Hahnel, Carolin; Goldhammer, Frank; Kröhne, Ulf; Naumann, Johannes

The role of reading skills in the evaluation of online information gathered from search engine environments

formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:
formally and content revised edition of the original source in:
Computers in human behavior (2018) 78, S. 223-234, 10.1016/j.chb.2017.10.004

The role of reading skills for the evaluation of online information gathered from search engine environments

Carolin Hahnel [ab], Frank Goldhammer [ab], Ulf Kröhne [a], & Johannes Naumann [c]

[a] German Institute for International Educational Research (DIPF), Frankfurt am Main, Germany

[b] Centre for International Student Assessment (ZIB), Germany

[c] Goethe University Frankfurt am Main, Germany

Correspondence concerning this article should be addressed to:

Carolin Hahnel, German Institute for International Educational Research (DIPF)

Schlossstraße 29, 60486 Frankfurt am Main, Germany

Phone: +49 69 24 708 727

E-mail: hahnel@dipf.de

Acknowledgements

This research was funded by the Centre for International Student Assessment (ZIB) in Germany.

Abstract

A critical evaluation to find useful information is essential when doing a web search. In this study, we investigated this evaluation skill of secondary school students, based on their behavior in selecting hyperlinks from a search engine result page (SERP). To clarify the role of reading for the evaluation of online information, we additionally assessed students' individual reading skills on word, sentence, and text level. Data of 416 15-year-old students participating in a computer based German add-on study to the *Programme for International Student Assessment* (PISA) in 2012 were investigated. Using generalized linear mixed models (GLMMs), effects of reading skills on the skill to evaluate online information were found. These effects were influenced by the similarity of SERP hyperlinks in relevance and students' navigation to subsequent SERPs or websites. The results are interpreted as skilled readers are able to allocate their cognitive resources more efficiently than less skilled readers when evaluating online information. Implications are discussed in terms of underlying cognitive processes in making web search decisions.

Keywords: evaluating online information; hyperlink selection; reading; semantic similarity; navigation behavior.

Highlights:

- Secondary school students' skill to evaluate online information was examined.
- Reading skills on word, sentence, and text level predicted students' link selection.
- Reading skills allowed better selection, the more the links varied semantically.
- Students selected more adequate hyperlinks when they accessed further information.
- Reading skills facilitated the evaluation when further information was accessed.

Search engines have become a ubiquitous tool in using the World Wide Web. As a broad information resource, they provide easy access for web users who seek information for any purpose, such as educational, occupational, and private. Search engine environments are frequently used by secondary school students (e.g., Feierabend, Karg & Rathgeb, 2013; OECD [Organisation for Economic Co-operation and Development], 2011).Yet, they reveal a vast amount of information considerably varying in relevance and quality. A critical evaluation of information in terms of relevance and credibility is crucial since an incorrect use of information can result in inappropriate decisions and serious consequences (Brand-Gruwel & Stadtler, 2011). Many students, though, show difficulties in selecting adequate online information (e.g., Brand-Gruwel, Wopereis & Walraven, 2009; Lucassen, Muilwijk, Noordzij & Schraagen, 2013; Walraven, Brand-Gruwel & Boshuizen, 2008).

Evaluating the appropriateness of information for solving a search task requires information to be identified and comprehended. Students and even adults differ in their reading proficiency on word, sentence, and text level (e.g., Perfetti 2007; Sabatini, 2015) which raises questions of whether and how reading skills affect their selection of online information. Therefore, the present study seeks to shed light on the role of reading as a conditioning factor of success when evaluating information from search engine result pages (SERPs). We investigated whether or not hierarchically related reading skills on word, sentence, and text level affect students' evaluation of online information. Furthermore, we examined if these reading effects were influenced by characteristics of SERP hyperlinks and individual user behavior of students. As characteristic of SERP hyperlinks, we considered how similar they were in terms of their relevance to a search task. As individual user behavior, students' navigation behavior to other SERPs as well as to websites connected to SERP hyperlinks was investigated.

1.1 Reading and processing web search information

An information-based web search usually starts by identifying a gap of knowledge (cf. Brand-Gruwel et al., 2009; Gerjets, Kammerer & Werner, 2011). Web users define a search task, verbalize a query, and enter it into a chosen search engine like Google. A SERP appears that lists several text abstracts with hyperlinks leading to websites of potential interest. Search engines offer a first classification but it is people who need to decide if the listed information meets the requirements of their search task. Therefore, web users are assumed to use criteria of information relevance and credibility affecting their processing and efforts in evaluating online information (Flanagin & Metzger, 2007; Metzger, 2007; Hilligoss & Rieh, 2008; Rieh, 2002). According to dual processing theory (Evans, 2008; Wirth, Böcking, Karnowski, & von Pape, 2007), web users will evaluate SERP information either heuristically or systematically. Systematic processing means web users perform an extensive evaluation of collected information, based on various characteristics (e.g., topic relevance, trustworthiness, completeness; Salmerón, Kammerer & García-Carrión, 2013). These processes are slow and deliberate, making them cognitively demanding. Especially when dealing with the "information flood" provided by search engines, mental costs of thorough search might be severe (Rieh, Kim, & Markey, 2012). In contrast, heuristic processing is fast and automatic and demands fewer cognitive resources to operate (Gigerenzer & Gaissmaier, 2011), but only focuses on limited characteristics of existing information (e.g., does information confirms one's expectation; Metzger, Flanagin, & Medders, 2010) that might be insufficient for adequate information selection.

When evaluating online information, reading is the essential component in receiving and processing written information. While reading, individuals are assumed to actively construct a mental representation that integrates a text representation of word structures and propositional meaning with one's general knowledge (Kintsch, 1998; Perfetti & Stafura, 2014; Rouet, 2006). This process is supposed to be semi-hierarchically organized on word, sentence, and text level (e.g., Hamilton, Freed & Long, 2013; Perfetti & Stafura, 2014; Richter, Isberner, Naumann & Kutzner, 2012). For experienced readers, basic reading activities like lexical access and propositional integration occur automatically (Perfetti, 2007; Samuels & Flor, 1997; Walczyk, 2000). Nevertheless, controlled processes (e.g., reflection, evaluation of text) are necessary for deep elaboration of text meaning.

