

The role of thinking-aloud instructions and prior domain knowledge in information processing and source evaluation during Web search

Yvonne Kammerer (y.kammerer@iwm-kmrc.de)

Knowledge Media Research Center, Schleichstr. 6, 72076 Tuebingen, Germany

Peter Gerjets (p.gerjets@iwm-kmrc.de)

Knowledge Media Research Center, Schleichstr. 6, 72076 Tuebingen, Germany

Abstract

This paper examines the impact of thinking-aloud (TA) instructions as well as of individuals' prior domain knowledge on information processing and source evaluation during Web search on a health-related topic. With regard to TA instructions, prompted instructions that entailed evaluation prompts (as used in some previous Web search studies) were compared to neutral instructions (in line with the standards defined by Ericsson & Simon, 1993) and to a silent condition. To measure participants' (N = 44) information processing and source evaluation we used a rich multi-method approach including eye-tracking methodologies, log file data, and verbal protocols. Results indicate that prompted TA instructions as compared to neutral instructions significantly increased participants' verbal reflections on information quality and on structural aspects of Web pages, given that participants possessed at least a moderate level of prior domain knowledge. In addition, prompted instructions resulted in less linear viewing sequences on the search engine results pages than the silent condition. Finally, the higher participants' prior domain knowledge the more intensely they scrutinized the search results presented by the search engine and the smaller were their average pupil sizes, which indicated lower cognitive load. The significance of the results is considered in light of methodological as well as educational implications.

Keywords: Web search; source evaluation; prior domain knowledge; thinking-aloud instructions; eye tracking

Introduction

In recent years, the World Wide Web (WWW) has evolved into a major information resource offering easy access to billions of Web pages on almost any topic. However, as anyone can publish virtually any information on the WWW, the quality of Web pages can vary widely. That is, misleading and low-quality Web pages, for example, in the field of medicine and health care, are as common as those providing neutral, high-quality information. Hence, to avoid the selection and use of doubtful or even false information, it is important that Web users themselves critically evaluate the quality of information they retrieve from the Web.

Previous empirical findings about Web users' source evaluation as indicated by verbal reports, however, are inconclusive. Whereas some studies indicate that Web users mainly evaluate search results and Web pages only based on the topical relevance or their ranking in the search engine results page (e.g., Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010; Savolainen & Kari, 2006), others suggest that Web users are also concerned to a substantial extent

about information quality (i.e., the accuracy, authority, objectivity, or currency of information; e.g. Rieh, 2002; Tombros, Ruthven, & Jose, 2005). According to Tombros et al. (2005) also the structure of Web pages (i.e., the clarity of the Web page or the organization of the information therein) is evaluated extensively.

The aim of the present paper was to identify potential reasons for the divergent findings. Specifically, we examined the impact of the instructions used in the studies as well as of participants' prior domain knowledge on information processing and source evaluation during Web search, using a rich multi-method approach including eye tracking methodologies, log file data, and verbal protocols.

The Role of Prompted Thinking-Aloud Instructions

One reason for the divergent findings might be a methodological one, namely that in the studies by Rieh (2002) and Tombros et al. (2005) participants were instructed beforehand to explain what criteria they used to evaluate Web information (Tombros et al., 2005) or to select good or credible information during Web search and to explain their evaluation processes in the form of postsearch interviews including specific evaluation-related questions (Rieh, 2002).

According to the seminal work by Ericsson and Simon (1993) and the meta-analysis by Fox, Ericsson, and Best (2011), however, procedures for verbal reporting are only nonreactive (i.e., do not alter thought processes and task performance) when instructions to think aloud are given in a neutral way, by instructing participants to verbalize their thoughts per se. In contrast, procedures that entail describing or explaining thoughts and actions – as it was the case in the studies by Tombros et al. (2005) and Rieh (2002) – are significantly reactive, altering participants' course of cognitive processing and leading to higher task performance than silent conditions. Hence, the instructions used by Tombros et al. (2005) and Rieh (2002) might have increased participants' awareness of the necessity of critically evaluating the information retrieved during Web search. Indirect evidence for this assumption comes from the studies by Hargittai et al. (2010) and Savolainen and Kari (2006) that used neutral thinking-aloud (TA) instructions and only found few utterances related to information quality or structural aspects of Web pages. A first aim of the present study, thus, was to further prove this assumption by directly comparing participants' verbal utterances when given TA

instructions entailing explicit prompts to explain what criteria they used to evaluate search results and Web pages (hereinafter referred to as prompted thinking-aloud) with neutral TA instructions (in line with Ericsson & Simon, 1993).

