Author(s): Hautala, Jarkko; Kiili, Carita; Kammerer, Yvonne; Loberg, Otto; Hokkanen, Sanna; Leppänen, Paavo H.T.

Title: Sixth graders' evaluation strategies when reading Internet search results : an eye-tracking study

Year: 2018

Version: Accepted version (Final draft)

Please cite the original version:

Hautala, J., Kiili, C., Kammerer, Y., Loberg, O., Hokkanen, S., & Leppänen, P. H. (2018). Sixth graders' evaluation strategies when reading Internet search results : an eye-tracking study. Behaviour and Information Technology, 37(8), 761-773. https://doi.org/10.1080/0144929x.2018.1477992

"This is an Accepted Manuscript of an article published by Taylor & Francis in Behavior and Information Technology on 24.5.2018, available online: http://www.tandfonline.com/10.1080/0144929X.2018.1477992."

Sixth Graders' Evaluation Strategies when Reading Internet Search Results: An Eye Tracking Study

Jarkko Hautala[1,2], Carita Kiili[3], Yvonne Kammerer[4], Otto Loberg[2], Sanna Hokkanen[2] and Paavo H.T. Leppänen[2]

[1]Center for Applied Language Studies, University of Jyväskylä, Jyväskylä, Finland

[2]Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

[3]Department of Education, University of Oslo, Oslo, Norway

[4]Leibniz-Institut für Wissensmedien, Tuebingen, Germany

Corresponding author: Jarkko Hautala, Center for Applied Language Studies, POB 35, Mattilanniemi 6, FI-40014, University of Jyväskylä, Jyväskylä, Finland. Email: jarkko.v.hautala@jyu.fi, Phone: +358 40 805 3487, Fax: +358-14-260 4400

Sixth Graders' Evaluation Strategies when Reading Internet Search Results: An Eye Tracking Study

(8634 words)

Abstract

Evaluating search engine results is a crucial skill for finding relevant information on the Internet. In this study, we used eye-tracking technology to examine search result evaluation strategies adopted by sixth-grade students (N = 36). Students completed 10 search tasks where they were asked to select a search result among four options that would help them to answer the given task. To identify which information students used to evaluate search results, we manipulated the relevancy of the search result's title, URL, and snippet components. We then analyzed the selection of search results as well as looking probabilities on the search result components. The results revealed that during first-pass inspection, students read the search engine page by first looking at the title of a search result. If the title was relevant, the probability of looking at the snippet of the search result increased. During second-pass inspection, there was a high probability of students focusing on the most promising search result by inspecting all of its components before making their selection. A cluster analysis revealed three viewing strategies: half of the students looked mainly at the titles and snippets; one-third with high probability examined all components; and one-sixth mainly focused on titles, leading to more frequent errors in search result selection. The results indicate that students generally made a flexible use of both eliminative and confirmatory evaluation strategies when reading Internet search results, while some seemed to not pay attention to snippet and URL components of the search results.

Keywords: *information search, online reading, search engine results page, eye tracking*

1. Introduction

The ability to search for relevant information on the Internet using search engines is essential for 21st century literacy. However, research indicates that students of various ages face difficulties in locating and critically evaluating information (Bilal & Kirby, 2001; Leu, Kulikowich, Sedransk, & Coiro, 2009; Leu, Coiro, Castek, Hartman, Henry, & Reinking, 2008). For a successful internet search, search engine users need to set an information need, define appropriate search query terms, evaluate and select search results from the search engine result page (SERP), and process the selected webpage(s) until their information need is fulfilled (Brand-Gruwel, Wopereis, & Walraven, 2009; Dinet, Chevalier, & Tricot, 2012; Sharit, Hernández, Czaja, & Pirolli, 2008). The ability to select relevant links from an SERP is a key skill that can significantly increase the effectiveness of retrieving the desired information (Argelagos, & Pifarre, 2012; Brand-Gruwel et al., 2009; Rouet & Britt, 2011; Rieh, 2002). The present study applied eye-tracking recordings to examine what kinds of evaluation strategies sixth graders' spontaneously applied when reading search results.

1.1. Evaluation of Internet search results

Each search result comprises a title, a snippet (i.e., an excerpt of the webpage content), and the webpage's uniform resource locator (URL) address. However, people do not always systematically use all these components to make selections in web environments but tend to rely on cognitive heuristics; that is, they consider only a few aspects, rather than systematically analyzing all aspects of the material (Dinet et al., 2012; Metzger, Flanagin, & Medder, 2010; Salmerón, Kammerer, & García-Carrión, 2013).

Adult readers generally use efficient "satisficing" strategies (cf. Simon, 1955; satisficing = a combination of satisfy and suffice) when conducting Internet searches (Pirolli, 2007). For instance, they do not evaluate all links and the information available for them, but are likely to stop at the search result they consider "good enough" (Lorigo et al., 2008). They tend to first skim through SERPs and look at the first few search results (Pan et al., 2007; Kammerer & Gerjets, 2014) before scrolling further down, proceeding to the next SERP, or refining the query (Lorigo et al., 2008). These findings suggest that people often evaluate the success of their search query before engaging in a detailed evaluation of the search results on the SERP. In addition, when asked to bookmark webpages for further study, they often also select results located further down the SERP (Salmerón et al., 2013).

Next, both the perceived relevancy of search results for the topic at hand and their ranking position in the SERP affect link selection (Lorigo et al., 2008). Several studies show that people inspect more search results when the rank order of the results is reversed (Pan et al., 2007; Kammerer & Gerjets, 2014). Howeover, in the reversed condition, users also more often click on irrelevant links that are listed first on the SERP (Pan et al., 2007) Taken together, people seem to click on links they find most relevant while placing considerable trust in the search engine (Lorigo et al., 2008; Matsuda, Uwano, Ohira, & Matsumoto, 2009).

Within a search result, viewers spend most of their time reading title lines and pay less attention to text snippets and URLs (Dinet, Bastien, & Kitajima, 2010; Granka et al., 2008), particularly when letter-normalized viewing times are being analyzed. In addition to evaluating the semantic relevance of a search result, the expected quality of information (or credibility of an information source) can also play a role in a user’s selection decisions (e.g., Balatsoukas & Ruthven, 2012; Kammerer & Gerjets, 2014; Rieh, 2002). Because anyone can virtually publish any information on the Web, the quality of information varies widely and many websites provide incomplete and/or inaccurate information. In a search result, for example, the URL provides cues about the credibility of the information source (e.g., Kammerer, Bråten, Gerjets, & Strømsø, 2013). Accordingly, tasks that afford finding a specific webpage or that require finding credible information lead to more URL and snippet viewing (González-Caro & Marcos, 2011; Matsuda et al., 2009). A study that integrated user selections, eye movements, and think-aloud protocols, found that individuals used appropriate relevancy criteria, for example, topic relevance and scope for titles as well as information quality and domain expertise for URLs, when exploring search result components (Balatsoukas & Ruthven, 2012).

