

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

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)

Copyright: © 2018 Informa UK Limited, trading as Taylor & Francis Group

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

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>

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

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

6 Jarkko Hautala^{1,2}, Carita Kiili³, Yvonne Kammerer⁴, Otto Loberg², Sanna Hokkanen² and
7 Paavo H.T. Leppänen²

8

9

10 ¹Center for Applied Language Studies, University of Jyväskylä, Jyväskylä, Finland

11 ²Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

12 ³Department of Education, University of Oslo, Oslo, Norway

13 ⁴Leibniz-Institut für Wissensmedien, Tuebingen, Germany

14

15

16 Corresponding author: Jarkko Hautala, Center for Applied Language Studies, POB 35,

17 Mattilanniemi 6, FI-40014, University of Jyväskylä, Jyväskylä, Finland. Email:

18 jarkko.v.hautala@jyu.fi, Phone: [+358 40 805 3487](tel:+358408053487), Fax: +358-14-260 4400

19

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

22 (8634 words)

23 **Abstract**

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

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

43

44 **1. Introduction**

45 The ability to search for relevant information on the Internet using search engines is essential
46 for 21st century literacy. However, research indicates that students of various ages face
47 difficulties in locating and critically evaluating information (Bilal & Kirby, 2001; Leu,
48 Kulikowich, Sedransk, & Coiro, 2009; Leu, Coiro, Castek, Hartman, Henry, & Reinking,
49 2008). For a successful internet search, search engine users need to set an information need,
50 define appropriate search query terms, evaluate and select search results from the search
51 engine result page (SERP), and process the selected webpage(s) until their information need
52 is fulfilled (Brand-Gruwel, Wopereis, & Walraven, 2009; Dinet, Chevalier, & Tricot, 2012;
53 Sharit, Hernández, Czaja, & Pirolli, 2008). The ability to select relevant links from an SERP
54 is a key skill that can significantly increase the effectiveness of retrieving the desired
55 information (Argelagos, & Pifarre, 2012; Brand-Gruwel et al., 2009; Rouet & Britt, 2011;
56 Rieh, 2002). The present study applied eye-tracking recordings to examine what kinds of
57 evaluation strategies sixth graders' spontaneously applied when reading search results.

58 **1.1. Evaluation of Internet search results**

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

65 Adult readers generally use efficient "satisficing" strategies (cf. Simon, 1955;
66 satisficing = a combination of satisfy and suffice) when conducting Internet searches (Pirolli,
67 2007). For instance, they do not evaluate all links and the information available for them, but
68 are likely to stop at the search result they consider "good enough" (Lorigo et al., 2008). They

69 tend to first skim through SERPs and look at the first few search results (Pan et al., 2007;
70 Kammerer & Gerjets, 2014) before scrolling further down, proceeding to the next SERP, or
71 refining the query (Lorigo et al., 2008). These findings suggest that people often evaluate the
72 success of their search query before engaging in a detailed evaluation of the search results on
73 the SERP. In addition, when asked to bookmark webpages for further study, they often also
74 select results located further down the SERP (Salmerón et al., 2013).

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

82 Within a search result, viewers spend most of their time reading title lines and pay
83 less attention to text snippets and URLs (Dinet, Bastien, & Kitajima, 2010; Granka et al.,
84 2008), particularly when letter-normalized viewing times are being analyzed. In addition to
85 evaluating the semantic relevance of a search result, the expected quality of information (or
86 credibility of an information source) can also play a role in a user's selection decisions (e.g.,
87 Balatsoukas & Ruthven, 2012; Kammerer & Gerjets, 2014; Rieh, 2002). Because anyone can
88 virtually publish any information on the Web, the quality of information varies widely and
89 many websites provide incomplete and/or inaccurate information. In a search result, for
90 example, the URL provides cues about the credibility of the information source (e.g.,
91 Kammerer, Bråten, Gerjets, & Strømsø, 2013). Accordingly, tasks that afford finding a
92 specific webpage or that require finding credible information lead to more URL and snippet
93 viewing (González-Caro & Marcos, 2011; Matsuda et al., 2009). A study that integrated user

94 selections, eye movements, and think-aloud protocols, found that individuals used
95 appropriate relevancy criteria, for example, topic relevance and scope for titles as well as
96 information quality and domain expertise for URLs, when exploring search result
97 components (Balatsoukas & Ruthven, 2012).

