



Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories

Amber Walraven^{a,b,*}, Saskia Brand-Gruwel^b, Henny P.A. Boshuizen^b

^a University of Twente, The Netherlands

^b Open University of the Netherlands, The Netherlands

ARTICLE INFO

Article history:

Available online 20 February 2010

Keywords:

Information-problem solving
WWW
Transfer
Evaluation of information
Instruction

ABSTRACT

Transfer of complex cognitive skills is important when stimulating students to become life long learners. An example of a complex cognitive skill is the skill of evaluating results, information and source while solving information problems using the WWW. Especially the knowledge and use of evaluation criteria is crucial in this matter. Two educational programs to foster this skill in students were designed and evaluated. The programs were based on two different transfer theories. The first program was based on the theory that transfer of complex cognitive skills is fostered through the development of a rich knowledge structure. The second program is based on the theory that transfer is fostered by paying explicit attention to the various steps that have to be taken in a problem solving process. Effects of the two programs on increase in knowledge and use of criteria and degree of transfer were determined. Results show that both programs enable transfer.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The World Wide Web (WWW) has made its way to education. Most secondary education students use the WWW as their only source for information for class assignments and rarely use books or visit the library (Beljaarts, 2006; Jones, 2002). Although students are frequent users of the WWW, their search method and the way they use the WWW has several imperfections. Especially the way they evaluate search results (hit list), information and source (websites) is far from ideal (Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Koot & Hoveijn, 2005; Lorenzen, 2002; Lyons, Hoffman, Krajcik, & Soloway, 1997; MaKinster, Beghetto, & Plucker, 2002; Wallace, Kupperman, Krajcik, & Soloway, 2000; Walraven, Brand-Gruwel, & Boshuizen, 2009). Especially on the WWW it is crucial to evaluate search results, information and source, since the WWW lacks centralized control and regulation, and its contents can easily be altered (Metzger, Flanagan, & Zwarun, 2003). Research has shown that students do not so much lack the skills to evaluate search results, information and source, but they do not always see the need to use these skills and often have to be prompted to do so (Britt & Aglinskias, 2002; Metzger et al., 2003; Walraven et al., 2009). It is therefore important that students understand the need to evaluate and learn how to use evaluation skills, but also that they apply these skills in more courses, contexts

and tasks and not only in the tasks of one particular course. Since evaluation skills are important for all courses, it should be prevented that students evaluate search results, information and source when writing a history essay, and cut and paste without evaluating when writing a biology report. The goal of this study is to determine whether it is possible to teach the use of evaluation skills, in a way that they are transferred to new situations or tasks in other domains where they haven't been explicitly taught.

In the introduction of this article, first the complex skill to evaluate search results, information and source will be discussed. Second, theories to foster transfer of a complex cognitive skill like evaluating information will be addressed. Third, the research questions addressed in this study will be presented.

1.1. Evaluation of search results, information and source on the Internet

Imagine a 9th grade student, Sam. Sam has to write an essay on the life of philosopher Francis Bacon (1516–1626). Sam types the query 'Francis Bacon' in a search engine and views the hit list. There are several criteria Sam could use to evaluate the hit list: title and summary of the site, the kind, the address, the rank in the hit list, whether the site mentioned in the hit list is known to the user and the language (see Appendix). In this case evaluating the hit list and not just opening the first result is important, since the first hits all relate to the painter Francis Bacon (1909–1992). Sam evaluates the search results and decides to open <http://plato.stanford.edu/entries/francis-bacon/>. Now, the information on that site has to be

* Corresponding author. Address: University of Twente, Faculty of Behavioural Sciences, Curriculum Development & Educational Design, P.O. Box 217, 7500 AE Enschede, The Netherlands. Tel.: +31 534894875.

E-mail address: a.walraven@utwente.nl (A. Walraven).

evaluated. Sam can do this with several criteria, grouped in three categories: usability, verifiability and reliability. Criteria for evaluating usability are language, connection to task, audience, topicality and amount. The verifiability of information can be evaluated with the criteria author, references, information agrees with more sites, information agrees with prior knowledge, and organization. The reliability of information can be evaluated by kind of information, objectivity, primary/secondary and goal (see [Appendix](#)). Sam concludes that there is indeed information on the life of the philosopher Francis Bacon on this website. Furthermore, references and information are provided by the Stanford Encyclopaedia of Philosophy, an organization that can be contacted through this website. Sam also evaluates the *source*. He can do this on technical, usability, verifiability, and reliability grounds. Criteria for evaluating the technical side of a source are appearance and speed. The criteria for usability are language, connection to task, audience and topicality. Verifiability has only one criterion: reputation. Reliability of a source can be evaluated with the criteria kind of source and primary or secondary source. Sam notices that the site is a kind of web encyclopaedia, with a board of editors and that it has received several grants. Sam decides to copy the information from this page and conducts another search to verify and supplement the information he has found.