1.2. Internet search evaluation by adolescents

Generally, presumably due to the nonlinear nature of online reading (Sung, Wu, Chen, & Chang, 2015) acquiring effective Internet search skills takes years to develop and is greatly facilitated by proper instruction (Bannert & Reimann, 2012; Van Deursen et al., 2014). Already sixth-grade students could evaluate the relevancy of search results with respect to a given search problem, but this skill was not fully developed until the eighth grade (Keil & Kominsky, 2013). In another study, SERP reading became more efficient from sixth to eighth grade, with faster response times and fewer clicks on search results (Gwidzka & Bilal, 2017) . Moreover, several studies have suggested that adolescents do not typically assess the reliability or credibility of information during web searches (Jochmann-Mannak, Huibers, Lentz, & Sanders, 2010; Kiili, Laurinen, & Marttunen, 2008; Walraven, Brand-Gruwel, & Boshuizen, 2009). For example, seventh grade students continued to make search result selections on the basis of superficial cues, such as boldfaced keywords, instead of semantic information (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011).

Eye movement studies have shown that adolescent begin by reading almost all results listed on an SERP, after which they pay more individual attention to them (Bilal & Gwidzka, 2016). In addition, eighth graders have been shown to start reading SERPs more consistently from the first ranked search result to the bottom, while sixth graders made more premature clicks on search results before reading them (Bilal & Gwidzka, 2016). Further, younger children looked at fewer snippets and instead looked more at thumbnail images, suggesting that children find it difficult to read long texts in the SERPs (Gossen, Höbel, & Nürnberger, 2014). Eye movement analysis of SERP reading among fifth, seventh, ninth, and eleventh grade students performing simple fact-finding tasks indicated that the typographical cueing of boldfaced search words seemed to attract the younger readers' gaze (Dinet et al., 2010; cf. also Rouet et al., 2011). Older students, on the other hand, were attracted by such cues only when the information search task was based on unfamiliar topics.

There is also considerable age and grade related variation in viewing strategies. For instance, while fifth and seventh grade students gazed mostly the boldfaced keywords, ninth and eleventh grade students individually read each search result (Dinet et al., 2010). Challenging search tasks also induced extensive reading of the search results in adolescents, until the task became too difficult causing effort decline (Walhout & Ooomen, Jarodzka, & Brand-Guwel, 2017).

In sum, these results indicate slow and gradual development of search result evaluation skills. However, very few studies (Dinet et al., 20120) have attempted to study to what extend children or adolescents use different types of information (title, URL, snippet) embedded in the search results, which is the main objective of the present study.

1.3. Interactive search framework

Cognitively, SERP reading can be considered an interactive search of a target item on a list, that is, a relevant search result among those less relevant. According to Brumby and Howes' (2008) interactive search framework, whether individuals pursue an exhaustive evaluation or satisficing strategy depends on the similarity or distinctiveness of a set of items encountered (e.g., with respect to relevance or credibility). Items that are sufficiently distinct are selected without others being inspected; however, if none of the items stands out, readers may browse all the items and/or re-inspect a subset of items considered the most relevant.

Applying this interactive search model to SERP reading suggests that encountering a highly relevant search result will inhibit the processing of subsequent search results. In addition, the interactive search process may affect the manner in which the components of a search result are inspected. Readers typically begin reading a search result from the title. If the title is relevant, they may proceed to reading the search result’s snippet and/or URL components. If not, they may eliminate this search result without inspecting its snippet and URL components. At some point, especially when the search results on a SERP become exceedingly irrelevant to the task at hand, the readers may enter a re-inspection phase to re-evaluate the search results considered the most relevant. During this re-inspection, they may re-read only the titles or deepen their evaluation using information provided in the search result’s snippet and/or URL components.

1.4. Research questions and hypotheses

We used eye movement recordings to examine sixth graders' spontaneous evaluation strategies during reading search results. The eye-tracking method is well suited for this, as it allows tracing the target of visual attention during task performance by following gaze location on the screen (see Rayner, 2012) .

We posed the following research questions (RQ) and hypotheses (H).

RQ 1: Are sixth-grade students able to utilize information provided by each search result component (i.e., title, URL, and snippet) as reflected in the selection rates of search results with (a) all components being relevant, (b) a result with an irrelevant snippet, (c) a result with an unreliable URL, or (d) a result with an irrelevant title?

H1: Given the finding that students do not systematically evaluate the credibility of information, it was expected that sixth-grade students can eliminate search results on the basis of irrelevant title or snippet information, but not on the basis of unreliable URL information.

RQ 2: What information sources do the students pay attention to and which evaluation strategies do they use during their selection? We operationalized this as the first-pass (i.e., initial inspection) and second-pass (i.e., re-inspection) looking probabilities of various components in search results (as defined in RQ1).

H2: In line with the interactive search model (Brumby & Howes, 2008), we hypothesized that the relevancy of a search result's title determines whether its snippet and URL address will be inspected. We separately examined the presence of these effects for initial inspection and re-inspection of search results. This is because readers may first eliminate poor-matching search results on the basis of title information only, whereas during the re-inspection, they may be more concerned with the snippet and URL components of relevant titles in the search results.

RQ 3: Does the early positioning of correct search results on the search list decrease the need to inspect other search results?

H3: According to the interactive search model, encountering a highly matching search result would reduce the need to inspect subsequent ones.

RQ 4: Are there differences between students in how they read and evaluate Internet search results?

H4: Previous studies have found that people use different heuristic in solving information problem solving tasks (e.g. Graff, 2005; Lawless & Kulikowich, 1996). Here, it is expected that students differ in the extent of attention they pay to the title, URL, and snippet components of the search results. We explored this using a cluster analysis that included the number of times students looked at the title, snippet, and URL of the search results.

2. Materials and methods

2.1. Participants

The participants were 36 students (age: *M* = 12.5 years, *SD* = 3.6 months, 18 males) on their last, i.e. sixth, primary school year, recruited from five schools in Central Finland. Students of this age are in the transition phase to adolescence. These students were also participants of our larger research project concerning Internet reading skills among students with and without learning disabilities. The present study focuses on search result evaluations by students without learning disabilities and thus, the following commonly used exclusion criteria were applied:

1) Reading difficulties, which were defined as a reading fluency performance score below the 15th percentile (based on the factor score derived from three reading measures: Lindeman, 1998; Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015; Holopainen, Kairaluoma, Nevala, Ahonen, & Aro, 2004) or as a parental report of the student with a reading disability diagnosis.

2) An attention-deficit scale score below the 25th percentile in a questionnaire using teacher ratings (Kesky; Klenberg, Jämsä, Häyrinen, & Korkman, 2010).

3) A nonverbal IQ performance result below the 7th percentile based on a 15-minute, 30-item version of the Raven matrices (Raven, Court, and Raven, 1992).

Written consent was obtained from all participants and their caregivers prior to the study. Ethical approval was derived from the Ethical Board of University of Jyväskylä.

2.2. Apparatus

Eye movements were recorded using a table-mounted EyeLink 1000 eye-tracker (SR Research) with forehead- and chin-rest. The stimuli were presented on a Dell Precision T5500 workstation with an Asus VG-236 monitor (1920 x 1080, 120 Hz, 52 x 29 cm) at 60 cm viewing distance. 13-point calibration with a one-degree visual angle as the acceptance criterion was applied. We conducted the calibration prior to the experiment and repeated it between trials when visible (a) head movements were made, (b) a drift was seen on the researcher's screen where the subjects' eye movements were overlaid on experimental stimuli, or (c) the calibration error exceeded .30 visual degrees.