98 **1.2. Internet search evaluation by adolescents**

99 Generally, presumably due to the nonlinear nature of online reading (Sung, Wu,
100 Chen, & Chang, 2015) acquiring effective Internet search skills takes years to develop and is
101 greatly facilitated by proper instruction (Bannert & Reimann, 2012; Van Deursen et al.,
102 2014). Already sixth-grade students could evaluate the relevancy of search results with
103 respect to a given search problem, but this skill was not fully developed until the eighth grade
104 (Keil & Kominsky, 2013). In another study, SERP reading became more efficient from sixth
105 to eighth grade, with faster response times and fewer clicks on search results (Gwidzka &
106 Bilal, 2017) . Moreover, several studies have suggested that adolescents do not typically
107 assess the reliability or credibility of information during web searches (Jochmann-Mannak,
108 Huibers, Lentz, & Sanders, 2010; Kiili, Laurinen, & Marttunen, 2008; Walraven, Brand-
109 Gruwel, & Boshuizen, 2009). For example, seventh grade students continued to make search
110 result selections on the basis of superficial cues, such as boldfaced keywords, instead of
111 semantic information (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011).

112 Eye movement studies have shown that adolescent begin by reading almost all results
113 listed on an SERP, after which they pay more individual attention to them (Bilal & Gwidzka,
114 2016). In addition, eighth graders have been shown to start reading SERPs more consistently
115 from the first ranked search result to the bottom, while sixth graders made more premature
116 clicks on search results before reading them (Bilal & Gwidzka, 2016). Further, younger
117 children looked at fewer snippets and instead looked more at thumbnail images, suggesting
118 that children find it difficult to read long texts in the SERPs (Gossen, Höbel, & Nürnberger,

119 2014). Eye movement analysis of SERP reading among fifth, seventh, ninth, and eleventh
120 grade students performing simple fact-finding tasks indicated that the typographical cueing of
121 boldfaced search words seemed to attract the younger readers' gaze (Dinet et al., 2010; cf.
122 also Rouet et al., 2011). Older students, on the other hand, were attracted by such cues only
123 when the information search task was based on unfamiliar topics.

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

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

134 **1.3. Interactive search framework**

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

142 Applying this interactive search model to SERP reading suggests that encountering a
143 highly relevant search result will inhibit the processing of subsequent search results. In

144 addition, the interactive search process may affect the manner in which the components of a
145 search result are inspected. Readers typically begin reading a search result from the title. If
146 the title is relevant, they may proceed to reading the search result's snippet and/or URL
147 components. If not, they may eliminate this search result without inspecting its snippet and
148 URL components. At some point, especially when the search results on a SERP become
149 exceedingly irrelevant to the task at hand, the readers may enter a re-inspection phase to re-
150 evaluate the search results considered the most relevant. During this re-inspection, they may
151 re-read only the titles or deepen their evaluation using information provided in the search
152 result's snippet and/or URL components.

153 **1.4. Research questions and hypotheses**

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

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

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

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

167 RQ 2: What information sources do the students pay attention to and which evaluation
168 strategies do they use during their selection? We operationalized this as the first-pass (i.e.,

169 initial inspection) and second-pass (i.e., re-inspection) looking probabilities of various
170 components in search results (as defined in RQ1).

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

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

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

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

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

190 **2. Materials and methods**

191 **2.1. Participants**

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

199 1) Reading difficulties, which were defined as a reading fluency performance score below the
200 15th percentile (based on the factor score derived from three reading measures: Lindeman,
201 1998; Eklund, Torppa, Aro, Leppänen, & Lyytinen, 2015; Holopainen, Kairaluoma, Nevala,
202 Ahonen, & Aro, 2004) or as a parental report of the student with a reading disability
203 diagnosis.

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

206 3) A nonverbal IQ performance result below the 7th percentile based on a 15-minute,
207 30-item version of the Raven matrices (Raven, Court, and Raven, 1992).

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

210 **2.2. Apparatus**

211 Eye movements were recorded using a table-mounted EyeLink 1000 eye-tracker (SR
212 Research) with forehead- and chin-rest. The stimuli were presented on a Dell Precision
213 T5500 workstation with an Asus VG-236 monitor (1920 x 1080, 120 Hz, 52 x 29 cm) at 60

214 cm viewing distance. 13-point calibration with a one-degree visual angle as the acceptance
215 criterion was applied. We conducted the calibration prior to the experiment and repeated it
216 between trials when visible (a) head movements were made, (b) a drift was seen on the
217 researcher's screen where the subjects' eye movements were overlaid on experimental
218 stimuli, or (c) the calibration error exceeded .30 visual degrees.

219 **2.3. Tasks and Materials**

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

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

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

235

236

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

[Learn now about Gold Rush](#)

<http://www.historychannel.fi/gold>

Alaska Gold Rush is a television program running in Discovery –channel. It shows the life of gold miners as they seek gold all around the Alaska...

[Gold Rush consequences](#)

<http://www.history.fi/goldrush>

Gold rush had diverse consequences for the Indians. Many kinds of theories has been suggested about how the Gold Rush affected the life of Indians. These web pages...

[Robot’s Gold Rush](#)

<http://m.player.fi/news/robots>

Robot's Gold Rush hits on PlayStation! Started as 3DS -game, Steam World Dig has been slowly but steadily conquering the world...