Sam is a fictive, ideal 9th grader. Unfortunately, not every student acts according to this ideal process. Moreover, research has shown that most evaluation criteria are not used by students and teachers when searching the WWW ([Walraven et al., 2009](#); [Walraven, Brand-Gruwel, & Boshuizen, submitted for publication](#)). Evaluating results is mostly done based on titles and summaries provided by the search engine, the number of results and the order of results ([Duijkers, Gulikers-Dinjens, & Boshuizen, 2001](#); [Fidel et al., 1999](#); [Hirsch, 1999](#); [Kafai & Bates, 1997](#); [Koot & Hoveijn, 2005](#); [Lyons et al., 1997](#); [Wallace et al., 2000](#)). Evaluating information and source is not always done based on clear, and well understood criteria but on intuition ([Koot & Hoveijn, 2005](#)).

A non-critical attitude towards information on the WWW can seduce students to cut and paste information without evaluating it ([Grimes & Boening, 2001](#); [Rothenberg, 1998](#)), resulting in reports and learning that lack quality ([Britt & Aglinskas, 2002](#)). The importance of instruction in an effective and critical use of the WWW for educational purposes has been recognized for several years, but instruction in information skills is rare and not always effective and hardly pays attention to transfer ([Walraven, Brand-Gruwel, & Boshuizen, 2008](#)). In the next section, theories to foster transfer of a complex cognitive skill like evaluating information will be addressed.

1.2. Transfer of training

Transfer, the issue of “how knowledge acquired in one situation applies (or fails to apply) in other situations” ([Singley & Anderson, 1989, p. 1](#)) is much debated in the scholarly community. Its nature, the extent to which it occurs and the nature of its underlying mechanisms is being discussed by several authors (e.g., [Barnett & Ceci, 2002](#); [Carraher & Schliemann, 2002](#); [Nokes, 2009](#); [Wagner, 2006](#)). According to several authors, transfer of complex cognitive skills can be measured in several dimensions (near versus far transfer and short-term versus long-term effects) and can be fostered in several ways (e.g., [Gick & Holyoak, 1983](#); [Thorndike & Woodworth, 1901](#); [Wertheimer, 1961](#)). A widely adopted view of transfer suggests that the “induction or construction of abstract rules and schemata, or other mental representations has been hypothesized to serve as the primary cognitive support for knowledge transfer” ([Wagner, 2006, p. 2](#)). An important theory in this view is the high road to transfer of [Perkins and Salomon \(1989\)](#), [Salomon and Perkins \(1989\)](#). According to this theory students have to be stimulated to pay

explicit attention to the various steps that have to be taken in a process and to the way these steps can be used flexibly in different situations. The high road to transfer depends on mindful abstraction from the context of learning. It is ‘the deliberate, usually metacognitively guided and effortful, de-contextualization of a principle, main idea, strategy, or procedure, which then becomes a candidate for transfer’ ([Salomon & Perkins, 1989, p. 126](#)). The conscious formulation of abstraction means answering questions like: what is the general pattern? What is needed? Which step can I take now? What rules or principles might apply? Abstracting is closely related to metacognitive skills like planning (what am I going to do), monitoring (is the process going according to plan?) and evaluating (what have I learned that I can use a next time?), thus high-road transfer can be fostered by stimulating a persons’ metacognitive skills.

The high-road transfer can be forward or backward reaching, with the present problem as point of reference. With forward reaching one abstracts situations from the current context to a potential transfer context. An example of forward reaching transfer is a child learning good study habits by setting aside a definite time for certain activities and sticking to it. The child actually schedules his or her activities in this way. When the child grows up and gets a busy job he or she still schedules priority projects in this way, so progress on that project is assured, no matter what happens. The principle (setting definite times) is so well learned that it simply suggests itself appropriately on later occasions ([Salomon & Perkins, 1989](#)). With backward reaching one abstracts in the transfer context, looking for features of the previous problem where new skills and knowledge were learned. An example of backward reaching transfer is having learned as a child to count to 10 when you felt you were losing your temper. Now, as an adult you notice that you are an impulse buyer and you want to find a way to inhibit your impulsiveness. You should try to hold back. When you think of remedies, the count to 10 strategy occurs to you. You try it out, and it helps.