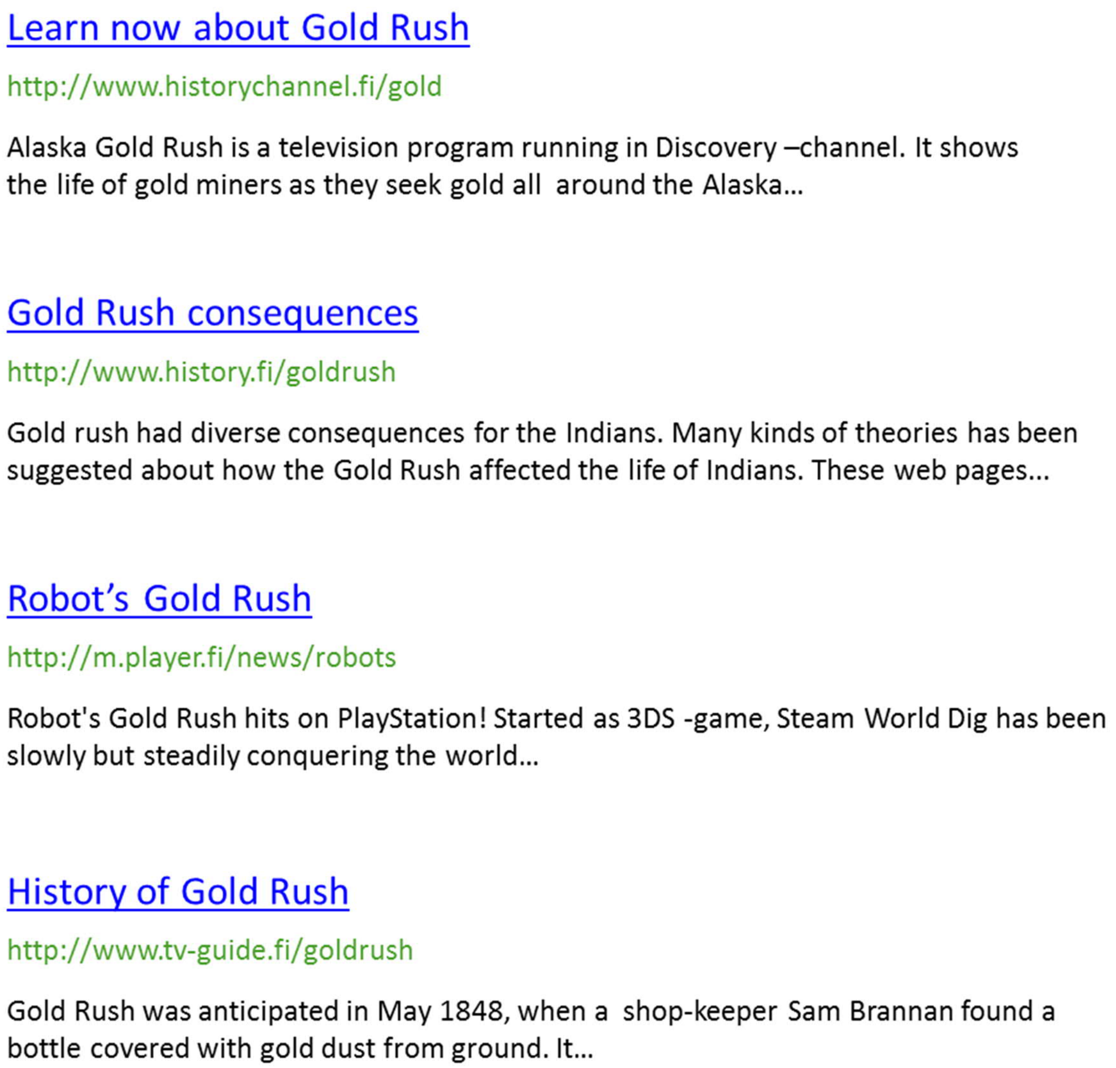
2.3. Tasks and Materials

The students completed a practice task and ten simulated information search tasks. To begin with, the students were shown a contextualized question (altogether four lines) on the screen. For example, the students were asked to find an answer to the question "Why was the Gold Rush harmful to Indians?". Then, they were shown four search results (see Figure 1) and asked to select one that would help them to answer the question.

The information search problem tasks focused on the following themes: coral reefs, gold nuggets, gold rush, placebo, doping, panda population, panda endangerment, vaccination rate, vaccination side-effects, and reasons for humpback whale migration and approximate distances. We excluded panda population task because it had a false constellation of search result types (two Irrelevant-Snippet items) owing to human error in stimuli preparation.

Each SERP (Fig 1) contained:. (A) a result with all the components being highly relevant to informational need (Correct), (B) a competing result with an irrelevant snippet (Irrelevant-Snippet) (C) a competing result with an untrustworthy URL address (Irrelevant - URL), and (D) a result with all the components being irrelevant (Distractor). The rank order of the different search result categories was counterbalanced across the tasks.

Figure 1. Translated example of a search result screen, preceded by a task assignment: "Find out, why was the Gold Rush harmful for Indians" On this screen, the rank order of the search result types was Irrelevant-Snippet, Correct, Distractor, and Irrelevant-URL. Note that the snippet texts extended over two complete lines of text in the Finnish language.



Two researchers rated the relevancy of each item with respect to the task assignment using a four-point scale (very relevant, probably relevant, probably irrelevant, and obviously irrelevant) with a Cronbach's alpha of .95 (Table 1). The length of the search result components (in characters) for all search result categories was equal within each task (Table

1). The titles were presented in Calibri 16-point font and the URL address and text field were in Calibri 12-point font with a line spacing of 6 points, equaling the minimum accuracy limits of the eye tracker's spatial accuracy of 0.5°.

Table 1. Descriptions of stimuli including number of written characters and relevancy ratings with means and standard deviations.

		Characters		Relevancy	
		Mean	SD	Mean	SD
Correct	Title	23.4	1.34	3.6	.70
	URL	31.4	2.31	3.2	1.3
	Snippet	156.0	8.0	3.6	.85
Irrelevant-Snippet	Title	22.8	3.34	3.1	.89
	URL	31.5	2.07	2.9	.99
	Snippet	160.2	9.4	1.6*	.60
Irrelevant-URL	Title	25.0	2.87	3.2	.44
	URL	32.2	2.28	1.9*	1.0
	Snippet	156.8	8.4	3.2	.79
Distractor	Title	24.1	2.96	1.4	1.1
	URL	32.4	2.72	1.6	1.2
	Snippet	158.5	8.8	1.4	.99

*$p < .05$ is a difference in the relevancy rating (on the scale 1 - 4) compared to the other components within the search result type, indicating that Irrelevant-Snippet items had an irrelevant snippet component, while Irrelevant-URL items had an irrelevant URL address.

