we limit our study to the thinking aloud and click-and-read methods. The thinking-aloud method lets subjects solve problems or use a document while verbalizing everything they are thinking. In addition, they read aloud the document parts they are using. An experimenter typically prompts subjects in a nondirective way to think and read aloud if they forget to do so. Subjects’ comments are typed out in protocols to make them suited for further analysis. Elaborate descriptions of the method and its use in several types of research can be found in Ericsson and Simon [3].

Subjects who work with the click-and-read method read a document from a 21 inch computer screen. Typically, the document on the screen is an exact copy of its paper equivalent: it shows two pages at a time, at the same size as a paper document. All text is initially blurred, except for certain parts that enable information selection, such as headings and titles. Subjects select what they want to read on the basis of these readable text parts. They click on the surrounding text block, which then becomes sharp. To select a new text block, they just click on another text block, and the former block becomes blurred again while the new one is brought into focus. Text blocks may be paragraphs, but they could also be longer or shorter passages, depending on specific research interests. All clicks are registered in log files, along with the time of clicking, thus representing a lot of information about the reader’s selection and reading behavior [4, pp. 65–67].

Both the thinking-aloud and the click-and-read techniques are suitable for collecting quantitative data on information selection and information use in professional documents. On the other hand, they also appear to have certain disadvantages that make experimental results subject to discussions about generalizability and reliability. An elaborate overview of the advantages and disadvantages of these techniques with respect to validity, reliability, and feasibility can be found elsewhere [5]. (Table I provides an overview.)

In this paper, we focus on the doubts that have been raised about the so-called reactivity of the instruments.

- Will the characteristics of the instrument affect experimental results on information selection and reading processes, or on performance on judgement tasks and carrying out instructions?
- Does the very processes of thinking and reading aloud lead readers to select and process more information than readers who read silently?
- Will a reader who uses the click-and-read technique be more selective than a reader with a paper text because the blurred text parts enhance this selectivity?

Currently, there are no definite answers to these questions. The thinking-aloud method has been subject to many investigations into its supposed reactivity (e.g., [6], [7]; many of them have been reported in [3]). The question of whether or not the thinking-aloud method affects the reading process can be answered by the general conclusion that reading and thinking aloud has proved to take more time than silent reading, but it does not alter or disturb the reading process [8, pp. 306–307]. This claim is supported by evidence that performance on an additional task, for instance a problem-solving task or a questionnaire, did not differ (and thus it is deduced that the preceding reading process did not differ either). Except for a registration of the total
time subjects spent reading and thinking, and except for a number of comparisons between eye movement registration and thinking aloud (e.g., [9]), actual differences between reading processes have not been measured.

The thinking-aloud method may not only affect the selection and reading process, but it may also affect the tasks that subjects want to perform. Ericsson and Simon’s [3] overall conclusion is that the thinking-aloud method only affects task outcome

• if this task or the instructions require subjects to verbalize thoughts that they do not immediately have available in an oral form (e.g., visible or motorial information);
• if the verbalization process requires intermediate processing and could change the course and structure of the thought process;
• if the thinking-aloud instructions elucidate extra inferences or reader-generated information.

However, for a number of reasons, we should interpret these conclusions carefully if we want to apply the thinking-aloud method to professional documents. First, due to the environment of the experiments, the conclusions are typically not based on lengthy professional documents, realistic tasks, or selective readers, but on short, relatively simple texts that were read linearly and completely. In realistic reading tasks, however, readers have very specific goals in mind. They are selective, and they use the document to carry out a task outside the document. Thinking aloud may have different effects under those circumstances. It may, for instance, cause readers to generate more information because readers may link the document content more actively and more explicitly to their task or previous knowledge than they would otherwise do. The earlier evidence thus may not be directly applicable to the type of document and reading process that professional communicators would like to test or investigate.

