**Author(s):** Kiili, Carita; Laurinen, Leena; Marttunen, Miika

**Title:** Students evaluating Internet sources – From versatile evaluators to uncritical readers

**Year:** 2008

**Version:**

**Students evaluating Internet sources:**

**From versatile evaluators to uncritical readers**

Manuscript for the Journal of Educational Computing Research

Vol. 39(1-2) 75-95, 2008-2009

Corresponding author:

Carita Kiili

Department of Educational Sciences, University of Jyväskylä, Finland
P.O.Box 35, FI- 40014 University of Jyväskylä, Finland
carita.kiili@edu.jyu.fi


Leena Laurinen

Department of Educational Sciences, University of Jyväskylä, Finland
P.O.Box 35, FI- 40014 University of Jyväskylä, Finland
leena.laurinen@edu.jyu.fi


Miika Marttunen

Department of Educational Sciences, University of Jyväskylä, Finland
P.O.Box 35, FI- 40014 University of Jyväskylä, Finland
miika.marttunen@edu.jyu.fi

Abstract

The Internet is a significant information resource for students due to the ease of access it allows to a vast amount of information. As the quality of the information on the Internet varies, it is important that students are able to evaluate such information critically. The aim of the study was to investigate how students evaluate Internet sources in an authentic learning task. Upper secondary school students ($n = 25$) were asked to look for source material on the Internet in order to write an essay. They were asked to verbalize their thoughts during the material gathering process. Their verbalizations and actions on the Internet were recorded and analyzed. The five evaluation profiles emerged: 1) versatile evaluators 2) relevance-orientated evaluators 3) limited evaluators 4) disorientated readers 5) uncritical readers.

Keywords: evaluation, relevance, credibility, critical reading, Internet, evaluation profiles

Introduction

The Internet is widely used for educational purposes and it is an important information resource for students. Instead of libraries and textbooks, students rely more and more on the Internet in their school work (Jones & Madden, 2002; Lenhart, Simon & Graziano, 2001; Metzger, Flanagin & Zwarun, 2003). Because anyone can publish anything on the Internet, the information obtained from this source can be inaccurate or even misleading. As in most cases teachers are unable to control the quality of Internet-derived information used by students, it is important to know how well students are able to evaluate that information.

Many researchers (Liu & Huang, 2005; Metzger et al., 2003; Whitmire, 2004) have investigated university or college students' evaluation of information on the Internet, but fewer studies (Brem, Russell, & Weems, 2001; Lorenzen, 2001) concern the evaluations of adolescent learners. Previous research has indicated that college students have shortcomings in evaluating Internet sources (Grimes & Boening, 2001; Metzger et al., 2003). Therefore, it is important to investigate the evaluation strategies used by younger students to be able to better scaffold them and prepare them for their future studies.

In the present study evaluation of information on the Internet is observed in an authentic learning context. The results provide information about the evaluation strategies used by upper secondary school students when searching the Internet for source material for an essay. Due to the cumbersomeness of the procedure, the number of participants in earlier think-aloud studies has ranged from 10 to 15 (Coiro & Dobler, 2007; Eveland & Dunwoody, 2000; Whitmire, 2004). More students are, however, needed in order properly to identify evaluation profiles among students. The purpose of this study is to meet this challenge.

*Literacy Skills and Critical Thinking*

Researchers agree that, with the rapid development of information and communication technologies, more diversified literacy skills are needed (Coiro, 2003; Leu, Kinzer, Coiro, & Cammack, 2004). Various new concepts, such as information literacy (Bruce, 2002; Doyle, 1994), digital literacy (Gilster, 1997), network literacy (McClure, 1994) and media literacy (Aufderheide & Firestone, 1993), have been used to describe these skills. Bawden (2001), in reviewing the new literacy concepts, noted that the concept of information literacy is the most widely used. Doyle (1994, p. 1) defines information literacy as the ability to access, evaluate and use information drawn from a variety of sources. The information-literate student accesses information efficiently, evaluates information critically and uses information accurately and creatively (Information literacy standards for student learning, 1998).

Because the Internet is full of partial and sometimes even misleading information, the most important skill when making use of it is critical thinking (Gilster 1997, p. 87). Ennis (1987, p. 10) crystallizes the concept of critical thinking as reasonable reflective thinking that is focused on what to believe or do. According to Facione (2006), critical thinking skills consist of inference, evaluation, interpretation, analysis, explanation, and self-regulation. In addition to critical thinking skills, a person can have a critical disposition (Ennis, 1987; Facione, 2006). A person with a critical disposition is truth seeking, confident in reasoning, judicious, inquisitive, systematic, analytical, and open-minded (Facione, 2006). Although finding credible and relevant information on the Internet is sometimes challenging, the Internet also creates new opportunities to access alternative viewpoints and to compare argumentation drawn from different

sources.

The study by Connor-Greene and Greene (2002) indicates that students need to be taught to evaluate texts critically. In their study 39 university students read an Internet article about the dangers of aspartame and wrote their reactions to that article. More than half of the students accepted the information in the article unquestioningly, a quarter expressed ambivalence, and a fifth of the students did not believe the article. After a 30-minute structured small-group discussion about the article all the groups listed 5–11 problems that appeared in the article. The students mentioned problems such as no research evidence to support allegations, no citation of sources or expert cited in the article, a sensational writing style, and sweeping statements.