[Perkins and Salomon \(1989\)](#) state that high-road transfer is important for skills that call upon strategic knowledge, like thinking skills and problem solving skills. Evaluating results, information and source when searching for information on Internet requires strategic knowledge, since it is part of the heuristic information-problem solving process. The basic assumptions of this transfer theory (conscious formulation of abstraction and stimulating metacognitive skills) could therefore be used to design instruction that fosters the transfer of evaluation skills. Instructional design based on this transfer theory should pay particular attention to strategy explication, emphasizing abstraction and de-contextualization. This means for the skills of evaluation results, information and sources that students should know the steps to be taken, strategies that can be used in the problem solving process, and how to regulate this process.

A contrasting theory on transfer does not focus on the steps of the problem solving process (by abstraction and metacognition) but emphasizes the importance of a good, extensive and well organized knowledge base and the domain specific interpretation of the skills ([Larkin, McDermott, Simoa, & Simon, 1980](#); [Perkins & Salomon, 1989](#); [Simons, Van der Linden, & Duffy, 2000](#)). This theory, which we will call the rich representation theory, is based on the way that experts organize their knowledge and tend to go about solving problems.

An expert’s extensive knowledge base includes three representations of the information: conceptual, episodic, and action representations. Conceptual representations refer to concepts and principles with their defining characteristics (like a cat is an animal with whiskers and a tail). Episodic representations refer to personal experiences with instances of concepts and principles (like I loved the cat I had when I was a child). Action representations

refer to the things one can do with the conceptual and episodic information, i.e., using that knowledge to solve a problem (like cats can be kept as a pet). When the three representations have many and strong relationships with each other (e.g., conceptual representations have a relation with concrete experiences) and with representations in other domains, the knowledge base has a high degree of connectedness. These connected, rich representations will make learning outcomes durable, flexible and generalizable. Knowledge and skills 'are not restricted to one context but reach out to other contexts and situations.' (p. 2), thereby fostering transfer.

For evaluation skills this would mean that students should have deep knowledge of concepts associated with the key concept evaluation. The instruction based on this theory should stimulate students to construct a well structured representation of the criteria to evaluate search results, information and source that can be used in different situations and while solving different tasks. Moreover, students must become aware of the usefulness of the criteria and they should experience that the use of the criteria helps to become critical websearchers. This experience makes the representation of criteria better anchored.

Both transfer theories can be used for instructional design. Research has shown that instruction indeed has positive effects on the use of evaluation skills in the same context or the same domain, but it is unknown whether transfer to new contexts or domains is achieved (e.g., Britt & Aglinskias, 2002; Gertjets & Helenthal-Schorr, 2008; Graesser, Wiley, Goldman, O'Reilly, Jeon, & McDaniel, 2007; Kuiper, Volman, & Terwel, 2008; Stadler & Broome, 2008). Furthermore, it is not clear whether instruction designed according to principles of the transfer theories actually improves evaluation skills within the domain of instruction, whether such instruction fosters transfer, and whether these theories are equally suitable for fostering transfer remains unclear.

1.3. Research questions

If students are to become critical users of the WWW in more than one domain, it is important that they use their evaluation skills in multiple contexts and various settings. And research so far does not provide much insight into how instruction must be designed to foster transfer of the complex cognitive skill of evaluation of results, information found and source of that information when searching the WWW. Furthermore we do not have insight in the instructional and transfer effect of two transfer theories, which is our focus of the present study. We will compare two educational programs based on the two described transfer theories. The first program will be based on the high-road transfer theory and will focus on the systematic approach and abstraction of general principles for evaluating search results, information and source, by stimulating metacognitive skills. This program will be called 'high road program'. The second program will be based on the rich representation theory and will focus on building a knowledge structure of criteria used in different settings for evaluating by mind mapping techniques. This program will be called 'rich representation program'. Our aim is to identify the effect and success and failure factors of both theories concerning use for instructional design. Analyses will be done both quantitative and qualitative. The research questions are: (1) What are the effects of instruction on students' evaluation behaviour, that is use of criteria for evaluating results, information and source when solving information problems on the WWW and on students task performance? (2) Do the effects of instruction based on two transfer theories (high road versus rich representation) differ in terms of transfer achieved?