2.3.1. Prior knowledge

The information-seeking tasks were unlikely to be solved with a sixth grader's prior knowledge. The degree of knowledge, however, was self-evaluated prior to the experiment using the following types of question: e.g., how much do you know about the threats to coral reefs? The response options were (1) I know nothing (47% of responses), (2) I know little (26%), (3) I know some (21%), and (4) I know a lot (7%). Students' self-reported prior knowledge was not correlated with their accuracy in selecting the correct search result, $r(36) = .234$, $p = .170$.

2.3.1. Internet search experience

Here, we asked students about their media usage, including the following questions about web searching (a) in general for acquiring information; (b) at school; (c) after school; and (d) during spare time. The response options were as follows: (1) hardly ever, (2) rarely, i.e. 1–2 times per month, (3) 1–2 times per week, (4) almost every day, (5) for less than two hours every day, and (6) more than two hours every day. We also asked questions on the extent of instructions they received on conducting Internet searches: (e) From their teacher; and (f) From their caregivers or other adults The response choices were as follows: (1) not even once, (2) at least once, and (3) more than once. The students' answers were summed to form a single measure of Internet search experience. Students' Internet search experience did not correlate with accuracy in selecting the correct search result ($r = -.153$, $p = .374$).

2.4. Procedure

One research assistant accompanied the participant in a laboratory room while the other assistant controlled the devices in the control room. The sequence of activities for a participant was: prior knowledge questionnaire, task instructions on paper, adjustments of the eye tracker's table height and the forehead- and chin-rest, calibration, practice task , and finally the ten experimental information search tasks, including at least one short or several breaks of a few minutes, depending on individual needs. Calibration was repeated after the breaks. The students completed the tasks using a mouse. The duration of each experiment session varied from 45 to 90 minutes on the basis of the participant.

2.5. Eye-movement data processing

Data was preprocessed using the Data Viewer program (SR Research Ltd., Canada). Saccade velocity threshold of 30 degrees/sec, and minimum fixation duration of 80 ms were applied. For each SERP, 12 predefined pixel-precise areas of interest (AOIs) corresponding to the three components (i.e., title, URL, and snippet) of the four search results were determined. Misaligned fixation locations on the vertical axis were subject of manual correction, with inter-rater agreement of 89.2% on whether to correct a trial or not. The correction was needed for 36% of the trials due to (1) spatially close AOIs, which frequently led to cases in which the fixation location fell on the wrong side of the AOI border, and/or (2) calibration errors, including spatially selective inaccuracies or drifts, for example, at the bottom of the screen.

First-pass runs with a single fixation on a search result were excluded (344 out of 6,231 passes; 5.5%). These passes reflect accidental visits of a search result, for example, when students shifted their attention to the first search result on the SERP, which could introduce a considerably large viewing probability error. To ensure that these passes did not contain a cognitive signal, we inspected the summed fixation durations of these passes, which were equal across search result types.

2.6. Data analyses

2.6.1. General viewing strategies

For a detailed analysis of the students' viewing strategies, we analyzed the first- and second-pass looking probabilities[1]. A first-pass look was defined as a first inspection of a search result, and a second-pass look included all the later inspections of a search result (i.e., after having inspected or re-inspected one or several other search results in between). Because the employed statistical method required integer values (counts), we conducted the analyses on the basis of the number of tasks (0–9) in which the participant looked at a component of each search result type (see section Tasks and Materials). To derive more illustrative looking probability values (0–1) for the figures, we divided these task counts by the total number of tasks (9). The first-pass looking probability indexed the likelihood of a search result component being looked at during the initial inspection of a search result. The second-pass looking probability indexed the likelihood of a component being viewed when re-inspecting a search result, including second and subsequent passes.

We conducted a generalized estimating equation analysis (GEE; for more information, see Hardin, 2005; Homish, Edwards, Eiden, & Leonard, 2010) with SPSS to examine the data. We selected the GEE approach instead of an analysis of variance (ANOVA) approach, because the looking probability variables were not normally distributed. Within the GEE analysis, we applied the robust estimator of covariance matrix, exchangeable correlation structure, Poisson loglinear model for counts, and hybrid method for parameter estimation. The analysis comprised a four-level within-subject factor of the search result type (i.e., Distractor, Irrelevant-URL, Irrelevant-Snippet, and Correct) and a three-level within-subject factor of the search result component (i.e., title, URL, and snippet). Finally, we performed

[1] Analysis of first- and second-pass summed fixation durations produced the identical pattern of results to looking probability analysis.

paired post-hoc comparisons with the least significant difference correction for significance levels.

2.6.2 Effects of correct search result position on competitor viewing

It is possible that when a search result that was well-matched to the information search task had already been read, the subsequent search results were inspected in less detail. To determine whether the students employed this strategy, we compared the probabilities of looking at the competitors presented before and after the Correct search result (Fig 3). To match the number of competitors (eight) presented before and after the Correct search result, an additional task needed to be excluded from the analysis; we excluded the first task (gold nugget) after the practice trial. The results revealed no difference in the perceived values of relevancy for the search result components between competitors, presented before and after the Correct search results ($F < 1$).

2.6.3. Individual viewing strategies

The results of the aforementioned analyses revealed that students differed most in their inspection of competing search results. Therefore, we conducted the cluster analysis for the mean number of tasks that each component was looked at within the Irrelevant-Snippet and Irrelevant-URL items. The analysis was conducted for the standardized values using the Ward method (cf.. Hyönä, Lorch, & Kaakinen, 2002).

4. Results

4.1. Search result selection

The students chose the Correct search result with high accuracy (M = 81.0%, SD = 17.0%). The probability of selecting the Irrelevant-URL was M = 6.8% (SD = 10.0%) and that of selecting the Irrelevant-Snippet was M = 12.0% (SD = 11.0%). The Wilcoxon's signed-ranks test for two related samples indicated that the participants chose Irrelevant-

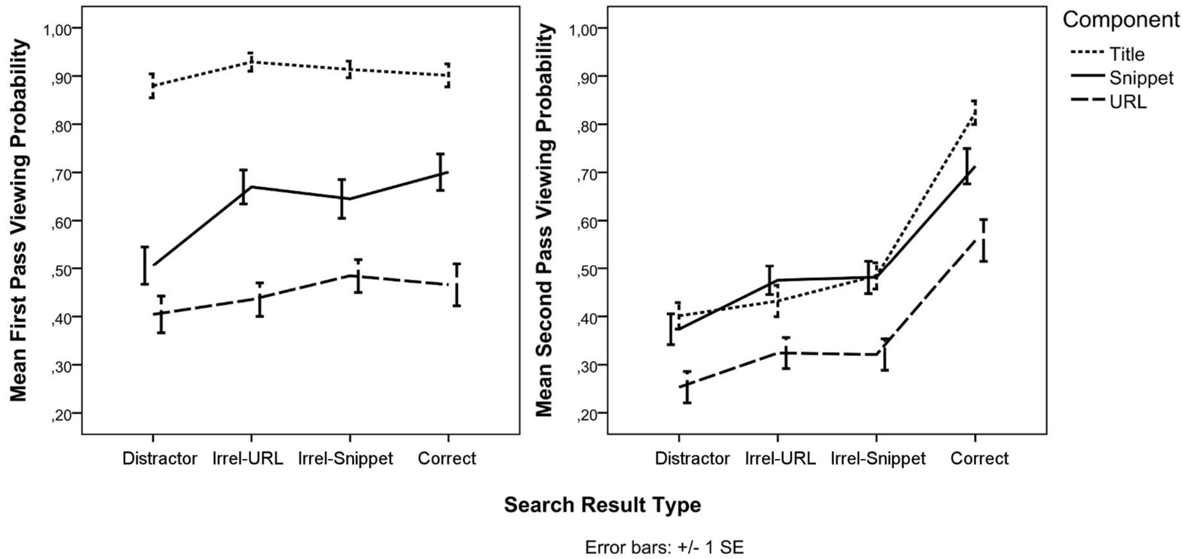
Snippet results more often than the Irrelevant-URL results ($Z(1, 35) = -2.28$, $p = .022$). No participant selected the Distractor search result in any of the nine tasks. The mean response time for the search result selection across the nine tasks was 23.4 s (*SD* = 7.7 s). There were no statistically significant gender differences in the accuracy of selecting the correct search result (p = .229) or in the response time (p = .566).