Furthermore, previous research that directly compared participants' information processing while thinking-aloud (with neutral or prompted instructions) to a silent condition, showed that with prompted TA instructions that entailed explanation prompts participants who had to solve a series of information tasks on a Website explored the information more extensively than when they worked in silence (Hertzum, Hansen, & Andersen, 2009). That is, in the former case participants showed a more distributed visual exploration of the screen (i.e., they searched or scanned more across the screen), scrolled more frequently within Web pages, and navigated to more Web pages than in the silent condition. Between the neutral TA instructions and the silent condition, in contrast, only few marginal differences were found. With the present study we aimed at expanding these findings to a Web search scenario including multiple Websites being accessed via a search engine results page (SERP), with participants either receiving prompted TA instructions (including evaluation prompts), neutral TA instructions, or working in silence.

The Role of Prior Domain Knowledge

Referring to the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986) from persuasion research, Metzger (2007) postulates that the extent to which Web users engage in source evaluations is dependent both on their motivation (which might be increased by evaluation prompts), but also on their ability (e.g. their prior domain knowledge). In this regard it should be noted that participants in Rieh's (2002) study reported to use their prior domain knowledge to evaluate information quality, indicating that they possessed a certain amount of prior knowledge on the search topics. Importantly, previous research has not only shown differences in source evaluations between domain experts and novices, but also between groups of novices varying in their level of prior domain knowledge. For example, a case-study by MaKinster, Beghetto, and Plucker (2002) that investigated undergraduate students' Web search on a complex science topic through postsearch interviews that entailed evaluation-related questions indicated that students with moderate domain knowledge scrutinized search results more thoroughly, by examining the titles, the page excerpts, and the URLs of the search results, than low-knowledge students. In addition, Bråten, Strømsø, and Salmerón (2011) found that when reading multiple documents dealing with a science-related topic undergraduates with low knowledge on the subject matter trusted the different documents to the same extent irrespective of the type of source, whereas students with higher knowledge judged an article from a company with vested interests about the addressed issue as less trustworthy than the other documents. According to Bråten et al. (2011) a possible explanation for the

undifferentiated or lacking trustworthiness evaluations of low-knowledge readers is that they have to invest more cognitive effort in comprehending the content of the documents than readers with higher domain knowledge. As a consequence, low-knowledge readers might have less cognitive resources available to engage in evaluations that go beyond content. Thus, another central aim of our study was to further examine the effects of prior domain knowledge on information processing and source evaluation during Web search as well as potential interactions between prior domain knowledge and TA instructions.

Hypotheses of the Present Study

Given the theoretical considerations and prior empirical findings our hypotheses were as follows:

First, with respect to the evaluation of information quality and structural aspects of Web pages as indicated by verbal utterances, we hypothesized that prompted TA instructions would increase the number of respective verbal utterances during Web search as compared to neutral instructions, but only when participants' possessed a certain level of prior knowledge (cf. Metzger, 2007) (*H1*). Low-knowledge participants' quality-related and structure-related verbal utterances should not be increased due to participants' lack of cognitive resources (cf. Bråten et al., 2011).

Second, based on the findings by Hertzum et al. (2009) we hypothesized that prompted TA instructions as compared to a silent condition would result in a more distributed scanning behavior on SERPs (*H2*) and a significant increase in Web pages selected from the SERPs during Web search (*H3*). In contrast, neutral instructions should only result in slight changes in regard to these measures as compared to a silent condition.