[History of Gold Rush](#)

<http://www.tv-guide.fi/goldrush>

Gold Rush was anticipated in May 1848, when a shop-keeper Sam Brannan found a bottle covered with gold dust from ground. It...

241

242

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

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

250

251 **Table 1. Descriptions of stimuli including number of written characters and relevancy**
 252 **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

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

256

257 **2.3.1. Prior knowledge**

258 The information-seeking tasks were unlikely to be solved with a sixth grader's prior
259 knowledge. The degree of knowledge, however, was self-evaluated prior to the experiment
260 using the following types of question: e.g., how much do you know about the threats to coral
261 reefs? The response options were (1) I know nothing (47% of responses), (2) I know little
262 (26%), (3) I know some (21%), and (4) I know a lot (7%). Students' self-reported prior
263 knowledge was not correlated with their accuracy in selecting the correct search result, $r(36)$
264 $= .234, p = .170$.

265 **2.3.1. Internet search experience**

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

276 **2.4. Procedure**

277 One research assistant accompanied the participant in a laboratory room while the other
278 assistant controlled the devices in the control room. The sequence of activities for a
279 participant was: prior knowledge questionnaire, task instructions on paper, adjustments of
280 the eye tracker's table height and the forehead- and chin-rest, calibration, practice task, and
281 finally the ten experimental information search tasks, including at least one short or several

282 breaks of a few minutes, depending on individual needs. Calibration was repeated after the
283 breaks. The students completed the tasks using a mouse. The duration of each experiment
284 session varied from 45 to 90 minutes on the basis of the participant.

285 **2.5. Eye-movement data processing**

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

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

302 **2.6. Data analyses**

303 **2.6.1. General viewing strategies**

304 For a detailed analysis of the students' viewing strategies, we analyzed the first- and
305 second-pass looking probabilities¹. A first-pass look was defined as a first inspection of a
306 search result, and a second-pass look included all the later inspections of a search result (i.e.,
307 after having inspected or re-inspected one or several other search results in between).
308 Because the employed statistical method required integer values (counts), we conducted the
309 analyses on the basis of the number of tasks (0–9) in which the participant looked at a
310 component of each search result type (see section Tasks and Materials). To derive more
311 illustrative looking probability values (0–1) for the figures, we divided these task counts by
312 the total number of tasks (9). The first-pass looking probability indexed the likelihood of a
313 search result component being looked at during the initial inspection of a search result. The
314 second-pass looking probability indexed the likelihood of a component being viewed when
315 re-inspecting a search result, including second and subsequent passes.

316 We conducted a generalized estimating equation analysis (GEE; for more information,
317 see Hardin, 2005; Homish, Edwards, Eiden, & Leonard, 2010) with SPSS to examine the
318 data. We selected the GEE approach instead of an analysis of variance (ANOVA) approach,
319 because the looking probability variables were not normally distributed. Within the GEE
320 analysis, we applied the robust estimator of covariance matrix, exchangeable correlation
321 structure, Poisson loglinear model for counts, and hybrid method for parameter estimation.
322 The analysis comprised a four-level within-subject factor of the search result type (i.e.,
323 Distractor, Irrelevant-URL, Irrelevant-Snippet, and Correct) and a three-level within-subject
324 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.

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

327 **2.6.2 Effects of correct search result position on competitor viewing**

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

337 **2.6.3. Individual viewing strategies**

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

343 **4. Results**

344 **4.1. Search result selection**

345 The students chose the Correct search result with high accuracy ($M = 81.0\%$, $SD =$
346 17.0%). The probability of selecting the Irrelevant-URL was $M = 6.8\%$ ($SD = 10.0\%$) and
347 that of selecting the Irrelevant-Snippet was $M = 12.0\%$ ($SD = 11.0\%$). The Wilcoxon's
348 signed-ranks test for two related samples indicated that the participants chose Irrelevant-

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

354 **4.2. General viewing strategies**

355 On average, the snippet was looked at the longest ($M = 3.32$ s, $SD = 1.76$), followed by
356 the title ($M = 1.55$ s, $SD = .49$) and URL ($M = .50$ s, $SD = .32$). Notably, 55% of the URLs
357 were never looked at, while this was true for only 11% of the snippets and 6% of the titles.
358 The Correct search results were looked at the longest ($M = 7.58$ s, $SD = 3.13$, with by average
359 2.8 viewing occasions) and both the Irrelevant-URL and Irrelevant-Snippet were looked at
360 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
361 viewing occasions), while the Distractor was looked at for the least duration ($M = 3.44$ s, SD
362 = 1.55, with 1.8 viewing occasions).