2. Method

2.1. Participants

Four classes (84 students, age 14–15) of three different secondary schools participated in this study. All schools were located in small towns and all schools were equal in size (1000–1500 students). Classes were randomly assigned to one of the educational programs. In the rich representation condition one class of 24 students (teacher A) and one class of 21 students participated (teacher B). In the high road condition one class of 19 students (teacher C) and one class of 20 students (teacher D) participated. Each class had their own teacher, so in total four teachers and four classes participated. Teacher A was a very experienced web user and maintained several websites, teacher B had less experience on the web. Both teachers worked in the same school. Teacher C did not have a lot of experience with ICT. Teacher D was an experienced web user and a teacher who liked integrating ICT in his lessons.

All schools and teachers volunteered to participate in this study and were convinced of the importance of teaching students to evaluate information found on the Internet. All schools had good ICT facilities and teachers and students were used to working with ICT and the WWW.

2.2. Materials

2.2.1. The educational programs

The educational programs were designed together with the classroom teachers. Table 1 gives an overview of the lessons in both programs. The characteristic differences between the programs will be discussed.

The general subject of both programs was World War II and both programs consisted of 15 lessons of 50 min. Students in both conditions received a reader on information-problem solving and how to evaluate search results, information and source. This reader was based on the skills decomposition of the information-problem solving skill by Brand-Gruwel, Wopereis, and Vermetten (2005). This reader described the necessary phases for information-problem solving (define the problem, search information, scan information, process information and organize and present information) and steps per phase (e.g., in the search information phase the steps are: select search strategy, define search terms, and evaluate search results). It also provided information on how and why the phases and steps should be taken and also provided rules of thumb concerning evaluation criteria.

2.2.2. High road program

In the high road program the focus was on the evaluation of results, sources and information, embedded in and linked to the whole process of information-problem solving. Research showed that working with process worksheets and modelling examples is effective (Brand-Gruwel & Wopereis, 2006; van Merriënboer, 1997) for the transfer of abstract principles and strategies or heuristics. Students in this program worked on several information problems during the lessons, like "Write an article for a newspaper in which you describe how Hitler gained control over Germany". Together with this task, they received a process worksheet. This worksheet provided the students with a step by step plan to solve their information problem. For instance, students were asked to write down their evaluations of the sites and information; in order to do so they could use the reader in which evaluation criteria were discussed. Students were also asked to reflect and to answer questions like "are you satisfied with the result so far?", "did you have difficulties with certain steps?"

Table 1

The lessons of the two educational programs.

Lesson	Rich representation program	High road program
1	Subject: Causes of World War 1 (WWI) Task: Find causes of WWI on the Internet	Subject: Introduction to World War 2 (WWII) Task: Combine facts and dates on WWII
2	Subject: The course of WWI Task: Make a presentation on the daily life of a French, British, Belgium or German soldier	Subject: Treaty of Versailles Task: Answer questions on the treaty of Versailles Students receive 5 websites and an evaluation form on evaluating websites
3	Subject: Treaty of Versailles Task: Find the terms of the treaty of Versailles Students write a newspaper with reactions on the treaty from a certain point of view (e.g., German or French)	Subject: The German woman after WWI Task: Find pictures of the new independent German woman between 1924 and 1929 Students received a process worksheet with a high amount of support
4	Subject: Treaty of Versailles Task: Write a newspaper. The students are divided in groups. Some groups write a gossip paper, other groups a serious newspaper. Students use the information found in lesson 3	Subject: Art in Germany 1900–1933 Task: Listen to a presentation and write an article on a person or artist connected to the Bauhaus Process worksheet with a high amount of support
5	Subject: Weimar politics Task: Find the political history of the Weimar republic	Subject: Art in Germany 1900–1933 Task: Write an article on a person or artist connected to the Bauhaus
6	Subject: Economics in the Weimar republic Task: Make a presentation with the title: 'Economic crisis in the Weimar republic, causes and social/political consequences' Students search pictures and have to explain connections orally	Subject: How Hitler gained control Task: Write a newspaper article on how Hitler gained control Process worksheet with less support
7	Subject: 1929, economical crisis, Hitler Chancellor of Germany Task: Find the election results in Germany between 1928 and 1933 and unemployment rates between 1928 and 1933 and connect this to the results of the NSDAP	Subject: How Hitler gained control Task: Write newspaper article and draw mind map on a totalitarian state
8	Subject: Hitler becomes a dictator: images of Hitler Task: Find pictures presenting a positive image of Hitler and pictures presenting a negative image of Hitler. Analyse the pictures	Subject: Germany becomes a totalitarian state Task: Which Word Out
9	Subject: Conference of Munich 1938 Task 1: Students view 3 cartoons and have to locate the cartoons in a database Task 2: Act out the conference of Munich. Every group finds arguments for the person they have to play	Subject: Conference of Munich 1938 Task: Answer questions on the conference of Munich Process worksheet with less support
10	Subject: Discuss the start of WWII Task: None, normal lesson	Subject: Conference of Munich 1938 Task: See lesson 9
11	Subject: The course of WWII till 1943 Task: None, normal lesson	Subject: Economics between 1920 and 1939 Task: Match quotes and years
12	Subject: What to do? Adjust, collaborate or resistance? Task: Every student is appointed a term: adjustment, collaboration or resistance. They have to project themselves into a role and explain why and how you react to war	Subject: The Netherlands during the occupation Task: Correct the site (students view a website and have to improve it)
13	Subject: Daily life in the war Task: Make an exhibition about the daily life of women in the Netherlands during the war	Subject: Monument Task: Take a picture of a war monument and write down what it is and why it was created