Third, according to the case-study results of MaKinster et al. (2002) we hypothesized that the higher participants' domain knowledge the more thoroughly they would scrutinize the search results, as indicated by longer total fixation times (*H4*). Finally, based on the theoretical assumption provided by Bråten et al. (2011) we hypothesized that the lower participants' domain knowledge, the more cognitively demanding the Web search would be for them, which should result in larger mean pupil sizes (*H5*) which is known to be an indicator of increased cognitive load (for review e.g. see Hyönä, Tommola, & Alaja, 1995; Wang, 2010).

Method

Participants and Design

Participants¹ were 44 undergraduates (18 male, $M = 25.02$ years, $SD = 3.68$) from different majors at a large German

¹ Note that parts of the empirical research reported in this paper have been published in Gerjets, Kammerer, and Werner (2011) who compared the two thinking-aloud groups (irrespective of prior knowledge). The data of the silent group were gathered at the same time as the data of the two thinking-aloud groups.

university, who were rewarded with either course credit or payment. Participants had normal or corrected to normal vision. They reported to use Google as their primary search engine and judged their computer- and Web search experience and skills between intermediate and high ($M = 3.50$, $SD = 0.67$, 4 items, on a scale from 1 = very low to 5 = very high, Cronbach's $\alpha = .78$).

Thinking-aloud condition served as a between-subjects factor, with participants being randomly assigned to a neutral TA condition, a prompted TA condition, or a silent condition. Neutral instructions to think aloud were worded in line with the standards described by Ericsson and Simon (1993), that is "Please think aloud during your Web search, that is, *verbalize everything that comes to your mind.*" In contrast, the prompted TA instructions were similar to the instructions used, for instance, by Tombros et al. (2005) or Rieh (2002) including evaluation prompts, that is "Please think aloud during your Web search, that is, *mention the evaluation criteria you apply to select search results and to assess Web pages.*" In the silent condition participants performed the task silently.

There were no differences between the three conditions regarding participants' age ($\chi^2(2, N = 44) = 2.24, p = .33$), gender ($F < 1$), or computer- and Web search experience and skills ($F < 1$).

As a second factor participants' self-reported *prior knowledge* on the subject matter of the task (i.e., diets and nutrition) was assessed (see 'Measures' for details) and used as a continuous between-subjects factor.

Tasks and Web Materials

Participants were given the task of seeking information on the WWW about two competing weight loss methods, namely low carb(ohydrate) diets and low fat diets, in order to give informed advice to a fictitious overweight friend who wants to lose weight by changing her diet.

To complete the experimental task, which was limited to 20 minutes, participants were provided with three prefabricated Google-like SERPs (with ten search results each), retrieved for the search terms "low fat", "low carb", and "low carb versus low fat". Participants were instructed to access all three SERPs during their Web search. They could access all 30 Websites corresponding to the list of search results presented on the SERPs, but were not allowed to generate new search results by changing the search terms. All search results and Websites were relevant to the search topic in regard to the content of information provided, but differed with regard to the type of sources including Websites provided by scientific institutions, journalists, industry and companies, and laypeople.

Measures

Self-reported prior domain knowledge To assess participants' prior knowledge on diets and nutrition, participants were administered a questionnaire with eight statements that had to be rated on five-point scales (1=totally disagree; 5=totally agree); example items are "I

know more about diets and nutrition than my family and friends" or "I can describe the concept of low carb diets". Cronbach's alpha was .87 for the eight items. The range of participants' prior domain knowledge was 1 to 4.38 ($M = 2.48$, $SD = 0.70$). There were no differences between the three experimental conditions regarding participants' prior knowledge on diets and nutrition ($F < 1$).