*Evaluation of Information*

Critical thinking skills are important when students evaluate information on the Internet. Judd, Farrow and Tims (2006) used credibility and relevance as factors in their instrument designed to teach undergraduates to evaluate the information value of web pages. The dimensions of credibility in their model are expertise (knowledgeable source and factual information) and trustworthiness (unbiased). The dimensions of information relevance are importance and currency.

According to Harris (1997), evaluation of credibility involves evaluation of the authority, quality, style, and accuracy of the text, evaluation of the purpose of the text, and verification of information. In the information science literature the relevance of information has been defined in various ways (Barry & Schamber, 1998; Borlund, 2003; Saracevic, 1996). Relevance can be system-orientated, i.e. the connection between a query and a document, or user-orientated, dealing with

the user's subjective decision-making. Saracevic (1996, pp. 212–213) divides relevance into five types of which three fall into the user-orientated category. *Cognitive relevance* refers to how well a document satisfies the user's information need. It includes perceived informativeness, novelty, and information quality. *Situational relevance* is associated with a task and user's interpretation of it. Finally, *motivational relevance* concerns the intents and goals of the user as well as the satisfaction experienced by him or her.

In the present study credibility refers to whether a student is paying attention to distinguishing reliable from unreliable information. Relevance, in turn, refers to whether a student is paying attention to distinguishing essential from non-essential information.

*Evaluation Studies*

Information evaluation on the Internet has been studied, for example, by surveys (Metzger et al., 2003; Wu & Tsai, 2005). Metzger et al. (2003) found that college students relied extensively on the Internet as a source of information in their academic work. Although students did not consider Web-based information to be as credible as traditional sources of information, they reported verifying information found on the Internet only rarely or occasionally. Surveys, however, do not reveal what kind of evaluation strategies students actually use when they are working on a school assignment and how those strategies are reflected in the quality of material studied.

When Grimes and Boening (2001) interviewed 11 college students about how they evaluated Web sources cited to in their research papers, they found that students evaluated Web sources only superficially, if at all. Owing to inadequate

evaluation skills or unwillingness to evaluate material, students used unauthenticated Web resources in their research papers. Whitmire (2004), in turn, studied the relationship of college students' (n = 15) epistemological beliefs (by questionnaire), reflective judgments (interview) and information search (think-aloud) on the Internet. The study indicated that the more advanced epistemological believers were better able to evaluate information sources and authority by using, for example, URL addresses and information of the institutional affiliation of the author. Brem, Russell, and Weems (2001) studied how 81 girls from the 9th, 11th and 12th grades evaluated scientific argumentation of websites of three kinds: hoaxes, weaker sincere sites and stronger sincere sites. They found that evaluation could, in part, be a situated activity. Students used more absolutist reasoning when evaluating weaker sincere sites than stronger sincere sites. By contrast, stronger sincere sites were accepted more easily on the basis of surface markers whereas weaker sincere sites received more attention. At hoax sites students used as little effort as possible to distinguish real from fake.

*Role of Evaluation in the Search Process*

According to Rieh (2000; 2002), people's judgments of information quality and cognitive authority on the Internet are either predictive or evaluative. Predictive judgments are based on a person's expectations of a web page prior to opening the page. Predictive judgment directs the decision-making and it can be based on knowledge, experience, recommendation, or on the information given in the search results. After opening a new web page an evaluative judgment is made on the basis of the nature of the information, its source, and the searcher's prior knowledge. In their interview study with seven students, MaKinster, Beghetto and

Plucker (2002, p. 167) described the usefulness of the predictive strategy in an information search task. They found a student who carefully evaluated the content of search results and URL addresses and was able effectively to recognize what web pages were useful.

Additionally, evaluation directs the information search process. Sutcliffe and Ennis (1998) presented a cyclic model of information search on the Internet. The cycle starts with problem identification, followed by needs articulation, query formulation, query reformulation and evaluation. Furthermore, the model stresses the user's previous domain and device knowledge on information search activities.

According to Wu and Tsai (2005), Web users' evaluative standards of accuracy and usefulness of information affect their information search strategies in Web-based learning environments. The researchers conducted a survey (610 university students) and analyzed the data with structural equation modeling. Learners who evaluated the accuracy of information by using multiple sources and concentrated on the relevance of the content used elaboration as a search strategy. In the e*laboration-strategy* a leaner tries to integrate information from multiple websites to meet his or her information need. In contrast, if a learner has adopted a *match strategy* for searching for information, she or he is trying to find a few relevant websites only. In the match strategy the learner evaluates the accuracy of information by reference to the authority of the website and evaluates the usefulness of information according to technical features (such as easy retrieval).

*Research Problems*

The research problems addressed by this study are as follows:

a) How do upper secondary school students allocate their working time among different kinds of actions on the Internet, such as browsing search results, reading texts, and navigating links?

b) How do students evaluate the credibility and relevance of information on the Internet when searching for source material for an essay?

c) What kinds of texts do students choose to read?

Method

*Participants*

Students, 25 in total (14 female and 11 male) from an upper secondary school in Finland volunteered to participate in the study. The participants were all born during the same calendar year and they were either 16 or 17 years. A signed consent was sought and received from the parents of the students.