(continued on next page)

Table 1 (continued)

Lesson	Rich representation program	High road program
14	Subject: Daily life in the war Task: See lesson 13	Subject: Concentration camps in the Netherlands Task: Write an article on the history of one of the concentration camps in the Netherlands Process worksheet without support
15	Subject: The war in our own region Task: Write an article about a war monument in your home town	Subject: Concentration camps in the Netherlands Task: See lesson 14

During the 15 lessons the support by the worksheets was faded. In the first lessons every step was explained and students received instruction on how to fill out the sheet (“Evaluate the information. Read pages 24–27 of your reader and study the worked example. Then write down the criteria you used”). Later on, worked examples were no longer presented and eventually only the main questions were presented. Fig. 1 shows the translation of two pages of a process worksheet.

2.2.3. Rich representation program

In this program rich representation will be visualized by using mind map techniques, because making mind maps or knowledge structures is effective for the development of rich representations and the development of a good, extensive and well organized knowledge base (Ausubel, 1963, 1968; Ausubel, Novak, & Hane-

sian, 1978; Bransford, Brown, & Cocking, 1999; Buzan, 1995; Novak, 1990). In the rich representation program, criteria to evaluate results, information or source were addressed to give students insight in the evaluation criteria, how these criteria are inter-linked, and when it is relevant to use certain criteria. Each of the 15 lessons started with a short discussion concerning criteria that can be used to evaluate and why it is important to use these specific criteria. After the discussion, students worked on one or more small tasks. An example of a task is: “Find the terms of the treaty of Versailles”. While solving these tasks, students were asked to focus on a specific criterion, for instance whether the author of the content was mentioned on the website. The lesson ended with another discussion about what they had learned about the history content and about evaluation of the information. Together with the teacher, students constructed a gradually advancing knowl-

Worksheet

Processworksheet: Bauhaus

Before you start:

Solving an information-problem consists of several phases. Shortly describe which phases there are.

- Phase 1:
- Phase 2:
- Phase 3:
- Phase 4:
- Phase 5:

Phase 1: Defining the problem

Which steps do you take during this phase? Shortly describe them here.

Formulate a main question

Formulate a main question. Make a mindmap to structure your thoughts if necessary. Take another look at page 10 en 11 of the reader on information problem-solving.
You can draw your mind map here.

Fig. 1. Translation of Process.

edge structure (mind map) during the lessons. Fig. 2 shows the knowledge web of class A after 15 lessons.

The critical difference between the programs therefore lies in the guidance on the process provided by the worksheets in the high road program and the discussions on criteria and their relations in the rich representation program.

2.3. Measurements

2.3.1. Evaluation of hit list

Evaluation of hit list was measured with four different information problems. Two tasks were in the domain of history (domain of instruction) and two in the domain of biology (transfer domain). The topics of the history tasks were 'Anastasia Romanov' and 'the Watergate affair', and the topics of the biology tasks were 'Super Size Me' and 'influence of sex before a sports match'. For each task a manufactured hit lists of 14 results on paper was provided. To examine if students could correctly evaluate a hit list, students had to select three sites they would open and three sites they would not open for each task. They could highlight and circle the parts of the hit list they based their decision on. Participants received a point per website if their evaluation was correct. That is, a point for choosing an appropriate site they wanted to open and a point for choosing not to open an inappropriate site. Maximum score was six points per hit list.