4.2. General viewing strategies

On average, the snippet was looked at the longest (*M* = 3.32 s, *SD* = 1.76), followed by the title (*M* = 1.55 s, , *SD* = .49) and URL (*M* = .50 s, , *SD* =.32). Notably, 55% of the URLs were never looked at, while this was true for only 11% of the snippets and 6% of the titles. The Correct search results were looked at the longest (*M* = 7.58 s, *SD* = 3.13, with by average 2.8 viewing occasions) and both the Irrelevant-URL and Irrelevant-Snippet were looked at for a nearly equal length of time (*M* = 5.25 s, *SD* = 2.06, and *M* = 5.16 s, *SD* = 2.24, with 2.0 viewing occasions), while the Distractor was looked at for the least duration (*M* = 3.44 s, *SD* = 1.55, with 1.8 viewing occasions).

Figure 2 presents the probabilities of looking at each search result component (title, URL, and snippet) within different search result types.

Figure 2. Probabilities of looking at different search result types and their components during first- (left panel) and second- (middle panel) pass viewing. Irrel is an abbreviation for Irrelevant.



4.3. First-pass looking probability

The GEE for the search results revealed significant main effects for both search result type ($\chi^2(3) = 37.6$, $p < .001$) and components ($\chi^2(2) = 86.9$, $p < .001$). These main effects were accompanied by a Type x Component interaction (χ^2 (6) = 27.4, $p < .001$). Pairwise comparisons revealed that the snippet for the Distractor search result was less likely to be looked at than those of the other types of search results (*p*s $\leq .007$), whereas the URL and titles of all types of search results were equally likely to be looked at.

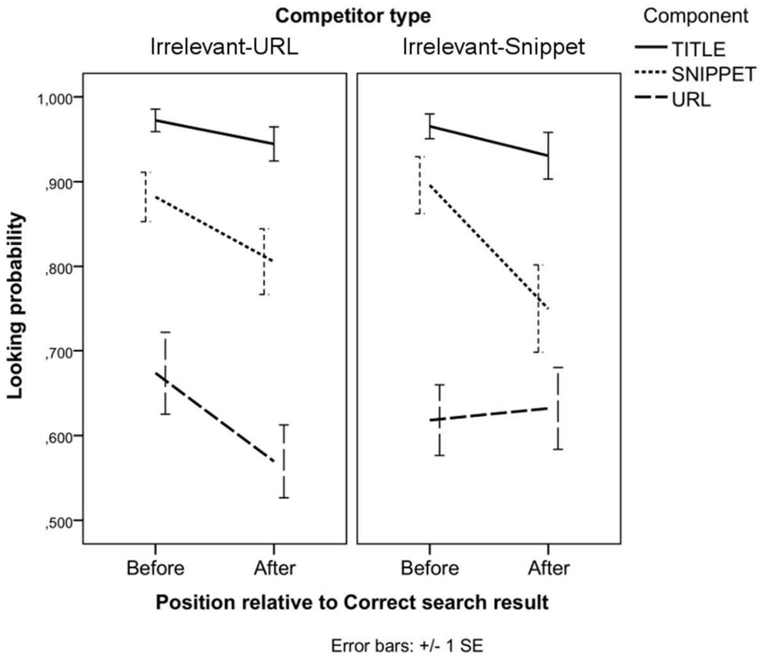
4.4. Second-pass looking probability

The main effects of both the search result type (χ^2 (3, $N = 36$) = 281.7, $p < .001$) and components (χ^2 (2, $N = 36$) = 40.5, $p < .001$) were significant. These main effects were accompanied by a Type x Component interaction (χ^2 (6, $N = 36$) = 14.3, $p = .026$). In general, all components of the Correct search result were looked at with much higher probability than those of the other search result types (*p*s < .001), which were looked at with equal probability. An exception was the snippet component of the Distractor search result, which was less likely to be looked at than the competitors' snippet components (*p*s ≤ .006).

4.5. Effects of Correct search result position on competitor viewing

The significant main effect of the Correct search result position (χ^2(1, *N* = 36) = 10.6, *p* = .001) was qualified by a significant three-way interaction of the Correct position (before, after), Competitor type (Irrelevant-Snippet and Irrelevant-URL), and Component (Title, URL, and Snippet) (χ^2(2, *N* = 36) = 6.20, *p* = .045). Titles were looked at with equal probability, regardless of whether the competitor preceded or followed the Correct search result (Fig 3). Irrelevant-Snippet's snippet component was less likely to be looked at if it appeared after the Correct search result, (*p* = .002). Irrelevant-URL's URL (*p* = .054) and snippet (*p* = 0.13) components were less likely to be viewed if they appeared after the Correct search result. These results suggest that after the students read the search results that strongly matched the information search task, they inspected the subsequent search results in less detail.

Figure 3. Probability of looking at competitors when Correct search result was positioned before (left panel) or after (right panel) a competitor.



4.6. Differences in viewing strategies

We identified the following three viewing strategy groups by conducting a cluster analysis (also see Fig 4): (1) six students (16.67%) who generally only looked at titles (Title readers); (2) 18 students (50%) who almost always looked at titles and snippets, but only occasionally looked at URLs (Title and Snippet readers); and (3) 12 students (33.33%) who almost always looked at all components (All Component readers). We validated the explanatory power of this cluster solution using a discriminant analysis and obtained an eigenvalue of 9.16, thus explaining 93.5% of the variance. A GEE analysis showed a significant two-way interaction between Component (Title, URL, and Snippet) and Viewing strategy Group ($\chi^2(4, N = 36) = 495$, $p < .001$). A post-hoc comparison indicated that the groups looked at titles with equal probability (ps > .99), but significantly differed from each other in their probability of looking at the snippet (ps < .001) and URL (ps < .040) components. An exception was that All Component readers and Title and Snippet readers looked at snippet components with an equal probability (*p* > .99).

Figure 4. Three-cluster solution based on probability of looking at competitors' URL and snippet components. The left panel: The probabilities of looking at the competitor's title, snippet, and URL components for the different viewing strategy groups. The right panel: the scatterplot of the probabilities of looking at the competitor's snippet and URL components (the numbers in the figure refer to number of participants with identical looking probability values).