*Task*

The study was integrated into process writing practice in the mother tongue (Finnish) class. Students were asked to look for source material on the Internet for 40 minutes in order to write an essay on a set topic. The topic of the essay was: *Sleeping as a human resource*. Sleeping is an issue of daily life, and thus all the students had at least some prior knowledge of the topic. Students were asked to verbalize their thoughts when gathering their source material. They then had 45 minutes to write a first draft of their essay, producing their further drafts at a later date.

When commencing their search for source material on the Internet, the students were allowed to choose one of the three alternative browsers: Internet Explorer, Mozilla Firefox, and Opera. The students were informed that they were free to use all the features of the browser. In each browser the starting page was empty and the students had to decide how to start the search task.

The students were allowed to make notes during the search, but not print the pages. They were, however, asked to choose two documents, which were then printed for them so that they could use them as source material in the second phase of the writing process. The researcher reminded the students choosing the two documents approximately ten minutes before the search time elapsed.

*Data Collection*

The study was conducted in the spring term of 2006. The researcher met each student individually. Each session lasted approximately 1 hour 45 minutes. The session started with a brief questionnaire. Students were asked for background information (age, sex) and information about their use of the Internet (number of hours weekly; familiarity with search engines; self-assessment of their information searching skills). After the students had completed the questionnaire the researcher described the task. The instructions were also given to the students in writing. After reading the instructions, students were asked to confirm whether they had understood the task. They were allowed to ask questions about the process.

The students were instructed to think aloud and report whatever they were thinking while searching for and reading information on the Internet. The students were informed that prompt questions (what is in your mind?; what are you

thinking?) would be asked if they remained silent for a longer period. The researcher did not model the thinking aloud, as the aim was to study students' spontaneous thoughts. If students asked what was meant by thinking aloud, they were answered. The sorts of questions asked were "Do I say, that now I'm thinking that I'm going to use Google to find information?" or "Do you mean speaking to myself?" Ericsson and Simon (1993) argue that thinking aloud resembles a natural process common in daily life. After the search task the students were asked how thinking aloud affected their performance.

Web actions as well as students' 'think-aloud' were recorded using Easy Video Capture software. Each session was replayed, transcribed, analyzed, and coded.

*Data analysis*

*Think-aloud protocols*

The 'think-alouds' were transcribed and evaluation strategies were separated from the protocols for further analysis. Evaluation strategies were divided into two main categories: evaluation of credibility (see Harris, 1997) and relevance (see Saracevic, 1996). The subcategories of credibility and relevance are presented in Table 1. The examples in Table 1 are taken from the present data. Silent rejections of the web page were not included in the analysis. The students (n = 24) rejected a page silently 110 times ($M = 4.58$, $SD = 3.80$). There was one exception among the students, a girl, who silently rejected a page 49 times. The reliability of analysis of think-aloud protocols was examined by having two persons to classify 16% of evaluative comments (the think-aloud protocols of three students) and we found 85.9% agreement in our analyses.

[Insert Table 1 here]

*Students' Internet actions*

Students' Internet actions were divided into seven categories: 1) formulating a query or using a URL address, 2) browsing search results, 3) downloading links and back-function, 4) reading texts, 5) skimming and rejecting the web page 6) navigating links, and 7) other. The time spent on each of these different actions was measured.

*Texts read by the students*

To begin with, we went through all the texts the students chose to read and formed a category for texts where the credibility of the publisher was somewhat questionable. Texts were classified into this category for the following reasons: information on the authority was not available on the website; the text was written by an anonymous person, novice or peer group; or the website was published by an ideological association or organization. The rest of the texts read by the students we divided into six categories according to the publisher of the website in which the text was included (public association, expert, expert organization or network of experts, Wikipedia, company or trade organization, media, educational institute). The time a student spent reading texts in each category was measured.

*Statistical analysis*

We conducted a hierarchical cluster analysis (Aldenderfer & Blashfield, 1984) to identify subgroups of students. The analysis was based on five variables: the number of credibility evaluations, the number of relevance evaluations, the sum of

different evaluation subcategories each student used, time spent on reading texts, and the proportion of the total text-reading time spent on reading texts with a questionable publisher credibility. Standardized scores for the five variables were used in the hierarchical cluster analysis employing the between-groups linkage method to identify an appropriate number of clusters. The analysis was based upon the increasing value of the Squared Euclidean Distance between clusters.

## Results

*Time Spent on Different Actions on the Internet*

Table 2 presents the mean percentages of the time the students allocated to different actions during the task. Most of the time (61.4%) was spent on reading texts. However, there were considerable inter-individual differences. The five most effective web-searchers spent more than 75% of their working time on reading texts, whereas the four weakest students spent so long searching for information that the time left to them for reading texts ranged from 30% to 45%.

[Insert Table 2 here]

The individual variation can be variously explained. For example, some students had difficulties in formulating effective queries and some in selecting relevant web pages from their search results. The latter problem can be considered as inability to predict the relevance of web pages, which in turn has an effect on the number of web pages first selected and then rejected as irrelevant.