2.3.2. Evaluation of websites

Evaluation of websites was measured with four information problems and four booklets with eight printed websites. Two tasks were in the domain of history (domain of instruction) and two in the domain of biology (transfer domain). The first history information problem regarded whether the Bush administration was be-

hind the attacks of 9/11, and the second regarded whether the NASA was responsible for the first landing on the moon. The biology tasks involved whether the Dutch non-smoking policy was effective enough and whether or not teenagers were more often infected with sexually transmitted diseases. To examine whether students could identify crucial features to base an evaluation on, students were asked which sites and what information they would or would not use, given the provided problem. They could highlight parts of information or features of the website they based their decision on. A list with criteria that could be noticed for the given websites was created. The score was based on how many of these criteria were noticed by participants. If students had circled a certain area on the site or written down a comment like: "Site is old" they received a point. The maximum scores for the history tasks were 30 and 26, and the maximum scores for the biology tasks were 23 and 29. Scores of participants were calculated as percentages of the maximum scores. So, if a participant scored 18 criteria on the history tasks were the maximum score was 30, his final score is 60.

2.3.3. Think-aloud protocols

To determine how students evaluate result, information and source online, 11 students were given two tasks (history and biology) before and after the intervention they had to solve while thinking aloud. The biology tasks used were about elections and the Cold War and were formulated as follows: 'Prior to governmental elections, a lot of polls are presented. What is the effect of these polls on the votes of the public?' and 'What was the influence of the "Pay of Pigs incident" on the course of the Cold War?' The two used biology tasks were about nourishment and pollution and the assignments were: 'What is the effect of genetically manipulated foods on the human body?' and 'What is the effect

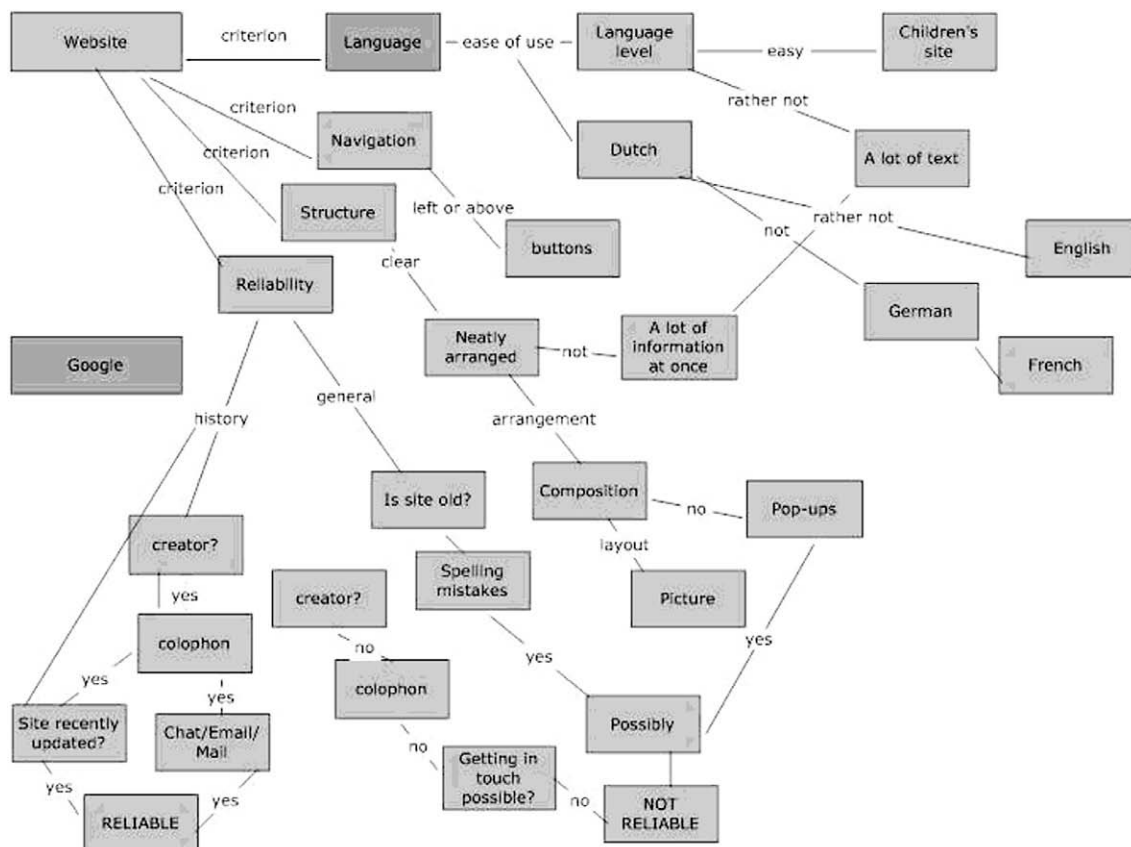


Fig. 2. Knowledge web of class A after 15 lessons (translated from Dutch).

of pollution on the quality of tap water?’ All tasks were preceded by the following text: “In this task you have to answer a question. You can search for information you need to answer the question on the Internet, and you can copy and paste the information in a word-file. Finally, use the information found to answer the question based on the information found in a few sentences. You have 30 min for this task. Think aloud while you are searching.”