Second, the indirect evidence, based on outcome measures, that reading and thinking aloud processes do not differ from silent reading processes should be reconfirmed by comparing selection and reading data of the thinking-aloud method to selection and reading data collected by means of other instruments that now have become available, such as modern techniques for eye movement registration or the click-and-read method.

The click-and-read method is much newer and has not yet been evaluated. Empirical evidence about possible reactivity effects should be collected for the same reasons, that is, to enable researchers and testers to make use of the technique in a well-considered way.

In summary, in order to decide which observation instruments can be selected for investigating and testing professional documents and their readers, it is necessary to know whether or not reactivity effects of both the thinking-aloud method and the click-and-read method will occur when they are used in professional situations. We planned a series of experiments

<table>
<thead>
<tr>
<th>Table I</th>
<th>Overview of the Characteristics of the Thinking-Aloud and Click-and-Read Instruments for Observing Reading and Selection Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of Instrument</td>
<td>Brief Description</td>
</tr>
<tr>
<td>Thinking and Reading Aloud</td>
<td>Subjects read a document and perform a task while thinking and reading out loud; the thinking aloud protocols are analyzed.</td>
</tr>
<tr>
<td>Click and Read</td>
<td>Subjects select information blocks from document on a large computer screen by clicking on those blocks. Mouse clicks and time spent using information are recorded and analyzed.</td>
</tr>
</tbody>
</table>
to investigate this. As a first step, we studied the effects of the thinking-aloud method and the click-and-read method within the same type of document and corresponding task: policy documents and judgment tasks. In follow-ups, we intend to investigate other professional reading tasks, such as carrying out instructions. The following are specific questions we have studied in this experiment:

(1) Does the thinking-aloud method in a professional reading situation, when compared to silent reading, lead to differences in information selection? Specifically, will it lead to the selection of fewer information items within a given amount of time, or will the time spent per selected information item be longer? This question can be answered by comparing the click-and-read method to a combination of the thinking-aloud and the click-and-read methods. By using the click-and-read method, the same selection measures can be taken with and without thinking and reading-aloud assignments.

(2) Do the thinking-aloud and the click-and-read techniques lead to differences in performance in both judgment and knowledge tasks, especially in comparison with unobserved reading situations?

**METHOD**

The study was conducted at the usability lab of Delft University, where 51 subjects were presented with an official 53-page Dutch policy document entitled “Studiefinanciering,” a policy report about grants provided by the Dutch government for scholarships. In The Netherlands, every university and college student receives a government grant. This policy document contained proposals for new grant systems. The subjects’ task was to promote students’ interests through formulating their opinion about the proposals in front of the Minister of Education (represented by the experimenter). They were told that their judgment should specifically concern one main issue: do the proposals allow every young person, especially those from low-income families, to become a student, or in other words, gain access to academic education? We will refer to this as the “accessibility of education” issue. In addition, subjects are asked to present their judgement about a second issue: plans of the government for combining work and study in the near future. As all subjects found these plans to be too general and too vague, these data were not suitable for analysis. However, this second goal probably caused subjects to be even more selective and goal-oriented than the time restriction, similarly to realistic situations.

Subjects were told that they would have one hour to read information, take notes, and formulate their judgments about the “accessibility of education” issue. They were told to carry out the task the way they would normally do. Although this instruction was neutral and did not mention any reading strategies, most subjects were immediately aware that selecting information would be necessary, as the document contained far too much information to be processed in one hour. Moreover, a lot of information was not relevant for the “accessibility of education” issue. This way, we intended to approach a realistic situation of a reader with a specific goal in mind, who chooses a strategy based on circumstances (number of pages, available time) and experience (what are they used to doing in these situations?).

After one hour, subjects were requested to present their judgment as if they were talking to the Minister of Education. These judgments were recorded on tape. Afterwards, they completed a questionnaire about factual information in the report. This questionnaire represented the knowledge variable.