*Evaluation Strategies*

As Table 3 shows, the evaluation of relevance ($M = 16.9$) was more frequent than the evaluation of credibility ($M = 3.2$). There was relatively wide variation among the students in the use of evaluation strategies. The evaluation of credibility ranged from 0–16 comments and the evaluation of relevance ranged from 6–43 comments. For example, six students did not evaluate the credibility of any web page or text at all.

[Insert Table 3 here]

Table 4 shows the frequencies of the subcategories of the credibility evaluations and the number of students who used the strategy in question at least once. Most of the evaluative comments (33%) expressed by the students concerned the publisher or author of the text or the expert interviewed. The second most common strategy was evaluation of the author's argumentation (opinion 10% and arguments 15%); however, this strategy was used by only six students. In fact, one student evaluated the author's argumentation 12 times, thus accounting for most of the comments in this category. Six students evaluated credibility on the basis of the style or content of the text. This category accounted for 10% of the credibility evaluations. Relatively few other strategies were used by the students.

[Insert Table 4 here]

Predictive evaluation strategies are essential when selecting the most relevant web pages from several search results. As shown in Table 5, most of the comments directed towards relevance were predictive in nature (31%) or concerned the

evaluation of topicality (27%) either on the textual or paragraph level (17%).

[Insert Table 5 here]

*The publishers of the texts read by students*

Table 6 reports the amount of reading time allocated by the students to the websites they read, sorted by publisher. Websites published by public associations, such as the well-known Finnish child welfare association, accounted for the highest proportion of students' reading time (33.5%). The second most-read type of website, accounting for 24.1% of reading time, were websites published by an expert, expert organization or network of experts. This category, which can be regarded as the most reliable source of information for the students' purpose, included research-based websites such as the Finnish Institute of Occupational Health or National Public Health Institute.

[Insert Table 6 here]

The students relied quite heavily on information presented in Wikipedia, 16.5% of total text reading time. Wikipedia was used by 80.0 % of the students.

The credibility of the publisher was somewhat questionable in nearly 10% of the websites read by the students. Reading texts from the websites with the questionable publisher credibility was not equally distributed across all students. Most of the time (76%) spent on reading these texts was accounted for by six students. Table 6 also presents the distribution of the texts students chose as source material, and it shows that 5 of the 50 chosen texts had somewhat

questionable publisher credibility.

*Evaluation profiles*

The cluster analysis indicated that the five-cluster solution was the most acceptable, as it best highlighted the salient differences between the students. Owing to the small number of students, the results of the cluster analysis must be interpreted with caution. The following five evaluation profiles were emerged (Table 7): versatile evaluators ($n = 3$), relevance-orientated evaluators ($n = 5$), limited evaluators ($n = 8$), disorientated readers ($n = 5$), and uncritical readers ($n = 3$). One student did not fit into any of the clusters.

[Insert Table 7 here]

The *versatile evaluators* used diversified evaluation strategies and their repertoire of strategies also included some cognitively demanding strategies, such as evaluation of argumentation and comparing two or more texts. The versatile evaluators used the time allocated to task effectively. They spent on average 73.7% of their time on reading texts. They mainly read texts from websites that could be regarded as having rather credible publisher. The following excerpt illustrates the evaluation of a versatile evaluator. The excerpt is a good example of an ongoing evaluation process when the reader finds a very promising website.

> *This is good information [marks part of the text]. Yes I'll take this. This REM sleep comes up all the time. This seems to be a good page,'cos it gives some reasons. This might be a good starting point. Yes, this answers this question; the title of the task [refers to the piece of the text, which he had marked]. Motor functions decline. This is a very good page. I will check*

*what kind of page this is [goes to the main page of the website]. I can trust this. The same kinds of issues have been discussed in other pages too and this sounds reasonable to me [continues reading].*

The *relevance-orientated evaluators* concentrated on evaluating relevance. They did not evaluate the credibility of information as frequently as the versatile evaluators and the strategies they used were not as diversified. The students succeeded in spending 69% of the task time reading texts and they also mainly read texts on websites where the publisher could be regarded as rather credible.

The *limited evaluators* hardly ever evaluated credibility. They evaluated relevance to a lesser degree than the versatile and relevance-orientated evaluators. They also read slightly more texts on websites with questionable publisher credibility than the versatile and relevance-orientated evaluators.

The *disorientated readers* had difficulties in locating relevant information on the Internet, therefore they spent less time on reading texts than the other students. Furthermore, the disorientated readers spent more time on formulating queries (11.1% vs. 5.2%) and on browsing search results (29.1% vs. 14.3%) than the students in the other clusters. Additionally, they spent more time on skimming and then rejecting web pages (10.4%) than other students (4.9%). One disorientated reader skimmed and rejected the same web pages repeatedly. Another disorientated reader had difficulties in formulating effective queries. When she did not find relevant web pages, she started reading her search results and making connective inferences between the example texts listed on the search results. In addition to showing disorientation, this student read her search results uncritically without looking the context of the sentences at all.

The *uncritical readers* spent more time on reading texts on websites where the

publisher had questionable credibility (43.9%) compared to the other students. Although the uncritical readers used more evaluation strategies than the limited evaluators and disorientated readers, they were not able to apply them constructively. For example, when the uncritical readers opened an unauthenticated web page they usually began to read it without paying any attention to the authority.