In web search, SERP information is usually presented in fragmented hyperlink abstracts, requiring readers to make selections based on sparse information. If web information is interpreted improperly, web users will come to incorrect conclusions in their evaluations, among other consequences. Since verbal SERP information is often just skimmed for keywords and phrases (Coiro & Dobler, 2007; Salmerón, Naumann, García & Fajardo, 2016), skills in retrieving meaning and comprehension are indispensable. For example, Rouet, Ros, Goumi, Macedo-Rouet and Dinet (2011) found that students matching for exact words selected more irrelevant link titles than students using semantic cues. Using surface cues like in word matching might spare cognitive resources, but it is often not an appropriate heuristic strategy for assessing the relevance of information.

1.2 Influences of the information basis

The way of individual information processing can be influenced by semantic and structural characteristics of information. Web users, for example, invested more time and cognitive resources in the evaluation of online information when heuristic expectations about hyperlinks concerning their order (Pan, Hembrooke, Joachims, Lorigo, Gay & Granka, 2007) or presentation format (Kammerer & Gerjets, 2014) were not met (Metzger et al., 2010). According to current models of web navigation (e.g., CoLiDeS+; Juvina & van Oostendorp, 2008), hyperlink selection is driven by the semantic similarity – or information scent – between presented information and a pursued search task (Blackmon, 2012). Information scent is assumed to be delivered by the hyperlink’s abstract or contextual cues like page arrangements. Empirically, web users selected nearly perfectly adequate hyperlinks from websites if information was semantically close to a search task (Blackmon, 2012). Since proficient readers are able to retrieve meaning accurately, quickly and with lower mental effort (Samuels & Flor, 1997; Walczyk, 2000), they might have an advantage over less skilled readers in identifying relevant information. This could be especially the case when the task-specific relevance of information is comparably easy to identify and web users are not in need to extensively allocate their cognitive resources. For example, when seeking for information about migraine to prepare a talk in biology class, the relevance of advertisements for new pharmaceuticals would be easier to distinguish as irrelevant material than news reports of medicals journals. Although both information sources share the same keywords ("migraine"), the hyperlink context informs the web user about the proximity to the search goal. In case that all hyperlinks are semantically close to a search task, web users will rather take other criteria into account to judge the usefulness of particular information sources (e.g., layout cues, prior content knowledge, or personal experiences to determine if information from Wikipedia is as good as from medical journals for completing a specific search tasks).

From a structural perspective, readers create actively their own information basis by deciding which information to read and which not. The selection of text (fragments) including page transitions in hypertext is often connoted with the concept of navigation (Lawless & Schrader, 2008). For digital text in general, students' navigation behavior was found to be an important predictor of their comprehension (e.g., Hahnel, Goldhammer, Naumann & Kröhne, 2016; Naumann, Richter, Christmann & Groeben, 2008; Salmerón, Cañas, Kintsch & Farjado, 2005). For web search in particular, though, students often look intensively at the first three search results on a SERP, but mainly ignore the rest (Pan et al., 2007), and visits over and above the first result page are often not even performed (Van Deursen & van Dijk, 2009). When using a search engine, students can navigate from a SERP to other ones or click on SERP hyperlinks to visit the websites connected. Navigation to other SERPs can be required when no suitable information is found on a previous SERP and a search query is not reformulated. Navigation to websites of SERP hyperlinks, in contrast, might be needed for further inspection of information confirming or rejecting its relevance. However, enriching the information basis of text generally increases the complexity in reading processes and draws upon limited cognitive resources (cf. DeStefano & LeFevre, 2007; Hamilton et al., 2013; Liu, Chin, Payne, Fu, Morrow & Stine-Morrow, 2016; Perfetti & Stafura, 2014). Therefore, skilled readers might have an advantage over poor readers when more information is encountered.

1.3 Study rationale

This study aims at investigating how individual differences in reading skills on word, sentence, and text level affect the selection of information from SERPs and whether these relations are influenced by semantic and structural characteristic of information. Web users might identify relevant web information by just scanning links for semantically relevant keywords, but also by comparing and weighting the content of several hyperlink entries against each other. Therefore, we expected that reading skills on word, sentence, and text level predict positively information selection from SERPs (H1). Web users are likely to select information that is semantically close to their search task. Compared to poor readers, though, proficient readers might make better hyperlink selections because they invest less cognitive resources in text processing. This might be especially true when relevant hyperlinks are easily to distinguish from non-relevant hyperlinks. Therefore, positive effects of reading skills on evaluating online information should be more pronounced, the more the SERP hyperlink abstracts vary in their relevance to a search task (H2). Web users' navigation to other pages determines the information basis for their selection of online information. Examining hyperlinks by checking other available alternatives (i.e., navigation to other SERPs) or verifying its fit to a search task (i.e., navigation to SERP websites) should positively predict students' evaluation of online information (H3). Since students encounter new written information through navigation, though, skilled readers should be at advantage over less skilled readers since they can process text information more efficiently. Therefore, we expected the relation of reading skills and evaluating online to increase when students navigate to other SERPs or SERP websites (H4).

Method

2.1 Sample

The analyzed data originated from a subsample of the *Programme for International Student Assessment* 2012 (PISA; OECD, 2013). Participating students were randomly sampled from a systematic sample of schools having 15-year-old students (details see OECD, 2014). For hypotheses testing, the data of 416 German students from 75 schools were considered (45.19% female, M_{age} = 15.84, SD_{age} = 0.29) who participated in PISA 2012 as well as in a national add-on study on computer based assessment (CBA).

2.2 Measures

2.2.1 Evaluating online information

The 16 items of the Test for the Evaluation of Online Information (TEO; Pfaff & Goldhammer, 2010) were used to measure individual skill in evaluating online information. The items simulate a search task together with a corresponding SERP (Figure 1). Search tasks vary from rather information-oriented (e.g., preparing a talk on autoimmune diseases for biology class) to problem-focused contents (e.g., seeking information about how to change a bicycle chain; see Appendix). The topics were chosen assuming students to have little or no prior knowledge about the content. Four TEO items displayed one SERP containing six hyperlinks; another four items provided an additional second SERP containing ten links in total. The remaining eight items consisted of a SERP of three or five hyperlinks leading to subsequent websites. The websites were static (e.g., hyperlinks, buttons, or menu entries were inactive). Task-specific and general instructions were presented on the left screen side.