Effects of thinking aloud on information selection (question 1) were tested by comparing two conditions: the click-and-read method, and a combination of the thinking-aloud and the click-and-read methods, called the click-and-think method. This new combination enabled us to see whether thinking aloud actually affects the amount of information selected in a given amount of time, and the amount of time spent per selected information item. The click-and-read method records those measurements in both conditions.

Furthermore, judgments were typed out and scored for value (positive or negative judgments), and factual knowledge questionnaires were collected and scored for correctness. These two outcome measures were compared over four conditions:

(1) Thinking Aloud
   (reading and thinking aloud with a paper text);
(2) Click and Read
   (reading silently while clicking on text on a computer screen);
(3) Click and Think
   (combination of 1 and 2: clicking text on a computer screen while thinking and reading out loud);
(4) Silent Reading
   (reading silently with a paper text).

Subjects were randomly divided over the conditions. Unfortunately, for practical reasons, we were not able to run equal numbers of subjects in all conditions: the silent reading condition contained fewer subjects than the other three (see Results section). Yet, after having analyzed the data, we decided there was enough reason to include this condition in the results as it showed relevant and consistent differences, even with small sample size.
If the conclusions for nonprofessional reading tasks also hold for the type of judgment task we use in this study, the following results might be expected:

- Within the one-hour time limit, thinking aloud subjects will select fewer text blocks and spend more time per text block than subjects who read silently (because if time were unrestricted, thinking-aloud subjects would need more time to complete their task).
- Performance on the judgment task (numbers of positive and negative judgments) and the knowledge task (number of correct questionnaire answers) may or may not differ between the four conditions. On the one hand, subjects do not have to verbalize information that is not available to them in a verbal encoding, nor is there any reason to assume that the click-and-read method will affect subjects' judgments. On the other hand, thinking aloud itself (not the instructions) may elucidate inferences, intermediate processing, and reader-generated thoughts in this context, which in turn may lead to effects on the judgment task outcome.

**Materials: Document and Instruments** The document was an official government document that had been distributed in 1996, two years before the experiment was conducted. None of the subjects had seen or read the document before. The document contained a table of contents, a summary, four chapters, and fifteen appendices, up to a total of 51 pages.

In the **thinking-aloud condition**, subjects were presented with a paper version of the report. They were requested to read aloud and think aloud throughout the session, and also when they were taking notes. When subjects fell silent for a minimum of 15–30 seconds, the experimenter would remind them in a nondirective way to read and think aloud from the control room via a microphone.

In the **click-and-read condition**, the report was presented on a 21-inch computer screen. On screen, subjects saw an identical copy of the text that subjects in the thinking-aloud condition saw on paper, two pages at once. The text in the click-and-read version of the report was blurred, except for all first sentences of a paragraph, all titles and headings, and all words that were somehow marked in the original text (e.g., underlined, italic, and boldface). The text was not heavily blurred, but just enough to make the original text too difficult to read; earlier studies with the click-and-read method and pilots of this study had shown that this way of blurring was enough to create an impulse to click, and it stayed as close to the original text as possible.

Subjects could select information on the basis of nonblurred headings, first sentences, and marked words and click on the accompanying text part to bring it into focus. One paragraph at a time was sharpened. As soon as another paragraph was clicked, the sharpened paragraph became blurring again and a new one was sharpened. All mouse clicks were registered in a log file, which contained the paragraph number and paragraph characteristics, along with the time of clicking.

Time data in the log file were used to calculate information using time. Note that this is not the same as reading time. During the time the information is on the screen, subjects not only read, but also think and elaborate about what they just read, and they take notes. The assumption underlying the information using times is that during these activities and during the preparation of the judgment, readers actually use the information.

In the **click-and-think condition**, subjects combine the materials of the click-and-read condition with the instructions for the thinking-aloud condition. They use the exact same text version as subjects in the click-and-read condition, but in addition, they read aloud and they think aloud.