The student who did not fit into any of the clusters concentrated on the evaluation of credibility; he made 16 comments on credibility (the overall mean was 3.2) and he particularly evaluated the credibility of the author's argumentation (12 out of 16 comments). This student could be regarded as a reader with a critical disposition, which, in turn, differentiated him from the other clusters.

## Discussion

The results indicated, consistently with previous studies (Grimes & Boening, 2001; Metzger et al., 2003), that most of the students only seldom evaluated the credibility of the obtained information ($M = 3.2$) when using the Internet as a source of information in an authentic learning task. With respect to credibility, evaluation of the publisher or author or expert interviewed was the most frequently used strategy. When students are trying to locate useful web pages from a huge amount of available information, the evaluation of relevance is obviously more necessary than the evaluation of credibility. Thus, the students evaluated relevance ($M = 16.9$) more frequently than credibility. When evaluating relevance, predictive evaluation and evaluation of the topicality of the text either on the textual or paragraph level were the most frequently used strategies. In this

study cognitively demanding strategies, such as the evaluation of argumentation or comparing texts at the inter-textual level, were only seldom used.

It is reasonable to assume that predictive evaluation strategies are associated with effective access to relevant information (see MaKinster et al., 2002). In this study 27% of all the evaluative strategies were predictive in nature. However, they were not able to explain how effectively students were able to use the task time reading texts. One explanation for this could be that for some students predictive evaluation strategies are difficult to verbalize. More information is needed about the interrelationship between predictive evaluation strategies and effective use of the Internet. However, it was evident that the ability of some students to predict the relevance of web pages was inadequate. These students did not consider the context of the search terms when browsing search results and might open similar irrelevant pages repeatedly. They were not able to reflect on their behavior and change their search strategies, or modify their ineffective queries. Whitmire (2004) found that the weakest students tended to choose the first hits from the search results because they assumed that search engines order the results according to their relevance and credibility.

Although the students did not frequently evaluate the credibility of information, most of the students succeeded in locating and reading texts which were posted by a rather credible publisher or author. The most frequently selected sources of information were public associations, expert organizations or networks of experts, as well as Wikipedia. The reading of texts where the credibility of the publisher was in doubt tended to be done by a small group of students.

Reading on the Internet is a complex process in which defining the need for information, searching for relevant information, and deploying evaluation

strategies are interwoven. Students with ineffective search strategies may be compelled to use lower standards of evaluation. In this study some students changed their standards of evaluation during the task because they were unable to access better material.

Furthermore, the results of this study showed considerable individual variation between students when they were using the Internet as a source for an essay. Students differed in how much and how versatile the evaluation strategies they used were, how effectively they managed to allocate the task time to reading and how much time they spent on reading texts where the credibility of the publisher was questionable. The study found five evaluation profiles: versatile evaluators, relevance-orientated evaluators, limited evaluators, disorientated readers and uncritical readers. Additionally, one student, who did not fit into any of the profiles, could be regarded as a reader with a disposition towards critical thinking. He paid active attention to the author's argumentation; to cite Ennis (1987), he was "focusing on what to believe". Evaluation profiles may have a profound effect what reading material is selected on the Internet; this in turn, has an influence on the quality of learning.

The present study raises some points of interest that could usefully be addressed in future research. Because of the need for critical thinking skills when using the Internet as a source of information, it would be important to learn more about the relationship between different evaluation profiles and learning. In this study, versatile evaluators and persons able to read with a critical disposition were in the minority. It would be interesting to know how differently the same profiles would be distributed among university students and adults and whether, in particular, having a critical disposition would be more representative.

The think-aloud method is regarded as an effective method for gaining an access to on-line processing (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). In spite of its advantages, there are also some methodological concerns. In the first place, there are individual differences in students' abilities to think aloud (Pressley & Afflerbach, 1995, p. 12). For example, some students had to be prompted more than others. The timing of prompts (was a student reading a text or browsing search results at the time) might also have an effect on students' thinking-aloud. Moreover, some students might not be willing to think aloud. For example, one student, who was constantly concerned about the time remaining for the task, said: "*Well, it is a little bit difficult to talk simultaneously; it takes all one's effort*".

Secondly, the think-aloud procedure might also encourage students to process material more actively and therefore thinking aloud might improve students' performance. In this study the students were asked afterwards about how thinking aloud affected their performance. One of the students answered: "*When I was thinking aloud, somehow I was thinking deeper. When I am only browsing pages I am not thinking too much*". In the main, however, the students felt that thinking aloud either did not affect their performance or that it impaired their performance. A few students reported that it was stress caused by the research situation which impaired their performance rather than thinking aloud per se.

Thirdly, some types of information are probably more representative than others in think-aloud protocols (Pressley & Afflerbach 1995, pp. 9–10). In this study the students were not instructed to concentrate on evaluation and therefore some evaluative thoughts they might otherwise have had may have been excluded from examination. For example, the students evaluated some web pages as

irrelevant very quickly, in one or two seconds. This was especially the case when it was obvious that the web page did not relate to the task at all. Thus, the reasons for these kinds of fast decisions were not spelled out. In the future, evaluation strategies could be studied by asking students to give reasons for their decision to read or to discard a web page. It is, however, possible that these kinds of instructions might direct students' thinking. It should perhaps be emphasized in defense of this study that the aim was to examine students' spontaneous attention to evaluation.