Dependent variables

Participants' thinking-aloud protocols (in the prompted and neutral TA conditions) were segmented at a small grain size: Each sentence or utterance preceded and followed by a pause was considered a separate segment. According to the studies by Rieh (2002) and Tombros et al. (2005) utterances were coded as *information quality* when participants on SERPs or Websites reflected on whether (or not) the information is good, valid, credible, or current, or the source provides trustworthy, reliable, or official information, and as *structure of Web pages* when participants addressed the clarity of a Web page or the organization of the information therein (cf. Tombros et al., 2005). Besides, utterances addressing the *topical relevance* (i.e., utterances on SERPs or Websites about whether or not the information matches with the search topic) were coded (e.g., Savolainen & Kari, 2006). Two raters familiar with the search task and the Web materials scored 30% of the protocols. Interrater reliability computed on this subsample of protocols yielded a Cohen's kappa of .76. Disagreements were resolved through discussion between the raters. One rater scored the remaining protocols.

In addition, for participants of all three conditions their viewing and selection behavior on the SERPs were analyzed. Eye movements and mouse clicks were recorded by a 50 Hz Tobii 1750 remote eye-tracking system supported by the software ClearView 2.7.1.

To analyze participants' eye-tracking data, for each of the ten search results on a SERP a polygonal area of interest (AOI) was defined around the search result (covering the title, the excerpt, and the URL), in order to determine for *how long* and *in which order* a participant was looking at a search result. The minimum fixation duration was set to 100 milliseconds with a fixation radius of 30 pixels. On this basis, the *total fixation time of each search result*, that is, the overall amount of time participants scrutinized a search result to decide whether to click on it or not, was calculated. Second, as a measure of distributed viewing behavior across the screen (cf. Hertzum et al., 2009) participants' viewing sequences on the SERPs, that is, the order in which the search results on a SERP were inspected, were computed, by comparing them with a perfectly linear sequence from top to bottom (search result 1, search result 2, search result 3, search result 4, etc.) by means of the Levenshtein distance (cf. Josephson & Holmes 2002). Levenshtein distance values were transformed into a similarity percentage: A *higher similarity percentage* indicated *more linear* top-to-bottom viewing sequences. The similarity percentages were averaged across the three SERPs. Third, participants' *mean*

pupil size (in mm) during task processing was computed (averaged across eyes).

Furthermore, based on the mouse click recordings, the number of Websites participants' accessed during their Web search was analyzed (cf. Hertzum et al., 2009).

Procedure

Participants were tested in individual sessions of approximately one hour. The lighting conditions were kept constant during all examinations. Before participants started on the search task, control variables and self-reported prior domain knowledge were assessed. Furthermore, participants received some general instructions about the Web search experiment as well as the TA instructions according to their experimental condition. Next, they performed a practice task (structured in the same way as the subsequent main task) for approximately five minutes to get acquainted with the Web search environment and with the TA method. After the practice task, participants received the instruction for the main task (i.e., the request of the fictitious friend). Then, they were calibrated on the eye-tracking system using a nine-point calibration and started their Web search. In the two TA conditions, whenever participants stopped verbalizing their thoughts, the experimenter reminded them (after 5 seconds) to keep thinking aloud or to mention the evaluation criteria, respectively. After 20 minutes participants were asked to stop the task. Subsequent to their Web search, they had to decide which of the two weight loss methods they would recommend. However, in the present study only process measures during Web search were analyzed.

Results

ANCOVAs with TA conditions (neutral, prompted, and silent) and prior domain knowledge (z-scored) as well as an interaction term between TA conditions and prior domain knowledge were conducted. Significant interaction effects between TA condition and prior domain knowledge were probed according to the procedure outlined by Aiken and West (1991). Table 1 shows means and standard errors for the dependent variables.

Verbal Utterances

With respect to verbal utterances about *topical relevance* neither TA condition nor prior domain knowledge had a significant effect and there was no significant interaction between the two factors (all $F_s < 1.90$, $p_s > .18$).