In the **silent-reading condition**, subjects read the paper version of the text, in the same room and under the same conditions as all other subjects, except that no observation instrument was used.

**Procedure** 51 students of Delft University of Technology participated in the experiment. They were paid a compensation fee. Subjects participated one at a time.

All subjects first practiced on the instruments they were about to use in the main task. In the practice task, they answered a set of questions after looking up the answers in a brochure about taxes. While carrying out this practice task, they had to read and think aloud, click to find the necessary information, or do both, depending on the observation instrument in the main task. The experimenter was present during the practice task to give instructions.

The main task was presented to subjects on a sheet of paper. In addition, it was verbally explained to them by the experimenter. Subjects had to pretend they had to replace a housemate: a member of the National Student Union who had been invited to the Minister of Education's office to comment on the official policy report about the future of the grant system, on behalf of the students. The housemate calls the subject two hours before the appointment because she is stuck in traffic and cannot make it in time. The subject is requested to replace his housemate and take the report from her desk to formulate an elucidated opinion about the accessibility of higher education in The Netherlands, all within one hour.
While the subject read his 
assignment, the experimenter left 
the room. The subject started his 
task after the experimenter had 
given a sign through a microphone 
in the control room, where the 
subject's computer screen and a 
video of the subject were displayed. 
The click-and-read computer 
logged the clicks. Video and audio 
recordings were made to register 
thinking-aloud protocols.

Subjects were allowed to take 
notes during the reading process 
and use those notes during the 
presentation of their judgments. 
After 45 minutes, the experimenter 
in the control room warned 
subjects that they had 15 minutes 
left. After 55 minutes, they 
were warned that they had 5 
minutes left. After one hour, the 
experimenter went to the testing 
room, took away or closed the 
document, and asked the subject 
 to present his or her judgment 
about the report as if he or she 
were talking to the Minister of 
Education.

Finally, subjects completed a 
questionnaire. It contained 25 
multiple-choice questions about a 
variety of facts mentioned in the 
report, both on the “accessibility 
of education” issue and on other 
issues that were irrelevant for the 
judgment task. Subjects had to 
remember the information in the 
text in order to choose the correct 
answers: they were not able to use 
the document again at this point.

**RESULTS**

**Question 1: Does the thinking-aloud method in a 
professional reading situation lead to the selection of fewer 
information items within a given 
amount of time, or will the time spent per selected information 
item be longer?**

By comparing data collected by 
means of the click-and-read and 
the click-and-think methods, 
we could find out whether 
or not thinking and reading 
 aloud takes up more time than 
silent reading in professional 
situations. The log files generated 
by the click-and-read technique 
represented the number of 
selected information items. The 
average number of seconds people 
needed to read one information 
item represented the time per 
information item. Table II presents 
an overview of the time and 
selection data involved in the 
comparison.

The total time subjects spent 
using the information and 
preparing their judgment was 
limited to one hour to approach 
a situation professional readers 
often find themselves in. Most 
subjects needed the full hour to 
complete their task. There were no 
significant differences in the total 
amount of time spent on using 
information and preparing for the 
judgment.

If thinking aloud caused a 
delay in this situation, as it did 
in previous experiments, we 
would expect to find that the 
click-and-think subjects selected 
fewer items, or that they spent 
more time per selected item 
than the click-and-read subjects. 
None of these expectations 
were confirmed by the data.

No significant differences were 
found between the click-and-read 
condition and the click-and-think 
condition on any of the selection 
or time measurements, neither for 
the complete text nor for those 
parts that were relevant for the 
"accessibility of education" issue. 
Except for the measurements 
in Table II, we also tested for 
differences in the number of clicks 
and time spent on summary, 
introduction, content page, 
conclusions, and tables. No 
differences were found. The 
paragraphs that were relevant in 
this respect had been (invisibly) 
coded in the click-and-read log 
files and could thus be analyzed 
separately.