Insert Figure 1 about here

2.2.1.1 Dependent variable. Students were requested to select the most informative and credible SERP hyperlink given a particular search task. They were asked to work as accurately and quickly as possible. Students' responses were scored by a pre-defined dichotomous scoring scheme (incorrect vs. correct). The dichotomous responses served as dependent variable to model the students' probability of successful item solution (see section 2.4 Data Analysis).

2.2.1.2 SERP characteristics. In order to determine to what extent SERP hyperlinks differ in their relevance to the search task, we built the indicator 'variability in relevance' as independent item variable. We asked 7 PhD students (85.71% female; M_{age} = 29.43, SD_{age} = 3.41) with a background in computer based assessment to rate each of the in total 96 hyperlinks. On a 4-point Likert scale, they were asked to rate how relevant a particular hyperlink is for a given search task (0 = 'not at all relevant' to 3 = 'absolutely relevant'). Kendall's coefficient of concordance for ordinal data revealed an acceptable interrater agreement (W_t = .69). The ratings were averaged across raters for each hyperlink and their variances were determined for each item. A high variability in an item points out that hyperlinks are highly distinguishable in relevance and heterogeneous in their semantic fit to a search task (*z*-standardized; *Min* = -1.91, *Max* = 1.86; Appendix).

2.2.1.3 Navigation behavior. Dichotomous indicators of navigation behavior reflected whether or not students performed a page transition at least once – either a transition to a second SERP or to SERP link websites. Therefore, they could only be derived for the four items containing two SERPs as well as the 8 items containing one SERP with websites (Appendix).

2.2.2 Reading skills on word and sentence level

To measure reading skills on word and sentence level, two subscales of the ProDi-L reading inventory were used (Richter et al., 2012). A lexical decision task assessed students' word recognition, asking whether a presented letter combination was a word or not (16 words vs. 16 non-words). The length of the stimulus material was about 3 to 10 letters and 1 to 3 syllables. Words were nouns varying in type, frequency, regularity, and amount of orthographic neighbors. Non-words were created to vary orthographically and phonologically (e.g., changing the onset, “bame” instead of “name”; cf. Balota, Cortese, Sergent-Marshall & Yap, 2004). Using a sentence verification task, semantic integration was assessed in students. They were asked if a presented sentence was either true or false (12 true vs. 12 false sentences; e.g., “Sugar is sweet”, “A cactus is a little furry animal”). Stimuli varied in their number of propositions and semantic abstractness. The sentence length varied between 16 to 61 characters and 1 to 3 propositions. Items of both tests were presented successively on a laptop screen. Students were asked to respond as accurately and quickly as possible. Their dichotomously coded responses (correct vs. incorrect) as well as their reaction times were collected.

The ProDi-L reading tests were originally developed for primary school children. To have skill indicators that comprise both students’ response accuracy and processing speed, drift rates were derived from diffusion models as skill indicators (Ratcliff, Gomez & McKoon, 2004; Schroeder, 2011). Diffusion models are based on the assumption that a decision results from an accumulation process of information over time. To decide between two response alternatives, information about a presented stimulus is collected until a defined decision criterion is reached. As a consequence, associated response behavior is shown (e.g., a student recognizes “hedgehog” as a word and responds with “word”). The efficiency of the information accumulation process is the drift rate. Individuals with higher drift rates show faster and a more accurate decision behavior than individuals with lower drift rates (Voss, Rothermund & Voss, 2004). Using the software *fast-dm* (Voss & Voss, 2007), the drift rates for students’ word recognition and semantic integration were estimated. High scores reflected that students were more accurate and faster in retrieving words from their mental lexicon and evaluating the semantic context of short statements, respectively.

2.2.3 Reading skills on text level

For the assessment of text level reading skills, two reading clusters from PISA 2009 were used to measure reading comprehension. PISA reading tasks are organized into units, that is, a text with three to five subsequent items. Four reading units are summarized into one cluster. The unit texts are designed to take several formats (e.g., continuous and non-continuous text) and types (e.g., description, narration, argumentation), and to cover several reading situations (e.g., personal, public, educational). The items request explicit and implicit information of the unit text, and can also require the student to reflect on a text. Item formats included multiple choice as well as open response formats. Text responses were coded according to standardized coding guidelines by trained and supervised coders of the Data Processing and Research Center (DPC) of the International Association for the Evaluation of Educational Achievement (IEA). Released items can be retrieved from the OECD website (http://www.oecd.org/pisa/38709396.pdf).

Note that for another study on equivalence between computer and paper based assessment, the reading clusters were administered on computer and paper in a randomized within- and between group balanced design. Item responses were examined for differences between these modes (Kröhne, Hahnel, Schiepe-Tiska & Goldhammer, 2013). Only items that did not show statistically significant differences in difficulty were regarded for further analyses. Therefore, responses to 18 out of 29 items were used in this study. WLE scores were derived from a partial credit item response model (Masters, 2010; scaling sample N = 880; estimated with *TAM*, Kiefer, Robitzsch & Wu, 2016) and served as ability estimates of students' reading comprehension. High scores reflected that students do well in comprehending texts. Reliability was acceptable (WLE reliability = .69, Cronbach's α = .83).

2.3 Procedure

The data originated from the German CBA add-on study that examined computer based assessment in the context of PISA (cf. Hahnel et al., 2016). This study took place a week after the PISA 2012 main study and was predominantly computer based. Trained test administrators tested about 14 students per school (data of 888 students was collected in total). All students were asked to complete one reading cluster within the first 30 minutes. After that, they either received (1) the TEO and the tests on word recognition and semantic integration, or (2) a second reading cluster and other tests. Only the former condition included the assessment of all variables investigated. Students were randomly assigned to all conditions. Comprehensive tutorials on the structure and functionalities of item surfaces and practice tasks were given for each test.