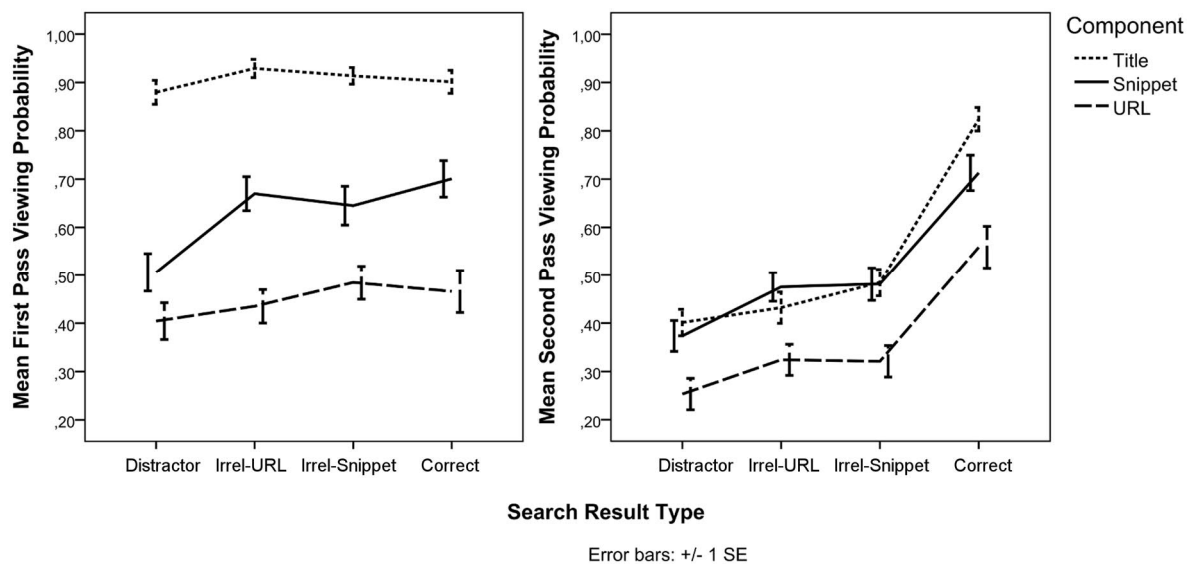
363 Figure 2 presents the probabilities of looking at each search result component (title,
364 URL, and snippet) within different search result types.

365

366

367 **Figure 2. Probabilities of looking at different search result types and their components**368 **during first- (left panel) and second- (middle panel) pass viewing.** Irrel is an abbreviation

369 for Irrelevant.



370

371 **4.3. First-pass looking probability**

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

378 **4.4. Second-pass looking probability**

379 The main effects of both the search result type ($\chi^2(3, N = 36) = 281.7, p < .001$) and
 380 components ($\chi^2(2, N = 36) = 40.5, p < .001$) were significant. These main effects were
 381 accompanied by a Type x Component interaction ($\chi^2(6, N = 36) = 14.3, p = .026$). In general,

382 all components of the Correct search result were looked at with much higher probability than
383 those of the other search result types ($ps < .001$), which were looked at with equal
384 probability. An exception was the snippet component of the Distractor search result, which
385 was less likely to be looked at than the competitors' snippet components ($ps \leq .006$).

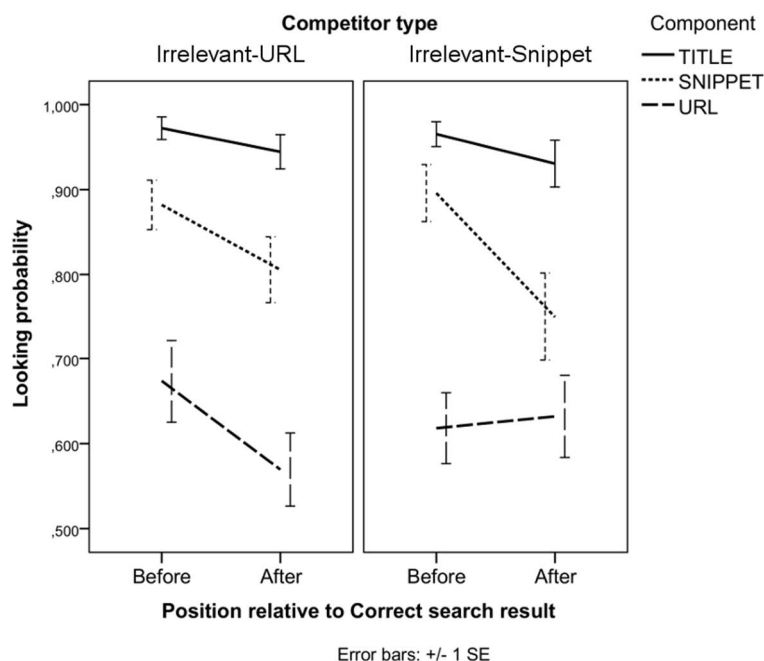
386 **4.5. Effects of Correct search result position on competitor viewing**