**Question 2: Do the thinking-aloud and 
click-and-read techniques lead 
to differences in performance on both judgment and knowledge 
tasks?**

Task performance in this study 
consisted of two parts: the 
judgment task (verbalization of a 
judgment on the proposals in the 
policy report) and the knowledge 
task (number of correct answers to 
a multiple-choice questionnaire).

We tested for differences between 
four conditions. Table III shows 
how many subjects in every 
condition were against (negative 
judgment) or in favor of (positive 
judgment) the proposals in the 
report.

Subjects in the thinking-aloud 
condition tend to judge the 
proposals positively. Subjects

---

**TABLE II**

**Mean Number of Clicks (Selected Items) and Mean Time 
(Seconds) Spent Reading and Using Information from the Policy Report, in the 
Click–And–Read and Click–And–Think Conditions. (Standard Deviations in Parentheses)**

<table>
<thead>
<tr>
<th>Name of Instrument</th>
<th>Total Time (limit 1 hour)</th>
<th>Total No. Selected Items</th>
<th>Time per Item</th>
<th>No. Items Selected on ‘Accessibility’ Issue</th>
<th>Time on ‘Accessibility’ Issue</th>
<th>Time per Item on ‘Accessibility’ Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click and Read (n = 16)</td>
<td>3562 (527)</td>
<td>101 (41.6)</td>
<td>47 (39.7)</td>
<td>37 (13.3)</td>
<td>1281 (462)</td>
<td>39 (27.3)</td>
</tr>
<tr>
<td>Click and Think (n = 14)</td>
<td>3490 (231)</td>
<td>99 (42.9)</td>
<td>42 (19.2)</td>
<td>40 (16.2)</td>
<td>1380 (275)</td>
<td>40 (18.7)</td>
</tr>
</tbody>
</table>
in the silent reading condition show the opposite tendency; they verbalized negative judgments relatively frequently. Click-and-read subjects (whether they had to think aloud or not) judged positively and negatively in more or less equal numbers. The differences between the conditions appeared to be significant ($\chi^2 = 10.29$, df = 3, $p < 0.05$).

Paired tests showed differences between the thinking-aloud condition and the silent-reading condition ($\chi^2 = 9.37$, df = 1, $p < 0.01$) and between the thinking-aloud and the click-and-think conditions ($\chi^2 = 5.04$, df = 1, $p < 0.05$). There was a tendency toward a significant difference between the click-and-read and the silent-reading conditions ($\chi^2 = 4.02$, df = 1, $p < 0.06$). Both the thinking-aloud and the click-and-read methods affected the judgment results.

Factual knowledge was the second performance variable. To test for differences in knowledge, the 25-item questionnaire was scored for correctness. We distinguished between questions that were relevant and irrelevant for the “accessibility of education” issue. Table IV shows an overview of these results.

The instruments appeared not to have any effect on knowledge; subjects in the four conditions approximately gave the same amount and the same percentage of correct answers, irrespective of the relevance of the question to the specific reading goal (the “accessibility of education” issue). Statistical tests showed no significant differences, except for one. Overall, the percentage of correct answers to relevant questions was higher than the percentage of correct answers to irrelevant questions ($F(1, 47) = 28.86, p < 0.001$), which is a natural consequence of the subjects’ focus of attention, but no effect of condition could be found. Factual knowledge does not seem to be affected by different observation instruments.