With respect to verbal utterances about *information quality* the ANCOVA showed a significant effect of TA condition ($F(1, 25) = 10.95$, $p = .003$, $\eta_p^2 = .30$) and of prior knowledge ($F(1, 25) = 6.22$, $p = .02$, $\eta_p^2 = .20$). These effects were qualified by a significant interaction between the two factors ($F(1, 25) = 4.62$, $p = .04$, $\eta_p^2 = .16$). Simple comparisons according to the procedure outlined by Aiken and West (1991) revealed that, consistent with the predictions of *H1*, only high-knowledge participants (i.e., 1

SD above the sample mean) and moderate-knowledge participants (i.e., at the sample mean) expressed more information-quality utterances during prompted TA than during neutral TA ($\beta = .80$, $t(25) = 3.85$, $p = .001$ and $\beta = .48$, $t(25) = 3.31$, $p = .003$, respectively). In contrast, low-knowledge participants (i.e., 1 *SD* below the sample mean) did not significantly differ in the two TA conditions ($\beta = .16$, $t(25) = 0.79$, $p = .44$). In addition, in the prompted TA condition prior knowledge was significantly positively related to the number of information-quality utterances ($\beta = .67$, $t(25) = 3.30$, $p = .001$), whereas in the neutral TA condition it wasn't ($\beta = .05$, $t(25) = 0.24$, $p = .81$).

With respect to verbal utterances regarding the *structure of Web pages* the ANCOVA showed similar effects: a significant effect of TA condition ($F(1, 25) = 13.69$, $p = .001$, $\eta_p^2 = .35$), a marginal effect of prior knowledge ($F(1, 25) = 3.49$, $p = .07$, $\eta_p^2 = .12$), and a significant interaction between the two factors ($F(1, 25) = 6.01$, $p = .02$, $\eta_p^2 = .19$). Simple comparisons revealed that only high-knowledge participants and moderate-knowledge participants expressed more structure-related utterances during prompted TA than during neutral TA ($\beta = .89$, $t(25) = 4.34$, $p < .001$ and $\beta = .53$, $t(25) = 3.70$, $p = .001$, respectively). In contrast, low-knowledge participants did not differ in the two TA conditions ($\beta = .17$, $t(25) = 0.85$, $p = .41$). In addition, as for the information-quality utterances, in the prompted TA condition prior knowledge was significantly positively related to the number of structure-related utterances ($\beta = .62$, $t(25) = 3.07$, $p = .01$), whereas in the neutral condition it wasn't ($\beta = -.08$, $t(25) = -0.41$, $p = .69$).

Eye-Tracking and Mouse Clicks

With regard to the *linearity of participants' viewing sequences* on the SERPs, as measured by the similarity percentages of participants' string to a linear top-to-bottom string, the ANCOVA showed a significant main effect of TA condition ($F(2, 38) = 6.19$, $p = .01$, $\eta_p^2 = .25$). Bonferroni post-hoc tests revealed that in line with the predictions of *H2*, in the prompted TA condition participants' viewing sequences were significantly less linear than in the silent condition ($p = .004$). The differences between the neutral TA condition and the silent condition were only marginally significant ($p = .08$). The two TA conditions did not differ significantly ($p = .89$). Prior knowledge had no effect on participants' viewing sequences and there was no interaction with TA conditions (both $F_s < 1$).

Regarding the *total fixation time of each search result* the ANCOVA showed a significant main effect of prior knowledge ($F(1, 38) = 4.00$, $p = .05$, $\eta_p^2 = .10$). In line with the predictions by *H4*, the higher participants' domain knowledge, the more intensely they scrutinized the search results to decide whether or not to click on them. TA conditions had no effect ($F < 1$) and there was no interaction with prior knowledge ($F(1, 38) = 1.52$, $p = .23$).

Table 1: Means (and standard errors) of the dependent variables as a function of thinking-aloud (TA) conditions.