387 The significant main effect of the Correct search result position ($\chi^2(1, N = 36) = 10.6, p$
388 $= .001$) was qualified by a significant three-way interaction of the Correct position (before,
389 after), Competitor type (Irrelevant-Snippet and Irrelevant-URL), and Component (Title,
390 URL, and Snippet) ($\chi^2(2, N = 36) = 6.20, p = .045$). Titles were looked at with equal
391 probability, regardless of whether the competitor preceded or followed the Correct search
392 result (Fig 3). Irrelevant-Snippet's snippet component was less likely to be looked at if it
393 appeared after the Correct search result, ($p = .002$). Irrelevant-URL's URL ($p = .054$) and
394 snippet ($p = 0.13$) components were less likely to be viewed if they appeared after the Correct
395 search result. These results suggest that after the students read the search results that strongly
396 matched the information search task, they inspected the subsequent search results in less
397 detail.

398

399

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



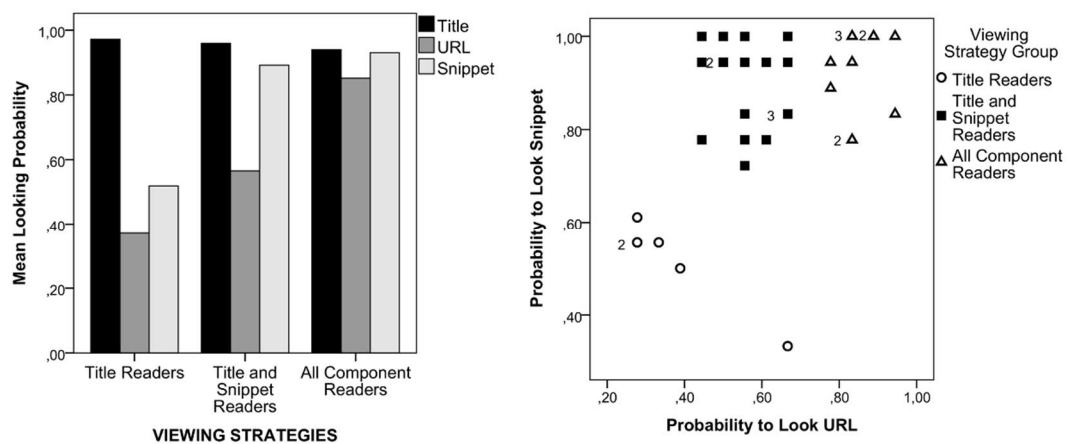
402

403 4.6. Differences in viewing strategies

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

415 components. An exception was that All Component readers and Title and Snippet readers
 416 looked at snippet components with an equal probability ($p > .99$).

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



423
 424 Table 2 lists the performance of the different viewing strategy groups. The groups
 425 differed in their accuracy of selecting the Correct search result ($\chi^2(2, N = 36) = 6.79, p =$
 426 $.033$), with Title readers making more errors than the two other groups ($ps \leq .007$), while All
 427 Component readers as well as Title and Snippet readers performed equally well ($p > .99$).
 428 Title readers also responded more quickly than the two other groups ($ps \leq .009$). There was
 429 no difference between cluster groups in prior knowledge ($p = .534$) and Internet search
 430 experience ($p \geq .85$), screening variables of reading fluency ($p = .394$) and nonverbal
 431 intelligence (Raven; $p = .179$), or the attention deficit scale ($p = .572$).

433 **Table 2. Performance of viewing strategy groups with means and standard deviations in**
 434 **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)

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

441 5. Discussion

442 This study aimed to explore strategies that typically developing sixth graders (12 to
 443 13 year olds in the last year of their primary school) use to inspect and select search engine

444 results, as a function of information value of the search result components (i.e., title, URL,
445 and snippet) both during inspection (first-pass viewing) and re-inspection (second-pass
446 viewing).

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

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

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

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

492 biased to confirm their opinions, even when faced with counterfactual evidence (Nickerson,
493 1998; White, 2013; Ashraf-Amri & Al-Sader, 2016).

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

506 Finally, the present study adds to the understanding of inter-individual differences in
507 the evaluation of search results (cf. Graff, 2005; Lawless & Kulikowich, 1996). As shown by
508 the cluster analysis, one group of students (i.e., the Title readers, comprising one-sixth of the
509 students) did not use the evaluation strategies as effectively as the other two groups. This
510 particular group of students predominantly looked only at titles and did not view other search
511 result components (i.e., snippet and URL) to make confirmatory or eliminative decisions.
512 Consequently, they performed less successfully than the other two groups (57% vs. 86% and
513 86%) who had a high probability of looking at both titles and snippets or all three
514 components, respectively. As these groups of students did not differ in the assessed cognitive
515 skills, or their self-reported prior knowledge or experience in conducting Internet searches, it
516 seems likely that these strategies result from students' individual learning history. Therefore,

517 it seems reasonable that students who do not utilize the snippet and URL -components in
518 their search result evaluation, would probably benefit from a targeted instruction on this skill
519 (cf. Coiro, 2011; Sung et al., 2015).