2.4 Data Analysis

Generalized linear mixed models (GLMM; De Boeck et al., 2011) were used for statistical analyses. These models assume that the probability of success can be expressed as a linear combination of fixed and random effects for predictor variables using a logit link function. Fixed effects do not differ for the observed units (e.g., students, items); random effects allow for variability across them. In this study, the logit of the probability *P* that a student correctly solves a TEO evaluation task was investigated, regarding the nested structure of items ($i = 1, \ldots, I$) in students ($j = 1, \ldots, N$) and schools ($k = 1, \ldots, K$). A baseline model was specified by

$$\text{logit}(P_{ijk}) = \beta_i + \theta_j + \theta_k \ , \quad (1)$$

where success in the TEO evaluation tasks is described as linear combination of the item easiness (fixed effect β_i), a student's skill in evaluating online information (random intercept θ_j) and a school's performance level (random intercept θ_k). This model was extended for hypothesis testing. First, the impact of the reading skills ($p = 1, \ldots, 3$) on the evaluation of online information was modeled as

$$\text{logit}(P_{ijk}) = \beta_i + \theta_j + \theta_k + \textstyle\sum_{p=1}^{P} \zeta_p Z_{(j,k)p} \ , \quad (2)$$

where ζ_p is the fixed effect of individual reading skills $Z_{(j,k)p}$ in word recognition ($p = 1$), semantic integration ($p = 2$), and reading comprehension ($p = 3$). Since reading skills are interrelated (e.g., Perfetti & Stafura, 2014), we examined models with Z_1, Z_2 and Z_3 separately as well as combined. Second, to investigate how the item variable 'variability in relevance' (X_i) affects the effects of reading on evaluating online information, an interaction term was added, resulting in

$$\text{logit}\left(P_{ijk}\right) = \beta_i + \theta_j + \theta_k + \sum_{p=1}^{P} \zeta_p Z_{(j,k)p} + \sum_{p=1}^{P} \zeta_p^* Z_{(j,k)p} X_i \,. \quad (3)$$

The coefficient ζ_p represents now the effects of reading in items with an average variability level in relevance, whereas ζ_p^* reflects how this effect changes when the variability in relevance increases. Note that a main effect of 'variability in relevance' is integrated in the item easiness parameters that are specified as fixed effects. Finally, the model was modified a last time to include the indicator of students' navigation behavior ($W_{(i,j,k)}$) to examine effects of navigation. Separate models for items containing two SERPs and items containing a SERP with websites were estimated since the interpretation of the indicators differ between types of navigation. The resulting model includes the effect of navigation on evaluating online information (ω) and how the effect of the reading skills changes when students navigate between pages (ζ_p^*):

$$\text{logit}\left(P_{ijk}\right) = \beta_i + \theta_j + \theta_k + \sum_{p=1}^{P} \zeta_p Z_{(j,k)p} + \omega W_{(i,j,k)} + \sum_{p=1}^{P} \zeta_p^* Z_{(j,k)p} W_{(i,j,k)} \,. \quad (4)$$

All analyses were carried out in *R 3.2.4* (R Core Team, 2016). For the estimation of GLMMs, the *R* package *lme4* was used (Bates, Mächler, Bolker & Walker, 2015; De Boeck et al., 2011). The hypotheses were tested one-sided at a significance level of 5%. Continuous variables were *z*-standardized. Regression coefficients can be interpreted as predicted changes in log odds for responding correctly if one predictor increases by one standard deviation.

Results

Estimating the baseline model showed that students vary in their skill to evaluate online information on an individual level ($Var\left(\theta_j\right) = 0.38$) as well as school level ($Var(\theta_k) =$

0.32). The estimated item easiness is presented in the Appendix for each item. Note that we refrain from reporting these parameters for all following models.

Concerning the impact of word, sentence, and text level reading skills on evaluating online information, the results of the models including the predictors separately and combined are presented in Table 1. They show that the separate predictors contributed significantly and positively to predict students' evaluation skill. Students selected rather relevant and credible hyperlinks when they were skilled in recognizing words, verifying semantic content of short statements, and comprehending texts. Although all reading skill components were positively correlated (word recognition - semantic integration: $r = .32$, $p < .001$; word recognition - reading comprehension: $r = .40$, $p < .001$; semantic integration - reading comprehension: $r = .38$, $p < .001$), both semantic integration and reading comprehension remained to be predictive in the combined model. The amount of variance explained was moderate in models including reading comprehension as predictor (R^2 in Table 1).

Insert Table 1 about here

To investigate how the similarity of SERP hyperlinks in relevance affect the association of reading skills and evaluating online information, models containing the item variable 'variability in relevance' were estimated. Table 2 shows the results. The main effects of all reading skills remained, given an average level of variability in relevance. The interaction effects show that only the effect of reading comprehension on evaluating online information rose with the an increasing variability of relevance in SERP hyperlinks. This was not true for the effects of reading skills on word and sentence level, neither in the separate models nor in the combined model.

Insert Table 2 about here

Concerning the influences of navigation, the effects of reading skills were examined for evaluation items containing two SERPs and, respectively, one SERP with websites. For the four items with two SERPs, a descriptive analysis of students' navigation behavior revealed that about half the students did not navigate at all in this condition (about 46% to 68% per item). Table 3 presents the results how SERP navigation behavior interacted with reading skills. Navigation had a positive effect indicating that students had better chances to solve the evaluation task correctly when they visited the second SERP. The models including reading skills separately showed main effects for all readings skills on evaluating online information when students did not navigate between two SERPs. When students navigate between the SERPs, the effects of reading on word and sentence level remained unchanged, but the effect of reading comprehension was increased. The combined model showed that only effects of navigation, reading comprehension and their interaction remained significant after taking all reading skills into account.

Insert Table 3 about here

In the eight items containing one SERP with websites, half the students showed no navigation behavior (about 46% to 67% per item). The results for models examining the impact of navigation on reading effects are presented in Table 4. Again, navigation was highly predictive in all models. The probability to solve an evaluation task rose when students navigated within the items. The models including just one reading skill showed only a main effect of word recognition when students did not visit any of the websites, but no interaction with students‘ navigation behavior was found. In contrast, semantic integration and reading comprehension changed to be predictive for the evaluation of online information when students visited websites. The combined model showed only a significant effect of word reading for students who did not leave the SERP as well as an effect of reading comprehension for students who visited websites connected to the hyperlinks on a SERP.