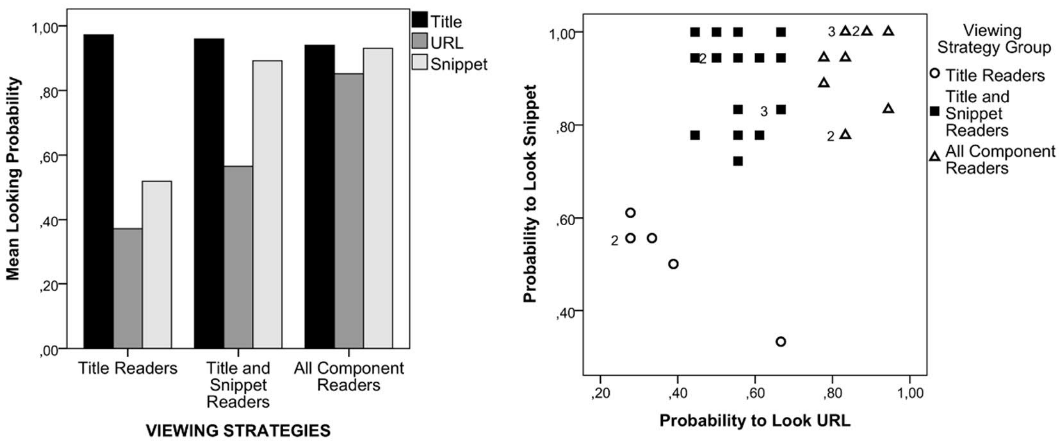


Table 2 lists the performance of the different viewing strategy groups. The groups differed in their accuracy of selecting the Correct search result ($\chi^2(2, N = 36) = 6.79$, $p = .033$), with Title readers making more errors than the two other groups (*p*s ≤ .007), while All Component readers as well as Title and Snippet readers performed equally well (*p* >.99). Title readers also responded more quickly than the two other groups (*p*s ≤ .009). There was no difference between cluster groups in prior knowledge (*p* = .534) and Internet search experience (*p* ≥ .85), screening variables of reading fluency (*p* = .394) and nonverbal intelligence (Raven; *p* = .179), or the attention deficit scale (*p* = .572).

Table 2. Performance of viewing strategy groups with means and standard deviations in parentheses.

	Viewing Strategy Groups					
	Title readers (n = 6)		Title and Snippet readers (n = 18)		All Component readers (n = 12)	
Accuracy Correct (%)	57	(24)	86	(12)	86	(7)
Irrelevant-Snippet error (%)	26	(13)	8	(9)	11	(8)
Irrelevant-URL error (%)	17	(14)	6	(10)	3	(5)
Response time (s)	16.9	(2.9)	23.8	(8.9)	26.1	(5.7)
Internet search experience (max 30 points)	13.8	(3.3)	13.1	(2.4)	13.4	(3.6)
Prior knowledge (max 36 points)	12.2	(3.3)	13.4	(2.9)	13.8	(2.5)
Reading fluency (factor score)	.46[1]	(.92)	.38	(.82)	.04	(.49)

Notes. [1] Here, n = 5 because one participant did not have data on the pseudo-word text reading subtask. Irrelevant-Snippet or Irrelevant-URL error is the average percentage of tasks students chose this type of search result in the experiment. Response time is the mean across all nine tasks. Reading fluency values are means of standardized factor scores with higher values indicating better reading fluency. In the Attention deficit scale larger value indicated poorer attention skills.

5. Discussion

This study aimed to explore strategies that typically developing sixth graders (12 to 13 year olds in the last year of their primary school) use to inspect and select search engine results, as a function of information value of the search result components (i.e., title, URL, and snippet) both during inspection (first-pass viewing) and re-inspection (second-pass viewing).

Students managed to choose the best-matching search result in 81% of the trials, paralleling previous findings that sixth-grade students are able to evaluate the relevancy of single search results to a given search problem (Keil & Kominsky, 2013). However, it is important to note that these results only tell us about students' ability to evaluate search results when they can focus on a limited amount of search results. It is thus possible that students might not perform as well in more complex information environments, such as the open Web. For example, Van Deursen et al., (2014) found that 9–13 years aged learners were able to find simple pieces of information from the open Web with 56 % accuracy.

Second, our students selected less often the irrelevant URL (6.8%) than irrelevant snippet (12%) competitor, contradicting previous findings that adolescents typically neglect credibility information such as URLs (Hirsch, 1999; Kroustallaki, Kokkinaki, Sideridis & Simos, 2015). A posthoc explanation for this finding might be that also properly reading and comprehending snippet information is a challenge in sixth graders' search result evaluation. This seemsreasonable considering that snippet text requires more careful reading than information in other components, which can be typically processed with a considerably lower number of fixations (Gossen, Höbel & Nürnberger, 2014). As a consequence, students may not adjust their reading style accordingly for snippets (Granka et al., 2008). The total fixation time provides supports for this interpretation, as titles were read at an average rate of 65 ms per letter, whereas snippets were read at an average rate of 20 ms per letter, which might explain why the students chose the competitor with irrelevant snippet relatively often in our study.

Third, students' search result evaluation strategies were analyzed based on the looking probabilities on the search results . During the initial (i.e., first-pass) inspection, the snippet of the Distractor search result was looked at with lower probability (50%) than the snippet of the three other types of search results (65–70%), indicating that students immediately eliminated irrelevant search results on the basis of their title information. . However, the lack of differences between the three relevantly-titled search results indicate that the students were not focusing on the Correst search results, suggesting that they were not yet about to make a selection during their first encounter with the search results. Moreover, the finding that URL addresses were looked at equally with a 40% probability across all search result types, suggests that URL information is not systematically used for elimination purpose at this stage.

However, students still seemed to detect the correct search result already during the first inspection, as evidenced by the much higher probability to return to look at the correct than the competing search results during second-pass inspection, and this was true for all components of the search result. This data pattern suggests that the students entered a phase of confirming their initial detection of the most relevant search result, after which they were likely to make a selection. In this phase, they seemed to exploit all possible information, by looking at both title and snippet and even URL components (Balatsoukas & Ruthven, 2012; González-Caro & Marcos, 2011; Matsuda et al., 2009). A noteworthy finding is that the students re-inspected only half of the competing search results, and equally on the two types of them. This finding further evidences that during re-inspection, students are more engaged in confirming their initial preferred search result, irrespective of it being right or wrong, rather than eliminating the competing search results. In addition, since the exhaustive elimination of competing search results requires greater cognitive effort, people are generally biased to confirm their opinions, even when faced with counterfactual evidence (Nickerson, 1998; White, 2013; Ashraf-Amri & Al-Sader, 2016).