In spite of its methodological limitations, the useful outcome of this study was the development of a coding scheme for the evaluation of Web sources, which can be taken advantage of in future research or teaching students evaluation skills. The study also provided descriptive information for teachers about the problems students encounter when using the Internet as an information source.

Traditionally students have read text books and articles selected by the teacher and thus, the appropriateness of the contents has been checked in advance. When using the Internet in schoolwork, there is a risk that some students may not be able to use the Internet effectively and might read and extract irrelevant or even misleading information. On the other hand, teachers might not be aware of the problems students encounter, because students are confident about their skills in using the Internet (Macklin, 2001, p. 306; Pierce, 1998). In this study as well, the disorientated and uncritical readers might have had an erroneous impression of their skills. They reported in the pre-questionnaire that their information searching skills are quite good, although in fact they had poor literacy skills.

This study showed that using Internet as an information source is a demanding task and students need to be scaffolded to be able to evaluate information

critically. The use of individual or collaborative scripts could be one means for supporting students. Collaborative scripts have previously been applied, among other uses, to support reading comprehension (see Kollar, Fischer & Hesse, 2006). Evaluation strategies can be considered epistemic scripts, as according to Weinberger, Ertl, Fisher and Mandl (2005), epistemic scripts can guide student's attention towards critical aspects of the task. Epistemic scripts are the kind of task strategy which could also be applied to individual learning (ibid). By means of epistemic scripts students could be guided to pay attention, for example, to the authority of websites and the purposes of the author. Stadtler and Bromme (2007) found that evaluative prompts received during Internet reading helped students to judge information critically. However, more research is needed to discover whether prompting has long-term effects on students' performance. Students rely frequently on the Internet in their schoolwork and therefore should be taught to be responsible in the use of information; in Salomon's (1998, p. 10) words: It is school's mission, particularly when colored by constructive hues, to enable students to think and to acquire skills of intelligently handling information, not to drown them in it.

## Acknowledgements

## References

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis.* Sage university papers series, no. 07-044. Beverly Hills: Sage Publications.

Aufderheide, P., & Firestone, C. M. (1993). *Media literacy: A report of the National Leadership Conference on Media Literacy, the Aspen Institute Wye Center, Queenstown Maryland, December 7–9, 1992*. Washington, D.C.: Communications and Society Program, the Aspen Institute.

Barry, C. L., & Schamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management, 34*(2–3), 219–236.

Bawden, D. (2001). Information and digital literacies: A review of concepts. *Journal of Documentation, 57*(2), 218–259.

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology, 54*(10), 913–925.

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*(2&3), 191–213.

Bruce, C. (2002). *"Information literacy as a catalyst for educational change: A background Paper,"* July 2002, White Paper prepared for UNESCO, the U.S. National Commission on Libraries and Information Science, and the National Forum on Information Literacy, for use at the Information Literacy Meeting of Experts, Prague, The Czech Republic. Retrieved September 2, 2006, from http://www.nclis.gov/libinter/infolitconf&meet/papers/bruce-fullpaper.pdf

Coiro, J. (2003). Reading comprehension on the Internet: Expanding our understanding of reading comprehension to encompass new literacies. *The Reading Teacher, 56*(5), 458–464.

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.

Connor-Greene, P. A., & Greene, D. J. (2002). Science or snake oil? Teaching critical evaluation of "research" reports on the Internet. *Teaching of Psychology, 29*(4), 321–324.

Doyle, C. S. (1994). *Information literacy in an information society: a concept for the information age.* Syracuse, NY: ERIC Clearinghouse.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.) *Teaching thinking skills: theory and practice.* A Series of books in psychology (pp. 9–26). New York: Freeman.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: verbal reports as data.* Cambridge, MA: MIT Press.

Eveland, W. P., & Dunwoody, S. (2000). Examining information processing on the World Wide Web using think aloud protocols. *Media Psychology, 2*(3), 219–244.

Facione, P. A. (2006). *Critical thinking: What it is and why it counts?* Insight Assessment. Retrieved November 1, 2006, from http://www.insightassessment. com/pdf_files/what&why2006.pdf

Gilster, P. (1997). *Digital literacy.* New York: Wiley Computer Pub.

Grimes, D. J., & Boening, C. H. (2001). Worries with the Web: A look at student use of Web resources. *College & Research Libraries, 62*(1), 11–23.

Harris, R. (1997). *Evaluating Internet research sources.* Retrieved September 12, 2006, from www.virtualsalt.com/evalu8it.htm.

*Information literacy standards for student learning. In Information power: Building partnerships for learning.* (1998). American Association of School Librarians. Chicago, IL: American Library Association and Association for Educational

Communications and Technology.

Jones, S., & Madden, M. (2002). *The Internet goes to College. How students are living in the future with today's technology.* Retrieved September 25, 2006, from http://207.21.232.103/pdfs/PIP_College_Report.pdf

Judd, V. C., Farrow, L. I., & Tims, B. J. (2006). Evaluating public web site information: a process and an instrument. *Reference Service Review, 34*(1), 12–32.

Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts–A conceptual analysis. *Educational Psychology Review, 18*(2), 159–185.

Lenhart, A. Simon, M., & Graziano, M. (2001). *Internet and education: Findings of the Pew Internet & American life project.* Retrieved September 25, 2006, from www.pewinternet.org./reports/toc.asp?Report=39

Leu, D. J., Kinzer, C. K, Coiro, J. L, & Cammack, D. W. (2004). Toward a theory of new literacies emerging from the Internet and other information and communication technologies. In R. B. Ruddell, & N. Unrau (Eds.) *Theoretical models and process of reading* (pp. 1570–1613). 5.ed. Newark, Del: International Reading Association.

Liu, Z., & Huang, X. (2005). Evaluating the credibility of scholarly information on the web: A cross cultural study. *The International Information & Library Review, 37*(2), 99–106.

Lorenzen, M. (2001). The land of confusion? High school students and their use of the World Wide Web for research. *Research Strategies, 18*(2), 151–163.

Macklin, A. S. (2001). Integrating information literacy using problem-based learning. *Reference Services Review, 29*(4), 306–313.

MaKinster, J. G., Beghetto, R. A., & Plucker, J. A. (2002). Why can't I find

Newton's third law? Case studies of students' use of the Web as a science resource. *Journal of Science Education and Technology, 11*(2), 155–172.

McClure, C. R. (1994). Network literacy: a role for libraries. *Information Technology and Libraries, 13*(2), 115–125.

Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student web use, perceptions of information credibility, and verification behavior. *Computers & Education, 41*(3), 271–290.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Pierce, A. F. (1998). *Improving the strategies high school students use to conduct research on the Internet by teaching essential skills and providing practical experience.* (ERIC Document Reproduction Service No. ED427756).

Rieh, S. Y. (2000). *Information quality and cognitive authority in the World Wide Web.* Thesis (Ph.D.). Rutgers, 2000. Retrieved February 4, 2007, from http://www.si.umich.edu/rieh/research_diss.html

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.

Salomon, G. (1998). Novel constructivist learning environments and novel technologies: Some issues to be concerned with. *Research Dialogue in Learning and Instruction, 1*(1), 3–12.

Saracevic, T. (1996). Relevance considered. In P. Ingwersen, & N. Ole Pors (Eds.) *Proceedings of the Second Conference on Conceptions of Library and Information Science CoLIS2* (pp. 201–218). Copenhagen. Denmark: Royal School for Librarianship.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2*(2–3), 191–210.

Sutcliffe, A., & Ennis, M. (1998). Towards a cognitive theory of information retrieval. *Interacting with Computers, 10*(3), 321–351.

Weinberger, A., Ertl, B., Fisher, F., & Mandl, H. (2005). Epistemic and social scripts in computer-supported collaborative learning. *Instructional Science, 33*(1), 1–30.

Whitmire, E. (2004). The relationship between undergraduates' epistemological beliefs, reflective judgment, and their information-seeking behavior. *Information Processing & Management, 40*(1), 97–111.

Wu, Y-T., & Tsai, C-C. (2005). Information commitments: evaluative standards and information searching strategies in web-based learning environments. *Journal of Computer Assisted Learning, 21*(5), 374–385.

Table 1. Subcategories of evaluation of credibility and relevance.

| Subcategory | Example |
| --- | --- |
| ***Evaluation of credibility*** | |
| Evaluation of publicist/author/expert | Here is at least somebody who is *a knowledgeable person, docent of psychology,* it is at least mentioned here, I don't know if he has been interviewed or what. |
| Paying attention to publicist without evaluative comments | We went to page nyyti.fi. Support Center for Students. |
| Evaluation of style or content | But this seems a little bit like, well I can't say, like *a story from a women's magazine.* |
| Evaluation of up-to-dateness | This is pretty *old* information |
| Predictive evaluation of credibility | This page will certainly contain some believable information [before opening the source] |
| Evaluation of writer´s argumentation | There is no evidence for this. This author could have made a reference to something. [reads following text: Continuous lack of sleep predisposes to cardiovascular diseases and even cancer.] |
| Evaluation of sources | Here the sources are marked well. This [source.] is at least reliable, *it is from science magazine.* |
| Verification of information | This Wikipedia is not so reliable, because everyone can freely add articles to it. *They talk about the same things as in the those previous pages,* all about feelings and thoughts and how they are analyzed during sleeping. |
| Evaluation of credibility without justification | Maybe that was not the most reliable page. |
| ***Evaluation of relevance*** | |
| Evaluation of topicality on the text level | Sleeping disorders [the title of the text] *are not so important for   this essay*, so I don't need any information from here. |
| Evaluation of topicality on the paragraph level | Well, the sleeping of animals has *nothing to do with this*. [a paragraph inside of the text] |
| Evaluation of relevance by comparing two or more texts | But then *this is like an encyclopedia* [opens Wikipedia from the other tab], *they just talk about sleep and this other one responds to this sleeping as a resourc*e, this one. I think I'll make some notes from this one. |
| Evaluation of novelty | Well, this information comes from a reliable source, but I think, that *there is not any essential new information.* |
| Evaluation of style or content | Well, this is maybe *too theoretical.* |
| Predictive evaluation of relevance | Then I'll check this, 'cos this Wikipedia has always had some information; and I have searched there before, so I'll check this now. |
| Evaluation of generalization | I am just thinking about, when they talked about the sleep of children, what you could take from this. |
| Evaluation of usability or mode of information | These are *audio clips; the wrong kind of material.* |
| Evaluating information as relevant by frequency of information | It was *mentioned for the third time already* in these pages, the increase in accident risk, *so it must be important.* |