Dependent variables	Thinking-aloud condition		
	Prompted instructions	Neutral instructions	Silent condition
# utterances about topical relevance	6.92 (1.21)	4.51 (1.26)	NA
# utterances about information quality	12.11 (1.55)	4.74 (1.60)	NA
# utterances about structure of Web pages	5.52 (0.84)	1.05 (0.87)	NA
% linearity of viewing sequences on SERPs	62.11 (3.90)	68.10 (4.10)	81.40 (4.00)
total fixation time (in s) of each search result	3.02 (0.35)	3.03 (0.36)	2.62 (0.36)
mean pupil size (in mm)	3.71 (0.10)	3.86 (0.11)	3.58 (0.10)
# Websites accessed (in %)	47.14 (4.09)	52.91 (4.30)	42.49 (4.20)

With regard to participants' *mean pupil size* during task processing, after controlling for participant's average eye-to-screen-distance (which had a strong positive effect on pupil size, $F(1, 37) = 7.89, p = .01, \eta_p^2 = .18$), the ANCOVA showed a significant effect of prior knowledge ($F(1, 37) = 5.87, p = .02, \eta_p^2 = .14$). In line with the predictions by *H5*, less prior domain knowledge was associated with larger pupil sizes. Besides, there was no significant effect of TA conditions ($F(2, 37) = 1.71, p = .20$), nor a significant interaction with prior knowledge ($F < 1$).

With regard to the *number of Websites participants accessed* during Web search, the ANCOVA showed no significant main effect of TA conditions ($F(1, 38) = 1.51, p = .24$), but of prior knowledge ($F(1, 38) = 5.40, p = .03, \eta_p^2 = .12$). This effect was qualified by a significant interaction with TA conditions ($F(2, 38) = 7.11, p = .002, \eta_p^2 = .27$). Simple comparisons revealed that participants with high prior knowledge accessed significantly more Websites in the prompted TA condition than in the silent or neutral TA condition ($\beta = .45, t(38) = 2.26, p = .03$ and $\beta = .53, t(38) = 2.32, p = .03$, respectively), which did not differ significantly ($\beta = .09, t(38) = 0.42, p = .68$). Thus, the assumptions of *H3* are at least partly confirmed. However, other than expected, low-knowledge participants accessed significantly more Websites in the neutral TA condition than in the silent or prompted TA condition ($\beta = .51, t(38) = 2.41, p = .02$ and $\beta = .72, t(38) = 3.65, p = .001$, respectively), which did not differ significantly ($\beta = .21, t(38) = 0.97, p = .34$).

In addition, in the prompted TA condition prior knowledge was significantly positively related to the number of selected search results ($\beta = .93, t(38) = 4.13, p < .001$), whereas in the silent and neutral TA condition no significant relationship with prior knowledge were found ($\beta = -.38, t(38) = -1.66, p = .11$ and $\beta = .25, t(38) = 1.17, p = .25$, respectively).

Discussion and Conclusion

With the present study we sought to examine the impact of TA instructions – especially of "prompted" TA instructions that entail evaluation prompts as used in some previous studies (e.g. Rieh, 2002; Tombros et al., 2005) – as well as of prior domain knowledge on participants' information processing and source evaluation during Web search.

First, with regard to verbal utterances about information quality and the structure of Web pages, in line with *H1*, our study showed that prompted TA instructions (i.e., to mention the evaluation criteria applied to select search results and to assess Web pages) as compared to neutral TA instructions (i.e., to verbalize everything that comes to their mind) significantly increased participants' verbal reflections on information quality and on structural aspects of Web pages, given that participants possessed at least a moderate level of prior knowledge on diets and nutrition. In contrast, irrespective of the instructions given, students with no or little prior domain knowledge only rarely expressed such utterances. Thus, to conclude, the findings by Rieh (2002) and Tombros et al. (2005) seem to have resulted from a combination of prompted TA instructions and a certain level of prior domain knowledge of the participants. Utterances about topical relevance, on the contrary were expressed by all participants to a similar extent, irrespective of prior knowledge or instructions. That is, the evaluation of whether or not a search result or a Web page addresses the search topic seems to be a default process that guides every Web search task.