520 **5.1. Theoretical implications**

521 Drawing on various theoretical accounts of information searches (Brumby & Howes,
522 2008; Dinet et al., 2012; Metzger et al., 2010), this study offers new evidence about semantic
523 control on looking behavior during information searches and SERP reading. Such behavior is
524 in line with the findings that reading comprehension processes sensitively affect which
525 portions of text are reread in particular during normal reading (Rayner, 2012). Our findings
526 add to this knowledge that when reading hierarchically organized materials, such as SERPs,
527 readers also routinely make decisions not to read certain parts of text, which are most likely
528 to be irrelevant for the task at hand.

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

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

549 **5.2. Limitations**

550 The present study adopted a highly controlled experimental approach in the laboratory
551 setting to examine the reading and evaluation of Internet search results. Consequently, some
552 caution is warranted in applying the findings into practice (Wopereis & van Merriënboer,
553 2011). It is likely that when searching information on the open Web the strategies reported
554 here will be applied to only a subset of search results included in SERPs. For example, users
555 may evaluate only a subset of the highest ranked search results and may discontinue the
556 evaluation process when the relevancy of the titles decreases (Bilal & Gwidzka, 2016; Pan et
557 al., 2007).

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

564 Finally, some of the findings might be specific to the task requirements and materials.
565 In the present study, the Correct search results had slightly higher relevancy ratings for each
566 component than the competing search results. Students apparently identified the Correct

567 search result already during first-pass reading, as they returned to this item with a higher
568 probability relative to the competitor items. In a complete orthogonal manipulation, the
569 Correct search result, however, would resolve only after a thorough inspection of the snippet
570 and URL components of competing search results, which would probably lead to even more
571 analytical evaluation strategies.

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

576 **6. Acknowledgements**

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

581 **7. References**

- 582 Ashrafi-Amiri, N., & Al-Sader, J. (2016). Effects of confirmation bias on Web search engine
583 results and a differentiation between non-assumptive versus assumptive search queries.
584 (thesis). University of Gothenburg, Gothenburg, Sweden.
- 585 Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in
586 secondary education through embedded instruction. *Computers in Human Behavior*, 28(2),
587 515–526.

- 588 Balatsoukas, P., & Ruthven, I. (2012). An Eye-tracking approach to the analysis of relevance
589 judgments on the Web: The case of Google search engine. *Journal of the American*
590 *Society for Information Science and Technology*, 63, 1728–1746.
- 591 Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through
592 prompts. *Instructional Science*, 40(1), 193–211.
- 593 Bilal, D., & Gwizdka, J. (2016). Children’s eye-fixations on Google search
594 results. *Proceedings of the Association for Information Science and Technology*, 53, 1–6.
- 595 Bilal, D., & Kirby, J. (2001). Factors influencing children’s and adults’ information seeking
596 on the Web: Results of two studies. In *Proceedings of the ASIST Annual Meeting*, 38, (pp.
597 126–140).
- 598 Bowler, L. (2010). A taxonomy of adolescent metacognitive knowledge during the
599 information search process. *Library & Information Science Research*, 32, 27–42.
- 600 Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information
601 problem solving while using Internet. *Computers & Education*, 53, 1207–1217.
- 602 Brumby, D. P., & Howes, A. (2008). Strategies for guiding interactive search: An empirical
603 investigation into the consequences of label relevance for assessment and
604 selection. *Human–Computer Interaction*, 23, 1–46.
- 605 Coiro, J. (2011). Talking about reading as thinking: Modeling the hidden complexities of
606 online reading comprehension. *Theory Into Practice*, 50, 107–115.
- 607 Dinet, J., Bastien, J. M., & Kitajima, M. (2010). What, where and how are young people
608 looking for in a search engine results page?: Impact of typographical cues and prior
609 domain knowledge. In *Conference Internationale Francophone Sur l'Interaction Homme-*
610 *Machine* (pp. 105–112). ACM.
- 611 Dinet, J., Chevalier., A, Tricot, A. (2012). Information search activity: An
612 overview. *European Review of Applied Psychology*, 62, 49–62.