Table 2. Mean percentages of time spent on different actions during the task

| Action | M% | SD | Min-Max % |
|---|---|---|---|
| Reading texts | 61.4 | 14.6 | 30–82 |
| Browsing search results | 17.1 | 10.3 | 5–46 |
| Formulating a query or using a URL address | 6.4 | 4.0 | 2–15 |
| Skimming and rejecting the web page | 6.1 | 3.7 | 1-18 |
| Downloading links and back-function | 4.1 | 2.1 | 1–8 |
| Navigating links | 3.8 | 3.5 | 0–12 |
| Other | 1.1 | 0.8 | 0–3 |

Table 3. Descriptive statistics for evaluation strategies

|                           | f   | M    | SD   | Min-Max |
|---------------------------|-----|------|------|---------|
| Evaluation of credibility | 81  | 3.2  | 3.8  | 0–16    |
| Evaluation of relevance   | 423 | 16.9 | 9.8  | 6–43    |
| Total                     | 504 | 20.1 | 11.2 | 7–44    |

Table 4. Subcategories of evaluation of credibility

| Subcategories of evaluation of credibility | Comments $f$ | % | Students who used a strategy at least once (% in parenthesis) |
|---|---|---|---|
| Evaluation of publisher/writer/expert | 27 | 33.3 | 12 (48%) |
| Evaluation of writer's argumentation | | | 6 (24%) |
| -writer's opinion | 8 | 9.9 | |
| -writer's arguments | 12 | 14.8 | |
| Evaluation of style or content of the text | 8 | 9.9 | 6 (24%) |
| Paying attention to publisher without evaluative comment | 6 | 7.4 | 4 (16%) |
| Evaluation of credibility without justification | 6 | 7.4 | 4 (16%) |
| Evaluation of up-to-dateness | 5 | 6.2 | 4 (16%) |
| Predictive evaluation of credibility | 4 | 4.9 | 4 (16%) |
| Verification of information | 3 | 3.7 | 3 (12%) |
| Evaluation of sources given in the text | 2 | 2.5 | 2 (8%) |
| Total | 81 | 100 | |

Table 5. Subcategories of evaluation of relevance

| Subcategories of evaluation of relevance | Comments $f$ | % | Students who used a strategy at least once (% in parenthesis) |
|---|---|---|---|
| Predictive evaluation of relevance | 130 | 30.7 | 23 (92%) |
| Evaluation of topicality on the textual level | 116 | 27.4 | 24 (96%) |
| Evaluation of topicality on the paragraph level | 73 | 17.3 | 19 (76%) |
| Evaluation of style or content of the text | 39 | 9.2 | 16 (64%) |
| Evaluation of novelty | 34 | 8.0 | 16 (64%) |
| Evaluation of usability or mode of information | 15 | 3.6 | 9 (36%) |
| Evaluation of relevance by comparing two or more texts | 8 | 1.9 | 6 (24%) |
| Evaluation of generalization | 6 | 1.4 | 5 (20%) |
| Evaluating text as relevant by frequency of information | 2 | 0.5 | 2 (8%) |
| Total | 423 | 100 | |

Table 6. Mean percentages of reading time on websites sorted according to publisher

| Type of publisher | Different publishers *f* | % of reading time | Number of texts chosen as source material (% in paranthesis) |
|---|---|---|---|
| Public association | 5 | 33.5 | 18 (36%) |
| Expert, expert organization or network of experts | 17 | 24.1 | 11 (22%) |
| Wikipedia | 1 | 16.5 | 8 (16%) |
| Company or trade organization | 12 | 9.8 | 4 (8%) |
| Media | 6 | 3.3 | 1 (2%) |
| Educational institute | 3 | 3.1 | 3 (6%) |
| Credibility of publisher was questionable | 14 | 9.7 | 5 (10%) |
| Total | 58 | 100 | 50 |

Table 7. Means of each cluster with respect to mean number of comments on credibility and relevance, the sum of the different evaluation subcategories used, proportion of time spent on reading and on reading texts with questionable publisher credibility

|  | Versatile evaluators | Relevance orientated evaluators | Limited evaluators | Disorientated readers | Uncritical readers |
|---|---|---|---|---|---|
| Number of comments on credibility | 7.7 | 3.0 | 0.6 | 2.6 | 3.0 |
| Number of comments on relevance | 24.0 | 30.4 | 11.8 | 10.0 | 13.0 |
| The sum of different evaluation subcategories used | 13.0 | 7.4 | 4.6 | 4.8 | 6.3 |
| Proportion of time spent on reading | 73.7 | 69.0 | 68.0 | 39.2 | 54.0 |
| The proportion of reading time spent on reading texts with questionable publisher credibility of the total reading time | 1.4 | 4.6 | 8.9 | 6.5 | 43.9 |