Fourth, in line with the prediction based on interactive search theory (Brumby & Howes, 2008), when the Correct search result appeared earlier in the list, the students were less likely to look at the snippet or URL of the competitor search results. In other words, once students spotted a well-matching search result, they were not interested in comprehensively examining the remainder of the search results. In contrast, when a competitor with an irrelevant URL address appeared before the Correct search result, the competitor's snippet and URL components were more likely to be looked at. This implies that when a search result is acceptable based on the title and snippet information, its URL address is more likely to be looked at, given that the more appropriate Correct search result has not yet been read. This demonstrates that some students in this study used highly sophisticated evaluation strategies for the search results, which also highlights the need for further research on the fine dynamics involved in SERP reading (cf. Dinet et al., 2010; Metzger et al., 2010).

Finally, the present study adds to the understanding of inter-individual differences in the evaluation of search results (cf. Graff, 2005; Lawless & Kulikowich, 1996). As shown by the cluster analysis, one group of students (i.e., the Title readers, comprising one-sixth of the students) did not use the evaluation strategies as effectively as the other two groups. This particular group of students predominantly looked only at titles and did not view other search result components (i.e., snippet and URL) to make confirmatory or eliminative decisions. Consequently, they performed less successfully than the other two groups (57% vs. 86 % and 86%) who had a high probability of looking at both titles and snippets or all three components, respectively. As these groups of students did not differ in the assessed cognitive skills, or their self-reported prior knowledge or experience in conducting Internet searches, it seems likely that these strategies result from students' individual learning history. Therefore, it seems reasonable that students who do not utilize the snippet and URL -components in their search result evaluation, would probably benefit from a targeted instruction on this skill (cf. Coiro, 2011; Sung et al., 2015).

5.1. Theoretical implications

Drawing on various theoretical accounts of information searches (Brumby & Howes, 2008; Dinet et al., 2012; Metzger et al., 2010), this study offers new evidence about semantic control on looking behavior during information searches and SERP reading. Such behavior is in line with the findings that reading comprehension processes sensitively affect which portions of text are reread in particular during normal reading (Rayner, 2012). Our findings add to this knowledge that when reading hierarchically organized materials, such as SERPs, readers also routinely make decisions not to read certain parts of text, which are most likely to be irrelevant for the task at hand.

In general, people may try to minimize their cognitive effort by predominantly relying on workable heuristics and strategies to solve problems (Metzger et al., 2010). A common aspect of current information searching models is that theemployment of an iterative process to analyze information until the user's information need is fulfilled or the process is aborted (Dinet et al., 2012). Our results specify the cognitive strategies involved in evaluating Internet search results, by providing evidence for a hierarchical, two-stage model of search result evaluation. During the first stage, that is, the initial inspection of search results, students attempt to reduce the problem space by eliminating poor search results on the basis of title information and spotting the most relevant search results using both title and snippet information. During the second stage or the re-inspection phase, students are concerned with confirming the relevancy of the most promising search results spotted during the initial inspection. In addition, when a highly promising search result is spotted, the analysis of the upcoming search results is somewhat inhibited.

This search behavior can be understood by the principles of the interactive search model (Brumby & Howes, 2008). If information provided in a title exceeds a dynamic threshold for relevancy, its snippet will be also inspected. If the snippet also provides relevant information, the search result may be stored in working memory as a search result for potential selection. The spotting of such a promising search result increases the relevancy threshold, rendering it more likely for the information provided by upcoming search results to be categorized as less relevant.

5.2. Limitations

The present study adopted a highly controlled experimental approach in the laboratory setting to examine the reading and evaluation of Internet search results. Consequently, some caution is warranted in applying the findings into practice (Wopereis & van Merriënboer, 2011). It is likely that when searching information on the open Web the strategies reported here will be applied to only a subset of search results included in SERPs. For example, users may evaluate only a subset of the highest ranked search results and may discontinue the evaluation process when the relevancy of the titles decreases (Bilal & Gwidzka, 2016; Pan et al., 2007).

Another limitation concerns the generalizability of our findings. The present study was conducted with typically developing 12-year-old Finnish students, who are relatively experienced information searchers compared to students from less advantaged backgrounds. Still, it is expected that adults and older students might exploit even more sophisticated evaluation strategies, for instance, a more systematic use of a URL address during initial inspection. Thus, future studies with older students or adults are desirable.

Finally, some of the findings might be specific to the task requirements and materials. In the present study, the Correct search results had slightly higher relevancy ratings for each component than the competing search results. Students apparently identified the Correct search result already during first-pass reading, as they returned to this item with a higher probability relative to the competitor items. In a complete orthogonal manipulation, the Correct search result, however, would resolve only after a thorough inspection of the snippet and URL components of competing search results, which would probably lead to even more analytical evaluation strategies.

Notwithstanding these limitations, it appears that when reading SERPs, students may consider two important heuristics: (1) eliminate a clearly irrelevant search result on the basis of a semantic analysis of its title information and (2) identify the most promising search results and conduct a full semantic analysis on them during re-inspection.

6. Acknowledgements

This research was funded by a grant 274022 from Academy of Finland to Paavo H. T. Leppänen. The authors would like to thank Sini Hjelm for organizing the data collection, and all of the research colleagues, research assistants and students involved in the research project.

7. References

Ashrafi-Amiri, N., & Al-Sader, J. (2016). Effects of confirmation bias on Web search engine results and a differentiation between non-assumptive versus assumptive search queries. (thesis). University of Gothenburg, Gothenburg, Sweden.

Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior*, *28*(2), 515–526.

Balatsoukas, P., & Ruthven, I. (2012). An Eye-tracking approach to the analysis of relevance judgments on the Web: The case of Google search engine. *Journal of the American Society for Information Science and Technology, 63,* 1728–1746.

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*(1), 193–211.

Bilal, D., & Gwizdka, J. (2016). Children’s eye-fixations on Google search results. *Proceedings of the Association for Information Science and Technology, 53,* 1–6.

Bilal, D., & Kirby, J. (2001). Factors influencing children’s and adults’ information seeking on the Web: Results of two studies. In *Proceedings of the ASIST Annual Meeting, 38,* (pp. 126–140).

Bowler, L. (2010). A taxonomy of adolescent metacognitive knowledge during the information search process. *Library & Information Science Research, 32,* 27–42.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education, 53,* 1207–1217.

Brumby, D. P., & Howes, A. (2008). Strategies for guiding interactive search: An empirical investigation into the consequences of label relevance for assessment and selection. *Human–Computer Interaction, 23,* 1–46.

Coiro, J. (2011). Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice, 50,* 107–115.

Dinet, J., Bastien, J. M., & Kitajima, M. (2010). What, where and how are young people looking for in a search engine results page?: Impact of typographical cues and prior domain knowledge. In *Conference Internationale Francophone Sur I'Interaction Homme-Machine (*pp. 105–112). *ACM*.

Dinet, J., Chevalier., A, Tricot, A. (2012). Information search activity: An overview. *European Review of Applied Psychology, 62,* 49–62.

Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through Grades 2, 3, and 8. *Journal of Educational Psychology, 107*(1), 126.

González-Caro, C., & Marcos, M. C. (2011). Different users and intents: An eyetracking analysis of Web search. *Proceedings of Web Search and Data Mining,* 9–12.

Gossen, T., Höbel, J., & Nürnberger, A. (2014). A comparative study about children’s and adults’ perception of targeted Web search engines. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1821–1824). ACM.