- 613 Eklund, K., Torppa, M., Aro, M., Leppänen, P. H. T., & Lyytinen, H. (2015). Literacy skill
614 development of children with familial risk for dyslexia through Grades 2, 3, and 8. *Journal*
615 *of Educational Psychology, 107*(1), 126.
- 616 González-Caro, C., & Marcos, M. C. (2011). Different users and intents: An eyetracking
617 analysis of Web search. *Proceedings of Web Search and Data Mining, 9–12*.
- 618 Gossen, T., Höbel, J., & Nürnberger, A. (2014). A comparative study about children's and
619 adults' perception of targeted Web search engines. In *Proceedings of the SIGCHI*
620 *Conference on Human Factors in Computing Systems* (pp. 1821–1824). ACM.
- 621 Gossen, T., Höbel, J., & Nürnberger, A. (2014). Usability and perception of young users and
622 adults on targeted web search engines. In *Proceedings of the 5th Information Interaction*
623 *in Context Symposium on - IiX '14* (pp. 18–27). ACM.
- 624 Graff M. (2005). Individual differences in hypertext browsing strategies. *Behaviour &*
625 *Information Technology, 24*, 93–99.
- 626 Granka, L., Feusner, M., & Lorigo, L. (2008). Eye monitoring in online search. In *Passive*
627 *Eye Monitoring* (pp. 347–372). Springer Berlin Heidelberg.
- 628 Hardin, J. W. (2005). Generalized estimating equations (GEE). *Encyclopedia of statistics in*
629 *behavioral science*.
- 630 Hirsh, S. G. (1999). Children's relevance criteria and information seeking on electronic
631 resources. *Journal of the Association for Information Science and Technology, 50*, 1265.
- 632 Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). Lukivakeuksien
633 seulontamenetelmä nuorille ja aikuisille [*Dyslexia screening test for youth and adults*].
634 Jyväskylä, Finland: Niilo Mäki Instituutti.
- 635 Homish, G. G., & Edwards, E. P., Eiden, R. D., & Leonard, K. E. (2010). Analyzing family
636 data: A GEE approach for substance use researchers. *Addictive Behaviors, 35*, 558–563.

- 637 Hyönä, J., Lorch, Jr. R. F, Kaakinen, J. K. (2002). Individual differences in reading to
638 summarize expository text: Evidence from eye fixation patterns. *Journal of Educational*
639 *Psychology, 94*, 44.
- 640 Jochmann-Mannak,H. E., Huibers, T. W. C., Lentz, L. R., & Sanders, T. J. M. (2010).
641 Children searching information on the Internet: Performance on children’s interfaces
642 compared to Google. *Towards accessible search systems - Workshop of the 33rd annual*
643 *international ACM SIGIR conference on research and development in information*
644 *retrieval* (pp. 27–35). ACM.
- 645 Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of Internet-specific
646 epistemic beliefs in laypersons’ source evaluations and decisions during Web search on a
647 medical issue. *Computers in Human Behavior, 29*, 1193–1203.
- 648 Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source
649 trustworthiness in the selection of Web search results when using a list or a grid interface.
650 *International Journal of Human-Computer Interaction, 30*, 177–191.
- 651 Keil, F. C., Kominsky, J. F. (2013). Missing links in middle school: Developing use of
652 disciplinary relatedness in evaluating Internet search results. *Plos One* ,8.
- 653 Kiili C, Laurinen L, Marttunen M. (2008). Students evaluating Internet sources: From
654 versatile evaluators to uncritical readers. *Journal of Educational Computing Research, 39*,
655 75–95.
- 656 Klenberg, L., Jämsä, S., Häyrynen, T., & Korkman, M. (2010). Keskitymiskysely,
657 Käsikirja. [*Concentration Questionnairy, Handbook*]. Helsinki: Psykologien Kustannus.
- 658 Kroustallaki, D., Kokkinaki, T., Sideridis, G. D., & Simos, P. G. (2015). Exploring students’
659 affect and achievement goals in the context of an intervention to improve Web searching
660 skills. *Computers in Human Behavior, 49*, 156–170.