**TABLE III**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Name of Instrument</th>
<th>No. of Positive Judgments</th>
<th>No. of Negative Judgments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Thinking Aloud (n = 13)</td>
<td>11</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>2 Click and Read (n = 15)</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>3 Click and Think (n = 14)</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>4 Silent Reading (n = 7)</td>
<td>1</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

**CONCLUSIONS AND DISCUSSION**

In this study, thinking aloud does not lead to a smaller number of selected information items, nor to less expected time spent per information item, whereas such differences might be on the basis of previous studies that report delays in thinking-aloud tasks. This finding may be explained by the difference between our experimental task and many other reading and problem-solving tasks that were used in thinking-aloud studies. The reading task in this study takes a lot of time and concentration, and is highly selective. Thinking aloud may obstruct readers only in the beginning, when they have to get used to this somewhat artificial behavior, but this effect may disappear once they get more engaged in their judgment task, a reading task that is similar to many other real-life reading tasks. Apparently, thinking aloud, selecting and reading text parts, and preparing for the judgment task become parallel tasks once the subject is really engaged in the assignment.

The results of this study lead to the conclusion that observation instruments affect the results of judgment tasks if the task is long and time-restricted, and if it requires selective reading. At the same time, observation instruments do not seem to affect knowledge. These findings raise several questions.

For instance, why is there an effect of instrument on judgments and not on knowledge? First, we should exclude the possibility of “ceiling effects” in the questionnaire results. The average number of correct answers was between 10.6 and 11.3. If all correct answers were the same (for instance, to the relevant questions), then the questionnaire may simply not have been able to discriminate between the conditions. This explanation, however, can be excluded because the analysis of the questionnaire results showed that only 4.3 correct answers out of an average of 11 were answers to relevant questions for the “accessibility of education” issue. Other correct

**TABLE IV**

<table>
<thead>
<tr>
<th>Condition</th>
<th>Name of Instrument</th>
<th>No. of Correct Answers</th>
<th>Correct Answers on Questions about ‘Accessibility’ (n=8)</th>
<th>Correct Answers on Other Questions (n=17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Thinking Aloud</td>
<td>10.9</td>
<td>4.6 (57.0%)</td>
<td>6.4 (37.5%)</td>
<td></td>
</tr>
<tr>
<td>2 Click and Read</td>
<td>11.3</td>
<td>4.2 (52.7%)</td>
<td>7.1 (41.6%)</td>
<td></td>
</tr>
<tr>
<td>3 Click and Think</td>
<td>10.6</td>
<td>3.9 (48.2%)</td>
<td>6.7 (39.5%)</td>
<td></td>
</tr>
<tr>
<td>4 No Instrument</td>
<td>11.0</td>
<td>5.0 (62.5%)</td>
<td>6.0 (35.3%)</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>10.9</strong></td>
<td><strong>4.3 (54.2%)</strong></td>
<td><strong>6.6 (38.9%)</strong></td>
<td></td>
</tr>
</tbody>
</table>
answers were spread over the questionnaire. There were no clear patterns of correct or wrong answers.

An alternative explanation is that acquiring knowledge by reading is a more automated process than generating an opinion about the subject matter in a document. Our subjects were students; they were intelligent and skilled readers who were used to extracting and processing information from long texts, even when working memory load is high because they are hearing or doing something else at the same time. In contrast to acquiring factual knowledge, judging the subject matter of a document with a specific goal in mind may not be an automated process, especially when the subject matter is not fully familiar. The results on the knowledge test in the first three experimental conditions may not differ from those in the silent reading condition simply because the instruments did not disrupt the automated process of knowledge acquisition, whereas they did disrupt the judgment task.

The results also raise a second question: how can the effects of the thinking-aloud method and the click-and-read method on the judgments be explained? Thinking aloud leads to more positive and fewer negative judgments than the silent-reading condition. An explanation begins by recognizing that thinking and reading aloud in judgment tasks causes more reader-generated elaborations than silent reading because of the relatively conscious way of verbalizing both judgments and arguments during the reading and preparation stage. As soon as they fall silent, subjects are reminded to keep reading and thinking aloud.

By instructing subjects to think aloud, we more or less force them to process the information intensively, while evaluating arguments and constantly verbalizing what they think (the so-called central route). In the silent reading conditions, subjects may have processed more information along the peripheral route, thereby paying attention to nonargumentative characteristics of the text [10].