Gossen, T., Höbel, J., & Nürnberger, A. (2014). Usability and perception of young users and adults on targeted web search engines. *In Proceedings of the 5th Information Interaction in Context Symposium on - IIiX ’14* (pp. 18–27). ACM.

Graff M. (2005). Individual differences in hypertext browsing strategies. *Behaviour & Information Technology, 24*, 93–99.

Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In *Passive Eye Monitoring* (pp. 347–372). Springer Berlin Heidelberg.

Hardin, J. W. (2005). Generalized estimating equations (GEE). *Encyclopedia of statistics in behavioral science*.

Hirsh, S. G. (1999). Children’s relevance criteria and information seeking on electronic resources. *Journal of the Association for Information Science and Technology, 50,* 1265.

Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). Lukivakeuksien seulontamenetelmä nuorille ja aikuisille [*Dyslexia screening test for youth and adults*]. Jyväskylä, Finland: Niilo Mäki Instituutti.

Homish, G. G., & Edwards, E. P., Eiden, R. D., & Leonard, K. E. (2010). Analyzing family data: A GEE approach for substance use researchers. *Addictive Behaviors, 35,* 558–563.

Hyönä, J., Lorch, Jr. R. F, Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94,* 44.

Jochmann-Mannak,H. E., Huibers, T. W. C., Lentz, L. R., & Sanders, T. J. M. (2010). Children searching information on the Internet: Performance on children's interfaces compared to Google. *Towards accessible search systems - Workshop of the 33rd annual international ACM SIGIR conference on research and development in information retrieval* (pp. 27–35). ACM.

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of Internet-specific epistemic beliefs in laypersons' source evaluations and decisions during Web search on a medical issue. *Computers in Human Behavior, 29,* 1193–1203.

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of Web search results when using a list or a grid interface. *International Journal of Human-Computer Interaction, 30,* 177–191.

Keil, F. C., Kominsky, J. F. (2013). Missing links in middle school: Developing use of disciplinary relatedness in evaluating Internet search results. *Plos One* ,8.

Kiili C, Laurinen L, Marttunen M. (2008). Students evaluating Internet sources: From versatile evaluators to uncritical readers. *Journal of Educational Computing Research, 39,* 75–95.

Klenberg, L., Jämsä, S., Häyrinen, T., & Korkman, M. (2010). Keskittymiskysely, Käsikirja. [*Concentration Questionnairy, Handbook*]. Helsinki: Psykologien Kustannus.

Kroustallaki, D., Kokkinaki, T., Sideridis, G. D., & Simos, P. G. (2015). Exploring students' affect and achievement goals in the context of an intervention to improve Web searching skills. *Computers in Human Behavior, 49,* 156–170.

Lawless, K. A., & Kulikowich, J. M. (1996). Understanding Hypertext Navigation Through Cluster Analysis. *Journal of Educational Computing Research*, *14*, 385-399.

Leu, D. J., Coiro, J., Castek, J., Hartman, D. K., Henry, L. A. & Reinking, D. (2009). Research on Instruction and Assessment in the New Literacies of Online Reading Comprehension. *Comprehension Instruction: Research-Based Correct Practices, 2*, 321-346.

Leu, D. J., Kulikowich, J., Sedransk, N. & Coiro, J. (2009). Assessing Online Reading Comprehension: the ORCA Project. *Research Grant Funded by the US Department of Education, Institute of Education Sciences.*

Lindeman J. (1998). ALLU: Ala-Asteen Lukutesti [*ALLU: Reading Test for Elementary School*]. Turku, Finland: University of Turku. Center for Learning Research.

Lorigo L, Haridasan M, Brynjarsdóttir H, Xia L, Joachims T, Gay G, Et Al.. (2008). Eye Tracking and Online Search: Lessons Learned and Challenges Ahead. *Journal of the American Society for Information Science and Technology, 59,* 1041-1052.

Matsuda Y, Uwano H, Ohira M, Matsumoto KI. (2009). An Analysis of Eye Movements During Browsing Multiple Search Results Pages. In *Human-Computer Interaction (*pp. 121-130). New Trends. Springer Berlin Heidelberg.

Metzger M J, Flanagin a J, Medders RB. (2010). Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication, 60,* 413-439.

Nickerson, R. S. (1998). Confirmation Bias: a Ubiquitous Phenomenon in Many Guises. *Review of General Psychology, 2*, 175.

Pan B, Hembrooke H, Joachims T, Lorigo L, Gay G, Granka L. (2007). In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication*, *12*, 801-823.

Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*. Oxford University Press.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist, 40*, 235–244.

Raven JC, Court JH, Raven J. (1992). *Standard Progressive Matrices,* Oxford Psychologists Press, Oxford, UK.

Rayner, K. (Ed.). (2012). *Eye Movements in Reading: Perceptual and Language Processes*. Elsevier.

Rieh SY. (2002). Judgment of Information Quality and Cognitive Authority in the Web. *Journal of the American Society for Information Science and Technology, 53*, 145-161.

Rouet J-F, Britt MA. (2011). *Relevance Processes in Multiple Document Comprehension*. In: Mccrudden MT, Magliano JP, Schraw G, Editors. Text Relevance and Learning from Text Greenwich, CT: Information Age Publishing; pp. 19–52.

Rouet JF, Ros C, Goumi A, Macedo-Rouet M, Dinet J. (2011). The Influence of Surface and Deep Cues on Primary and Secondary School Students' Assessment of Relevance in Web Menus. *Learning and Instruction, 21*, 205-219.

Salmerón L, Kammerer Y, García-Carrión P. (2013). Searching the Web for Conflicting topics: Page and User Factors. *Computers in Human Behavior, 29*, 2161-2171.

Sharit, J, Hernández, M. A., Czaja, S. J., & Pirolli, P. (2008). Investigating the Roles of Knowledge and Cognitive Abilities in Older Adult Information Seeking on the Web. *Transactions on Computer-Human Interaction (TOCHI), 15(3).* ACM.

Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of Economics, 69*, 99–118.

Sung, Y. T., Wu, M. D., Chen, C. K., & Chang, K. E. (2015). Examining the online reading behavior and performance of fifth-graders: evidence from eye-movement data. *Frontiers in psychology*, *6*, 665.

Van Deursen, A. J., Görzig, A., Van Delzen, M., Perik, H. T., & Stegeman, A. G. (2014). Primary School Children’s Internet Skills: A Report on Performance Tests of Operational, Formal, Information, and Strategic Internet Skills. *International Journal of Communication, 8,* 1343-1365.

Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity on online search behavior of adolescents. *Journal of the Association for Information Science and Technology*, *68*(6), 1449-1461.

Walraven, A. Brand-Gruwel, S., & Boshuizen, H.P.A. (2008). Information problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, *24*, 623-648.

Walraven A, Brand-Gruwel S, Boshuizen HP. (2009) How Students Evaluate Information and Sources When Searching the World Wide Web for Information. *Computers & Education*, *52*, 234-246.

White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 3-12). ACM.

Wopereis, I. G. J. H., & van Merriënboer, J. J. G. (2011). Evaluating text-based information on the World Wide Web. *Learning and Instruction, 21*, 232–237.