2.3.4. Field notes

In each class the implementation of the programs was observed three times. Field notes of these observations served as secondary material that could possibly explain the results. During these observations special attention was given to the interaction between the students and between the students and the teacher concerning evaluation behaviour and the use of evaluation criteria.

2.4. Design and procedure

A pre-test–post-test with two conditions was used to determine the effect of both programs on students’ evaluation behaviour (e.g., evaluation of hit list, websites and information). Table 2 presents the design of the experiment.

Before the first lesson, all students did a pre-test consisting of a hit list and website evaluation task. These tasks were counterbalanced and rotated. There was no maximum time to finish the tasks. Three students in three classes and two students in one class also solved the two tasks thinking aloud. After these students read the first task, they had to write down what they already knew about the topic. Then, while thinking aloud, they searched the WWW for a maximum of 30 min to solve the problem. Information found could be stored in a word-file. The second task followed the same procedure. The images of the computer screen and the audio were recorded on videotape. Extra audio files were created with a laptop and the program Audacity. After finishing the second task, participants were asked to verbalize their search procedure. (e.g., “Could you tell me how you searched to solve this last task?” “Which things did you pay attention to while searching?”).

After the pre-test the 15 lessons were given by the teachers. In each class three lessons were observed by the first author. A week after the last lesson the students completed the two evaluation tasks again (different information problem) and the same four students solved two tasks while thinking aloud (different tasks). Pre- and post-test tasks were counterbalanced and rotated. Half of the students received history task 1 (hit list and website) and biology task 1 (hit list and website) during the pre-test, and the remaining half received history task 2 and biology task 2. Furthermore, half of the students started with the history tasks, and the other half started with the biology tasks. During the post-test students received a different biology and history task than during the pre-test. Students who had made task 1 in a domain, now made task 2 and vice versa. Again, the order of tasks (starting with history or biology) differed between students.

Table 2
Design of the study.

O1	X1	O2	N = 39
O1	X2	O2	N = 45

O1 = two tasks evaluation hit list (history and biology), two tasks evaluation information and source (history and biology), two thinking aloud tasks (history and biology).

X1 = high road program (three observations per class).

X2 = rich representation program (three observations per class).

O2 = two tasks evaluation hit list (history and biology), two tasks evaluation information and source (history and biology), two thinking aloud tasks (history and biology).

2.5. Data analysis

2.5.1. Think-aloud protocols

All think-aloud protocols were transcribed verbatim. The use of criteria was analyzed in two contexts: (1) the use of criteria to evaluate search results, information and source and (2) the use of criteria other than to evaluate search results, information and source. To analyze which criteria students use to evaluate search results, information and source a coding scheme, developed by Walraven et al. (2009) was used. Only the evaluations that were explicitly mentioned were scored. Students also discarded pages without telling why. In those cases they evaluated the site, but it was not clear on what grounds. The scored criteria were counted.

To analyze the use of criteria other than to evaluate, a more qualitative approach was used. All utterances containing knowledge on criteria for evaluating search results, information or source were classified. An utterance is a sentence or a group of sentences focused on one subject. Furthermore, the utterances were divided based on the goal of the utterance: (1) justifying an action based on an evaluation criterion (e.g., I am scanning the page to see if I can find information that answers my question), (2) adjusting information-problem solving strategy based on an evaluation criterion (e.g., I have to compare this to more sites, so I am going to go back to Google and do another search) and (3) using an evaluation criterion without drawing conclusions (e.g., It is mentioned here who the author is).

2.5.2. Task performance on think aloud task

The solution to the task was judged based on three criteria: (1) quality of the sources, (2) triangulation of information and (3) coverage of information. The quality of sources was judged based on the evaluation criteria mentioned in the Appendix. To that end each website that was used was evaluated based on these criteria. Each website received a score of zero to four points. Next, a mean score per student per task was calculated by adding the scores of the used websites and dividing it by number of sites used. The triangulation of information refers to whether the answer contains information of more than one source. Zero points for only one site, one point for more than one site. Coverage of information used was determined by identifying content related elements. Coverage can be zero to four points. Zero points if no answer was provided or if the answer did not have any of the content items as specified in the correction model. One point if the answer contained only a part of a content item from the correction model. Two points if the answer contained a completed item from the correction model. Three points if the answer contained more than one item from the correction model. Four points if every item from the correction model was mentioned. Maximum score for task performance was nine.