- 661 Lawless, K. A., & Kulikowich, J. M. (1996). Understanding Hypertext Navigation Through
662 Cluster Analysis. *Journal of Educational Computing Research, 14*, 385-399.
- 663 Leu, D. J., Coiro, J., Castek, J., Hartman, D. K., Henry, L. A. & Reinking, D. (2009).
664 Research on Instruction and Assessment in the New Literacies of Online Reading
665 Comprehension. *Comprehension Instruction: Research-Based Correct Practices, 2*, 321-
666 346.
- 667 Leu, D. J., Kulikowich, J., Sedransk, N. & Coiro, J. (2009). Assessing Online Reading
668 Comprehension: the ORCA Project. *Research Grant Funded by the US Department of*
669 *Education, Institute of Education Sciences.*
- 670 Lindeman J. (1998). ALLU: Ala-Asteen Lukutesti [ALLU: Reading Test for Elementary
671 School]. Turku, Finland: University of Turku. Center for Learning Research.
- 672 Lorigo L, Haridasan M, Brynjarsdóttir H, Xia L, Joachims T, Gay G, Et Al.. (2008). Eye
673 Tracking and Online Search: Lessons Learned and Challenges Ahead. *Journal of the*
674 *American Society for Information Science and Technology, 59*, 1041-1052.
- 675 Matsuda Y, Uwano H, Ohira M, Matsumoto KI. (2009). An Analysis of Eye Movements
676 During Browsing Multiple Search Results Pages. In *Human-Computer Interaction* (pp.
677 121-130). New Trends. Springer Berlin Heidelberg.
- 678 Metzger M J, Flanagin a J, Medders RB. (2010). Social and Heuristic Approaches to
679 Credibility Evaluation Online. *Journal of Communication, 60*, 413-439.
- 680 Nickerson, R. S. (1998). Confirmation Bias: a Ubiquitous Phenomenon in Many
681 Guises. *Review of General Psychology, 2*, 175.
- 682 Pan B, Hembrooke H, Joachims T, Lorigo L, Gay G, Granka L. (2007). In Google We Trust:
683 Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated*
684 *Communication, 12*, 801-823.

- 685 Pirolli, P. (2007). *Information foraging theory: Adaptive interaction with information*.
686 Oxford University Press.
- 687 Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive
688 aspects of online inquiry through software-based scaffolding. *Educational Psychologist*,
689 *40*, 235–244.
- 690 Raven JC, Court JH, Raven J. (1992). *Standard Progressive Matrices*, Oxford Psychologists
691 Press, Oxford, UK.
- 692 Rayner, K. (Ed.). (2012). *Eye Movements in Reading: Perceptual and Language Processes*.
693 Elsevier.
- 694 Rieh SY. (2002). Judgment of Information Quality and Cognitive Authority in the Web.
695 *Journal of the American Society for Information Science and Technology*, *53*, 145-161.
- 696 Rouet J-F, Britt MA. (2011). *Relevance Processes in Multiple Document Comprehension*. In:
697 Mccrudden MT, Magliano JP, Schraw G, Editors. *Text Relevance and Learning from Text*
698 Greenwich, CT: Information Age Publishing; pp. 19–52.
- 699 Rouet JF, Ros C, Goumi A, Macedo-Rouet M, Dinet J. (2011). The Influence of Surface and
700 Deep Cues on Primary and Secondary School Students' Assessment of Relevance in Web
701 Menus. *Learning and Instruction*, *21*, 205-219.
- 702 Salmerón L, Kammerer Y, García-Carrión P. (2013). Searching the Web for Conflicting
703 topics: Page and User Factors. *Computers in Human Behavior*, *29*, 2161-2171.
- 704 Sharit, J, Hernández, M. A., Czaja, S. J., & Pirolli, P. (2008). Investigating the Roles of
705 Knowledge and Cognitive Abilities in Older Adult Information Seeking on the
706 Web. *Transactions on Computer-Human Interaction (TOCHI)*, *15*(3). ACM.
- 707 Simon, H. A. (1955). A Behavioral Model of Rational Choice. *Quarterly Journal of*
708 *Economics*, *69*, 99–118.

- 709 Sung, Y. T., Wu, M. D., Chen, C. K., & Chang, K. E. (2015). Examining the online reading
710 behavior and performance of fifth-graders: evidence from eye-movement data. *Frontiers*
711 *in psychology*, 6, 665.
- 712 Van Deursen, A. J., Görzig, A., Van Delzen, M., Perik, H. T., & Stegeman, A. G. (2014).
713 Primary School Children's Internet Skills: A Report on Performance Tests of Operational,
714 Formal, Information, and Strategic Internet Skills. *International Journal of*
715 *Communication*, 8, 1343-1365.
- 716 Walhout, J., Oomen, P., Jarodzka, H., & Brand-Gruwel, S. (2017). Effects of task complexity
717 on online search behavior of adolescents. *Journal of the Association for Information*
718 *Science and Technology*, 68(6), 1449-1461.
- 719 Walraven, A. Brand-Gruwel, S., & Boshuizen, H.P.A. (2008). Information problem solving:
720 A review of problems students encounter and instructional solutions. *Computers in Human*
721 *Behavior*, 24, 623-648.
- 722 Walraven A, Brand-Gruwel S, Boshuizen HP. (2009) How Students Evaluate Information
723 and Sources When Searching the World Wide Web for Information. *Computers &*
724 *Education*, 52, 234-246.
- 725 White, R. (2013). Beliefs and biases in web search. In *Proceedings of the 36th international*
726 *ACM SIGIR conference on Research and development in information retrieval* (pp. 3-12).
727 ACM.
- 728 Wopereis, I. G. J. H., & van Merriënboer, J. J. G. (2011). Evaluating text-based information
729 on the World Wide Web. *Learning and Instruction*, 21, 232–237.
- 730
- 731