Insert Table 4 about here

Discussion

The present study investigated the impact of reading skills on word, sentence, and text level when students evaluate information from search engine result pages (SERPs). In line with hypothesis H1, we found that skills in word recognition, semantic integration, and reading comprehension predicted students' selection behavior of SERP hyperlinks, although only semantic integration and reading comprehension made a unique contribution. These analyses were deepened by taking characteristics of hyperlinks and students' navigation behavior into account. Supporting H2 partly, the effect of reading comprehension on hyperlink selection was more pronounced, the more the links differed in their relevance to a search task. Effects of word and sentence level reading were not affected by this task characteristic. Confirming H3, navigation to other SERPs and SERP websites was positively related to students' evaluating skill. Partially supporting H4, reading comprehension was the only skill to interact with students' navigation between SERPs. An interaction of navigation with semantic integration and reading comprehension was also found in tasks providing a SERP with websites. These reading skills explained evaluations of online information only when navigation was actually performed. Interestingly, when students did not visit SERP websites, their word recognition skills were predictive for their evaluation skill.

4.1 Reading processes in web search

The study's results show that several reading processes are involved when students evaluate online information to select hyperlinks from SERPs. Moreover, they give evidence that different reading skills support different strategies that can be applied to process web information (cf. Coiro & Dobler, 2007; Evans, 2008; Wirth et al., 2007). The overall effects of both semantic integration and reading comprehension suggest differences in the width and depth of information processing. Students who are able to capture semantic contexts of fragmented texts like in SERP hyperlinks will select appropriate information. However, some students might just process hyperlink information superficially until they have found one matching the search task to a large degree semantically, whereas others compare and relate every hyperlink from a limited collection in order to identify the most suitable alternative.

As previous studies revealed, the adequate use of structural and semantic cues like the layout of a website or discrepancies in contents will determine the quality of hyperlink evaluations (e.g., Metzger et al., 2010; Rouet et al., 2011). Especially prior content knowledge may play the biggest part for judging and comparing such cues (Lucassen et al., 2013), but the interaction of reading comprehension skills with the task-specific relevance of SERP hyperlinks points at an advantage of skilled readers. It might be easier to identify relevant semantic cues and to make actually use of them for skilled readers than less experienced ones (Rouet et al., 2011). In this respect, reading comprehension could be an important prerequisite for identification and also a supporting skill for setting semantic cues into an appropriate context. Skilled readers might also use semantic cues to identify if information is irrelevant. In that case, they could turn away from irrelevant hyperlinks easier than struggling readers saving limited cognitive resources (cf. Walczyk, 2000).

Support for these interpretations is also given by the results involving students' navigation behavior. Navigation to other SERPs or SERP websites generally supported students' evaluations. Through navigation, students could confirm or falsify their expectations about a link and exclude that SERP information misled them to hasty conclusions (cf. Coiro & Dobler, 2007). When visiting a second SERP, effects of word and sentence level reading skills did not increase, but readers with good comprehension skills were in advantage in selecting useful hyperlinks. Although the direction of this positive effect remains unclear, it seems plausible that reading comprehension and navigation behavior are mutually dependent. On the one hand side, students might navigate because they could not find information that fits a search task semantically (cf. Blackmon, 2012) or is sufficiently relevant and credible (cf. Rieh, 2002). That would mean that navigation to another SERP would be only performed when students elaborated on and compared the hyperlinks of a particular SERP explaining why the effects of word recognition and semantic integration remained unchanged. On the other side, good readers might be able to process the amount of information that was added through navigation more efficiently than less skilled readers. Since constituent skills of reading are more fluent in skilled readers and can be processed largely automatically, they might enable readers to process text information more efficiently requiring less mental effort for the integration of information from multiple sources and decision making (cf. Hamilton et al., 2013; Rouet, 2006; Walczyk, 2000).

Fluency in reading and processing information also explains the increased effect of semantic integration and reading comprehension when students visited websites connected to SERPs. Presumably, experienced readers are in a better position to assess the need for additional information from SERP websites than less skilled readers. The missing effects of semantic integration and reading comprehension for students who did not visit connected websites may look odd at first, but together with the finding of a word recognition effect in the same condition they underline again that students apply different strategies to evaluate online information and that specific reading skills can be supportive depending on the strategy chosen. Although the results cannot uncover specific strategies of information processing, they complement studies that found students to use both heuristic and systematic processing strategies in order to select hyperlinks (e.g., Salmerón et al., 2015). The results point at the importance to regard what students actually did while task processing in terms of navigation (Salmerón et al., 2005). Evaluation strategies and their conditions, though, need to be investigated with more fine-grained data, for example, as from think-aloud studies that can provide valuable insights into reasons for specific navigation behaviors and strategies of information processing (e.g., Gerjets et al., 2011).

4.2 Limitations

There are at least three limitations that should be considered in this study. First, the Test for the Evaluation of Online Information (TEO), which was used to measure students' skills in evaluating online information, simulates a search engine environment but it also restricted to predefined functions of that closed environment. That is, test authors have decided on the quality and volume of information that can be accessed and have to be taken into account for evaluations. Therefore, the TEO cannot represent the variety of actions web users usually apply (e.g., free choice of search terms), the amount of available online information (e.g., unspecified information quantities), or the strategic use of web search features (e.g., making use of knowledge about search engine algorithms). As a consequence, TEO results cannot be generalized to give a comprehensive picture of behaviors when seeking information in an open web space. However, it is the major advantage of the TEO that the process of information seeking is standardized. Reproducible search results are normally not given in open search. Since both web users' actions and available information space are limited, the items spotlight micro-processes in web search and allow for in-depth analyses. To make selection criteria of students transparent (e.g., relevance, trustworthiness, credibility, prior knowledge), a further development of the TEO would be useful. Students could be asked for the reason why they have selected a particular hyperlink after task completion. Alternatively, they could also rate the provided hyperlink according to several quality aspects, comparable to the conducted relevance ratings in this study. Such an extension, however, should be considered carefully since students might align their response behavior according to given criteria.