Elaborations, especially reader-generated elaborations, are known to affect attitudes [10], [11], but that still does not explain the value of the judgments in different conditions. Why would thinking aloud radically lead to more positive and fewer negative judgments instead of the other way around? We can think of two explanations for this phenomenon. First, the silent reading group may have contained more subjects with a positive or tolerant attitude toward the document content. However, as we assigned subjects randomly to each condition, we think we can practically rule out this possibility. Another possible cause for the effect may be social pressure. Suppose the participants’ initial attitudes toward the document were negative. This is a realistic presupposition since the document contained plans about students’ future budgets and Dutch governmental plans about these issues are not known for their generosity. In this line of reasoning, silent readers feel negative about the document at the start of their task. During the process they at least partly base their judgement on peripheral cues (such as source characteristics, formal language) and thus strengthen their negative opinion; if they evaluate arguments, and they feel negative about them, they do not have to mention that aloud. In contrast to this view, thinking-aloud subjects (who in our hypothetical situation also start with a negative attitude) may feel obliged to work along the central route. Moreover, they may feel embarrassed if they constantly verbalize negative thoughts. They may formulate well-balanced comments because they know someone is recording what they say and thus end up with a more positive judgment.

What complicates this line of reasoning is that the click-and-read condition, in which subjects work silently also differs from the silent-reading condition, although this difference only shows a tendency toward significance. Though less pronounced, this difference also implies more positive judgments in the click-and-read condition.

The thought may come to mind that what we see here are simply testing effects; subjects are more positive because they are tested. However, the same holds for subjects in the silent-reading and click-and-think conditions, and thus testing effects are unlikely. All subjects participated in the same room with the exact same setting of tables, computer monitors, and video cameras.

The effect caused by the click-and-read method is probably a different effect than the effect caused by the thinking-aloud method; there is no specific reason here why reader-generated elaborations would occur more often than in the silent-reading condition. Starting from the same assumption as in our explanation of the thinking-aloud effect, i.e., that the text initially evokes negative attitudes, we suspect that the click-and-read method leads to more positive judgments because it makes the document more accessible for readers. In the click-and-read document, all first sentences of all paragraphs were readable and other information was blurred. This way, the click-and-read method may strongly support the selection and judgment-forming processes, thus avoiding the negative opinions evoked by the language or structure of the document, like in the other silent-reading condition. In other words, the click-and-read method may provide information selection support that is evaluated very positively. This support may
serve as an important peripheral cue, which also causes a difference with the silent-reading condition. These results mean that researchers and usability testers should be reluctant to use the thinking-aloud method and the click-and-read method in judgment tasks or other persuasive tasks, as the instruments are likely to affect the task outcome. For research questions or testing goals that only require quantitative measurements of information selection or knowledge measurements, and not judgment tasks, the instruments can be used; they do not seem to affect selection or knowledge measurements.

It goes without saying that these results do not automatically hold for other professional documents and tasks. Different document types are known to lead to different reading processes [12], and different reading processes may be affected by research instruments in different ways. The results of this experiment do not predict the effects of the instruments in a reading-to-do task, such as carrying out instructions for a complex device, or a reading-to-learn-to-do-task, such as learning to work with a software program. In a series of follow-up studies and analyses, this experiment should be repeated, replicated, or specified for judgment tasks and other types of tasks.

ACKNOWLEDGMENT

We would like to express our gratitude for very valuable comments by anonymous reviewers and the Guest Editors on an earlier draft of this article.

REFERENCES


Nicole Ummelen is an assistant professor at the department of Business Communication at Tilburg University in The Netherlands. Her research involves information selection and effects in software manuals, information use, and effects in electronic documents, and a comparison of instruments for registering information selection.
Rob Neutelings is currently managing a service unit in the Faculty of Architecture at Delft University of Technology. His research projects involve an analysis of the way policy documents are read by professional readers and a comparison of instruments for registering information selection.