Second, with regard to the effects of prompted or neutral TA instructions on participants' information processing during Web search as compared to a silent condition, in line with *H2* and the findings by Hertzum et al. (2009) about Web page inspection, we found that the prompted instructions resulted in less linear, that is, more distributed viewing sequences on the SERPs. That is, instead of simply following the list order (a typical effect on SERPs) participants examined the search results in a rather free way. This can be seen as an indication of own evaluations, instead of simply relying on the order presented by the search engine. Between the neutral TA instructions and the silent condition, in contrast, only marginal respective differences were found (cf. Hertzum et al., 2009). With regard to the number of Websites accessed during Web search the results showed a more complex pattern than expected. For participants with rather high prior knowledge similar effects were found as in the study by Hertzum et al. (2009), namely that participants in the prompted TA condition accessed more Websites than in the other conditions (in line with *H3*). However, for participants with a moderate level of prior knowledge no differences were

found between conditions, and participants with low prior knowledge accessed more Websites with neutral TA instructions than with prompted instructions or when working in silence. A post-hoc explanation for the latter finding might be that low-knowledge participants in the neutral TA condition were overwhelmed by the situation and did not really know what to say and to do and, thus, simply selected (almost) all Websites.

Third, with regard to further effects of prior knowledge, in line with *H4* we found that the higher students' self-reported prior knowledge on diets and nutrition the more intensively they scrutinized the search results on the SERPs to decide whether to select them (cf. MaKinster et al., 2002). Furthermore, in line with *H5* participants with lower prior knowledge across conditions had increased pupil sizes, which is an indicator of increased cognitive load. This effect provides further evidence for the assumption proposed by Bråten et al. (2011) that low-knowledge users have to employ more cognitive effort in order to comprehend the content of the Web pages than Web users with higher domain knowledge. As a consequence, they seem to lack cognitive resources to engage in evaluations that go beyond content. This is also indicated by the fact that in both TA conditions low-knowledge users hardly ever verbally reflected on the information quality or on the structure of Web pages. Future research should examine these effects in greater detail, analyzing more detailed moment-to-moment changes in participants' pupil sizes and using within-subjects task variations.

In sum, from a methodological perspective, the present findings indicate that prompted TA instructions increase participants' verbal reflections on information quality and on structural aspects of Web pages, and, therefore, might overestimate, users' spontaneous reflections about these issues. Yet, from an educational point of view the findings of the present study suggest that simple evaluation instructions that prompt Web users to evaluate search results and Web pages have the potential to increase their awareness about the necessity to critically evaluate the information found on the Web – given that they possess a certain amount of prior knowledge on the subject matter of the search task at hand. In order to improve source evaluations of searchers with no or little prior domain knowledge, however, more comprehensive instructional support seems to be required that also includes measures to reduce individuals' cognitive burden during task processing. Furthermore, when the WWW is used as a research tool in formal learning settings (e.g., in school), it seems advisable that the teacher first provides some general information on the topic before students start with their Web search.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Bråten, I., Strømsø, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21*, 180-192.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*, 316-344.
- Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*, 220-231.
- Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: young adults' evaluation of Web content. *International Journal of Communications, 4*, 468-494.
- Hertzum, M., Hansen, K. D., & Andersen, H. H. K. (2009). Scrutinising usability evaluation: Does thinking aloud affect behaviour and mental workload? *Behaviour and Information Technology, 28*, 165-181.
- Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology, 48A*, 598-612.
- Josephson, S. & Holmes, M. E. (2002). Visual attention to repeated internet images: testing the scanpath theory on the world wide web. In A.T. Duchowski (Ed), *Proceedings of the ETRA 2002 Symposium*. ACM Press.
- MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find Newton's Third Law?: Case studies of students using of the Web as a science resource. *Journal of Science Education and Technology, 11*, 155-172.
- Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*, 2078-2091.
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.
- Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology, 53*, 145-161.
- Savolainen, R., & Kari, J. (2006). User-defined relevance criteria in Web searching. *Journal of Documentation, 62*, 685-707.
- Tombros, A., Ruthven, I., & Jose, J. M. (2005). How users assess Web pages for information-seeking. *Journal of the American Society for Information Science and Technology, 56*, 327-344.
- Wang, J. T. (2010). Pupil dilation and eye-tracking. In M. Schulte-Mecklenbeck, A. Kuhberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research: a critical review and user's guide*. New York: Psychology Press.