Second, one could argue that many students did not follow the instructions because they did not navigate in tasks allowing for visiting other pages than the initial SERP. Indeed, students were not directly instructed to look at all available information but they were told that it is available. Additionally, a reminder was permanently visible (grey panel on the left side of Figure 1). Presumably, students did not ignore the instruction but acted naturally by not visiting all available information according to their experiences with web search and search engines. In that case, the TEO items would have worked well in providing face-valid conditions of web search.

Finally, reading skills are not the only important constituents for evaluating online information. In addition to prior knowledge about contents and structural features which has been found to affect the evaluation of surface and semantic cues fundamentally (e.g., Lucassen et al., 2013; Rouet et al., 2011), web users' memory skills, individual perceived cognitive load, or learnt strategies to overcome memory deficits (e.g., bookmarking, using tabs) can have a strong impact on the allocation of cognitive resources affecting how students compare and weight web information (cf. DeStefano & LeFevre, 2007). Correlation based analyses like we did can give first insights into relationships between individual cognitive skills, features of texts provided in web search and indicators of actual behavior of web users. Nevertheless, other empirical approaches like, for example, think-aloud studies or experimental designs stressing cognitive resources during web search tasks are needed to cross-validate and strengthen our interpretations of the present study's results.

4.3 Conclusions

The study shows that partial processes of reading (i.e., word recognition, semantic integration and reading comprehension) are essential when seeking information with the aid search engines. We interpreted the results in that reading skills on word, sentence, and text level support different information processing strategies that students can choose when evaluating online information from SERPs. Especially proficient text level skills seem to support readers in the identification and the use of semantic cues in SERP hyperlinks reflecting their relevance to a search task. Furthermore, students' individual reading skills interact with their behavior in visiting other SERPs or websites that are connected to SERP hyperlinks. This indicates that skilled readers in comparison to poor readers are able to allocate their cognitive resources more effectively when dealing with online information making navigation as an event of receiving more information less effortful. Uncovering in-depth mechanisms in this interplay of reading skills, navigation behavior and evaluation of online information will be a challenge for future research. In educational means, the results show in general that students' skill to sensibly select online information while doing a web search rests on component skills that can be learned and trained. Poor readers, though, will not get a fresh start in the information age (cf. Walraven et al., 2008). The deficits in their reading skills can impair their use of the 'new' media if not alleviated by adequate education and intervention.

References

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual Word Recognition of Single-Syllable Words. *Journal of Experimental Psychology: General*, *133*(2), 283–316.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blackmon, M. H. (2012). Information scent determines attention allocation and link selection among multiple information patches on a webpage. *Behaviour & Information Technology*, *31*(1), 3–15. https://doi.org/10.1080/0144929X.2011.599041

Brand-Gruwel, S., & Stadtler, M. (2011). Solving information-based problems: Evaluating sources and information. *Learning and Instruction*, *21*(2), 175–179. https://doi.org/10.1016/j.learninstruc.2010.02.008

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, *53*(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, *42*(2), 214–257. https://doi.org/10.1598/RRQ.42.2.2

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*(12), 1–28.

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616–1641. https://doi.org/10.1016/j.chb.2005.08.012

Evans, J. S. B. T. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

Feierabend, S., Karg, U., & Rathgeb, T. (2013). *JIM-Studie 2013. Jugend, Information, (Multi-) Media [JIM-Study. Youth, Information, (Multi-) Media]*. Stuttgart: MPFS.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, *9*(2), 319–342. https://doi.org/10.1177/1461444807075015

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*(2), 220–231. https://doi.org/10.1016/j.learninstruc.2010.02.005

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, *62*(1), 451–482. https://doi.org/10.1146/annurev-psych-120709-145346

Hahnel, C., Goldhammer, F., Naumann, J., & Kröhne, U. (2016). Effects of linear reading, basic computer skills, evaluating online information, and navigation on reading digital text. *Computers in Human Behavior*, *55*, 486–500. https://doi.org/10.1016/j.chb.2015.09.042

Hamilton, S. T., Freed, E. M., & Long, D. L. (2013). Modeling Reader and Text Interactions During Narrative Comprehension: A Test of the Lexical Quality Hypothesis. *Discourse Processes*, *50*(2), 139–163. https://doi.org/10.1080/0163853X.2012.742001

Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, *44*(4), 1467–1484. https://doi.org/10.1016/j.ipm.2007.10.001

Juvina, I., & van Oostendorp, H. (2008). Modeling Semantic and Structural Knowledge in Web Navigation. *Discourse Processes*, *45*(4–5), 346–364. https://doi.org/10.1080/01638530802145205

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*(3), 177–191. https://doi.org/10.1080/10447318.2013.846790

Kiefer, T., Robitzsch, A., & Wu, M. (2015). *TAM: Test Analysis Modules*. Retrieved from http://CRAN.R-project.org/package=TAM

Kintsch, W. (1998). *Comprehension: A Paradigm for Cognition*. Cambridge University Press.

Kröhne, U., Hahnel, C., Schiepe-Tiska, A., & Goldhammer, F. (2013, August). *Analyzing Mode Effects of PISA Print Reading Including a Comparison of Time-related Information.* Presented at the 15th Biennal Conference of the European Association for Research on Learning and Instruction (EARLI), Munich, Germany.

Lawless, K. A., & Schrader, P. G. (2008). Where do we go now? Understanding research on navigation in complex digital environments. In D. J. Leu & J. Coiro (Eds.), *Handbook of New Literacies* (pp. 267–296). Hillsdale, NJ: Lawrence Erlbaum Associates.

Liu, X., Chin, J., Payne, B. R., Fu, W.-T., Morrow, D. G., & Stine-Morrow, E. A. L. (2016). Adult age differences in information foraging in an interactive reading environment. *Psychology and Aging*, *31*(3), 211–223. https://doi.org/10.1037/pag0000079

Lucassen, T., Muilwijk, R., Noordzij, M. L., & Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology*, *64*(2), 254–264. https://doi.org/10.1002/asi.22743

Masters, G. N. (2010). The Partial Credit Model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 109–122). New York, NY: Routledge.

Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, *58*(13), 2078–2091. https://doi.org/10.1002/asi.20672

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and Heuristic Approaches to Credibility Evaluation Online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Naumann, J., Richter, T., Christmann, U., & Groeben, N. (2008). Working memory capacity and reading skill moderate the effectiveness of strategy training in learning from hypertext. *Learning and Individual Differences*, *18*(2), 197–213. https://doi.org/10.1016/j.lindif.2007.08.007

Organisation for Economic Co-operation and Development (OECD). (2011). *PISA 2009 Results: Students On Line*. OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/pisa-2009-results-students-on-line_9789264112995-en

Organisation for Economic Co-operation and Development (OECD). (2013). *PISA 2012 Assessment and Analytical Framework*. OECD Publishing. Retrieved from http://dx.doi.org/10.1787/9789264190511-en

Organisation for Economic Co-operation and Development (OECD). (2014). *PISA 2012 Technical Report*. OECD Publishing. Retrieved from http://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf

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*(3), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Pfaff, Y., & Goldhammer, F. (2012). *Measuring individual differences in ICT literacy: Accessing and Evaluating Information*. Paper presented at the 14th EARLI Conference, Exeter, UK.

R Core Team. (2015). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182. https://doi.org/10.1037/0033-295X.111.1.159

Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeiten bei Grundschulkindern. *Zeitschrift Für Pädagogische Psychologie*, *26*(4), 313–331. https://doi.org/10.1024/1010-0652/a000079

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, *53*(2), 145–161. https://doi.org/10.1002/asi.10017

Rieh, S. Y., Kim, Y.-M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing & Management*, *48*(6), 1136–1150. https://doi.org/10.1016/j.ipm.2012.05.001

Rouet, J.-F. (2006). *The skills of document use: from text comprehension to Web-based learning*. Mahwah, NJ: Erlbaum.

Rouet, J.-F., 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*(2), 205–219. https://doi.org/10.1016/j.learninstruc.2010.02.007

Sabatini, J. (2015). *Understanding the Basic Reading Skills of U.S. Adults: Reading Components in the PIAAC Literacy Survey*. Princeton, N.J.: Educational Testing Service.

Salmerón, L., Cañas, J. J., Kintsch, W., & Fajardo, I. (2005). Reading Strategies and Hypertext Comprehension. *Discourse Processes*, *40*(3), 171–191. https://doi.org/10.1207/s15326950dp4003_1

Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior*, *29*(6), 2161–2171. https://doi.org/10.1016/j.chb.2013.04.034

Salmerón, L., Naumann, J., García, V., & Fajardo, I. (2016). Scanning and deep processing of information in hypertext: an eye tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning*. https://doi.org/10.1111/jcal.12152

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly*, *13*, 107–121.

Schroeder, S. (2011). What Readers Have and Do: Effects of Students' Verbal Ability and Reading *Journal of Educational Psychology*, *103*(4), 877–896. https://doi.org/10.1037/a0023731

van Deursen, A. J. A. M., & van Dijk, J. A. G. M. (2009). Using the Internet: Skill related problems in users' online behavior. *Interacting with Computers*, *21*(5–6), 393–402. https://doi.org/10.1016/j.intcom.2009.06.005

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206–1220.

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775.

Walczyk, J. J. (2000). The interplay between automatic and control processes in reading. *Reading Research Quarterly*, *35*(4), 554–566.

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*(3), 623–648. https://doi.org/10.1016/j.chb.2007.01.030

Wirth, W., Böcking, T., Karnowski, V., & von Pape, T. (2007). Heuristic and Systematic Use of Search Engines. *Journal of Computer-Mediated Communication*, *12*(3), 778–800. https://doi.org/10.1111/j.1083-6101.2007.00350.x

Appendix

Overview of content, characteristics, and easiness of the TEO items

Item	Search task	Number of hyperlinks	Variability in relevance	Type	Item Easiness (*SE*)
item 1	How to stop smoking	6	0.76	SERP only	-0.57 (0.13)
item 2	Prepare a presentation on autoimmune diseases	6	0.08	SERP only	0.24 (0.13)
item 3	How to replace a light bulb	6	1.86	SERP only	-1.78 (0.16)
item 4	Prepare a presentation on migraine	6	0.17	SERP only	-0.45(0.14)
item 5	How to clean a drain	10	-0.37	two SERPs	-1.44 (0.15)
item 6	Learning about climbing sports	10	-0.39	two SERPs	0.14 (0.13)
item 7	Learning about sailing sports	10	-0.34	two SERPs	-1.03 (0.14)
item 8	Prepare a presentation on alternative energy sources	10	-0.45	two SERPs	-1.14 (0.14)
item 9	Learning about diving sports	5	0.57	SERP with websites	-1.31 (0.14)
item 10	Learning about lunar eclipse	3	-1.91	SERP with websites	0.30 (0.13)
item 11	How to treat a common cold	5	-1.17	SERP with websites	-2.42 (0.18)
item 12	How to change a bicycle chain	3	-0.85	SERP with websites	-0.66 (0.14)
item 13	Prepare a presentation on computer viruses	5	1.05	SERP with websites	0.54 (0.14)
item 14	How to treat a sunstroke	3	1.20	SERP with websites	-0.10 (0.13)
item 15	Learning about algae	5	0.81	SERP with websites	-2.59 (0.19)
item 16	How to treat acne	3	-1.02	SERP with websites	-0.91 (0.14)

Note. Values of ‘variability of variance’ are *z*-standardized. Item easiness was estimated from the baseline model.

Table 1

Effects of reading skills on evaluating online information

	WR				SI				RC				Combined model			
Fixed effects	**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***	
word recognition (WR)	.19	.05	<.001	***									.07	.04	.086	
semantic integration (SI)					.21	.05	<.001	***					.10	.05	.028	*
reading comprehension (RC)									.47	.05	<.001	***	.42	.05	<.001	***
ΔR^2	1.8%				0.9%				19.6%				19.3%			

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 2

Effects of reading skills on evaluating online information in interaction with the variability in relevance (VIR) within items

	WR				**SI**				**RC**				**Combined model**			
Fixed effects	**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***	
word recognition (WR)	.18	.05	<.001	***									.07	.05	.086	
WR : VIR	.03	.03	.141										.00	.03	.499	
semantic integration (SI)					.21	.05	<.001	***					.10	.05	.027	*
SI : VIR					.04	.03	.093						.01	.03	.392	
reading comprehension (RC)									.47	.05	<.001	***	.42	.05	<.001	***
RC : VIR									.10	.03	.001	**	.10	.04	.002	*

Notes. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 3

Effects of reading skills on evaluating online information depending on students' navigation behavior in items with two SERPs

	WR				SI				RC				Combined model			
Fixed effects	**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***	
navigated between SERPs	1.29	.16	<.001	***	1.29	0.16	<.001	***	1.13	0.16	<.001	***	1.08	.16	<.001	***
word recognition (WR)	.25	.12	.015	*									.16	.12	.088	
navigated between SERPs : WR	.09	.16	.276										-.05	.17	.384	
semantic integration (SI)					.25	.12	.015	*					.12	.12	.150	
navigated between SERPs : SI					.25	.17	.065						.18	.17	.149	
reading comprehension (RC)									.52	.13	<.001	***	.45	.14	.001	**
navigated between SERPs : RC									.34	.17	.021	*	.31	.18	.039	*

Notes. Reference category for the navigation variable is "no navigation". * $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4

Effects of reading skills on evaluating online information depending on students' navigation behavior in items containing one SERP with websites

	WR				**SI**				**RC**				**Combined model**			
Fixed effects	**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***p***		**Est.**	***SE***	***P***	
navigated to websites	1.05	.09	<.001	***	.99	.10	<.001	***	.87	.10	<.001	***	.85	.10	<.001	***
word recognition (WR)	.14	.07	.020	*									.14	.07	.026	*
navigated to websites : WR	-.03	.09	.371										-.12	.10	.120	
semantic integration (SI)					.01	.06	.412						-.03	.06	.344	
navigated to websites : SI					.20	.10	.022	*					.14	.10	.086	
reading comprehension (RC)									.07	.06	.120		.02	.07	.381	
navigated to websites : RC									.36	.10	<.001	***	.38	.11	<.001	***

Notes. Reference category for the navigation variable is "no navigation". * $p < .05$; ** $p < .01$; *** $p < .001$.

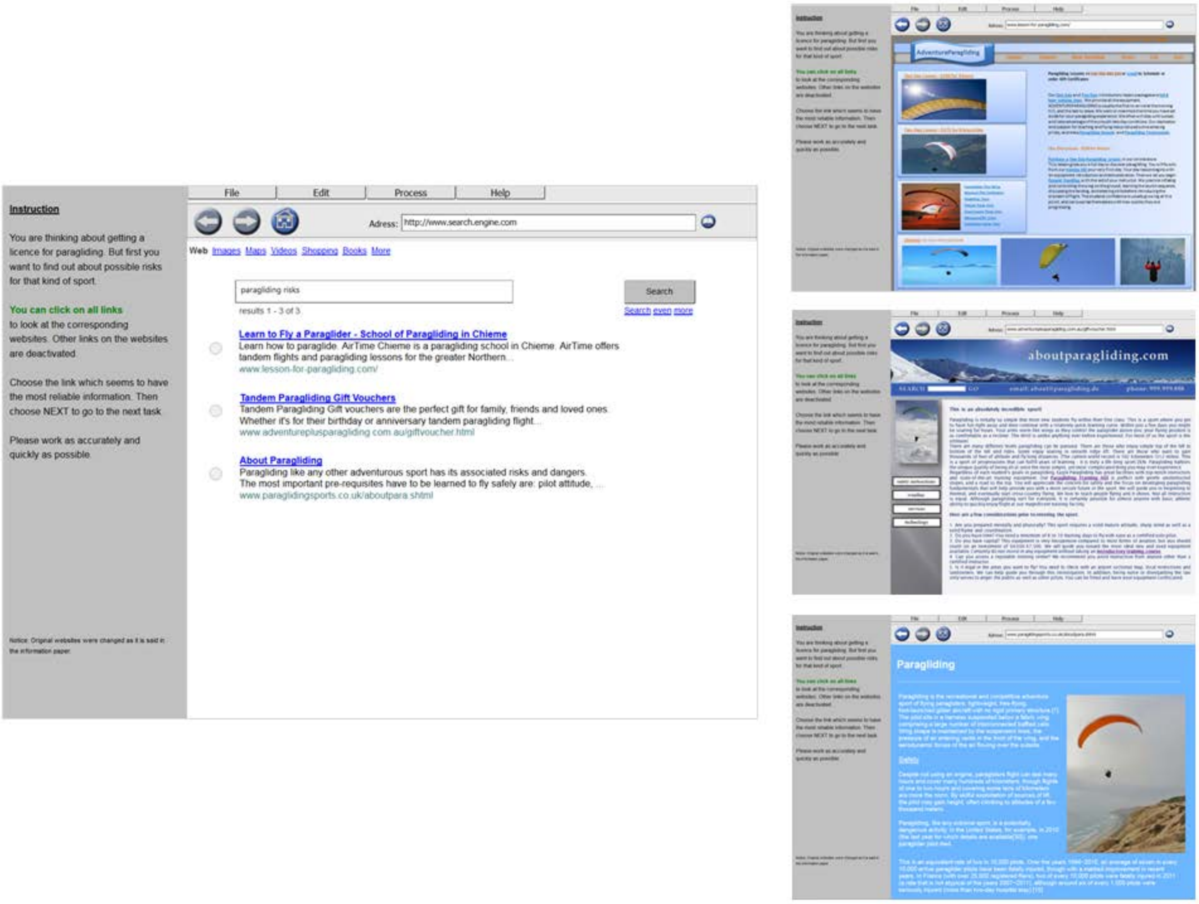

Figure 1. Screenshot of a TEO item example with active hyperlinks. The first page is a search engine result page (SERP) for the topic “paragliding” with three search results (left). If one of the three hyperlinks is clicked, students will be lead to the corresponding connected website (right). Backtracking to the SERP was possible.