3. Results

3.1. Evaluation tasks hit list and websites

Table 3 provides the means and standard deviations of hit list and website evaluation task score. Scores are provided for the history tasks and biology tasks. The latter are the transfer tasks.

3.1.1. Effects of the instruction

To determine the effects of the programs on students’ use of criteria for evaluating results (hit list) in the domain of instruction (history) a repeated measures ANOVA analysis with program as between factor was performed. There was no significant main effect on the factor ‘time’, $F(1,82) = 0.99$, $MSE = 1.20$, ns . The main effect for ‘program’ was significant, $F(1,82) = 3.38$, $MSE = 6.40$, $p = .05$, $\eta^2 = 0.05$. The high road students scored higher overall. A marginal

Table 3

Means and standard deviations of hit list and website evaluation task score.

	High road program (N = 39)		Rich representation program (N = 45)	
	Mean (SD)		Mean (SD)	
	Pre-test	Post-test	Pre-test	Post-test
Hit list history	5.0 (1.1)	4.9 (1.0)	4.3 (1.5)	4.8 (1.1)
Websites history	13.8 (8.7)	14.4 (9.7)	17.7 (9.4)	21.7 (10.0)
Hit list biology	4.8 (1.1)	4.5 (1.1)	4.6 (1.2)	4.8 (1.0)
Websites biology	16.5 (11.8)	19.2 (10.6)	17.6 (10.1)	21.6 (9.3)

interaction effect between 'time' and 'program' was found, $F(1,82) = 3.07$, $MSE = 3.70$, $p = .08$, $\eta^2 = 0.04$. This means that students in the rich representation condition learned slightly more than students in the high road condition. Because both programs were implemented in two different classes, it was determined if class effects occurred within conditions. No significant class effects were found.

Effects of the programs on students' use of criteria for evaluating information and source (websites) were determined by using a repeated measures ANOVA on the results of the history websites evaluation task with program as between factor. A marginal main effect was found for 'time', $F(1,82) = 3.65$, $MSE = 217.284$, $p = .06$, $\eta^2 = 0.04$. That is students in both programs slightly improved their evaluation scores. A significant main effect for 'program' was found, $F(1,82) = 11.07$, $MSE = 1325.64$, $p = .00$, $\eta^2 = 0.11$. The rich representation condition scored higher overall. No significant interaction effect between 'time' and 'program' was found, $F(1,82) = 2.13$, $MSE = 126.78$, ns .

Again it was determined if there were class effects within the conditions, because each condition existed of two classes. No significant difference between classes was found in the high road program. Within the rich representation condition a significant difference between classes was found, $F(1,43) = 7.03$, $MSE = 357.33$, $p = .01$, $\eta^2 = 0.14$. Students in class A increased from 14.29 (SD 8.7) to 22.04 (SD 9.1) while the scores of class B decreased from 21.62 (SD 8.7) to 21.38 (SD 11.1).

3.1.2. Transfer effects of instruction

To determine the effects of the programs on students' use of criteria for evaluating results (hit list) on the biology task a repeated measures ANOVA analysis with program as between factor was performed. There was no significant main effect on 'time', $F(1,82) = 0.40$, $MSE = 0.37$, ns , and also no main effect on the factor 'program', $F(1,82) = 0.02$, $MSE = 0.02$, ns . However, a significant interaction effect between 'time' and 'program' was found, $F(1,82) = 4.11$, $MSE = 3.57$, $p = .05$, $\eta^2 = 0.05$. The scores of the rich representation condition increased while the scores of the high road condition decreased. Furthermore, no class effects were found within the two conditions.

A repeated measures ANOVA analysis of the results on the evaluation of *biology websites* showed a significant main effect on the factor 'time', $F(1,82) = 5.79$, $MSE = 468.34$, $p = .02$, $\eta^2 = 0.07$. This means that both programs had a positive effect on students evaluation behaviour. No main effect was found for the factor 'program', $F(1,82) = 0.96$, $MSE = 130.67$, ns , and also no interaction between 'time' and 'program' was found, $F(1,82) = 0.21$, $MSE = 16.58$, ns .

Moreover, it was examined if there were class effects within the conditions. No significant difference between classes was found in the high road program. Within the rich representation condition a significant difference between classes was found, $F(1,43) = 3.82$, $MSE = 289.54$, $p = .06$, $\eta^2 = 0.08$. Students in class A increased from 15.46 (SD 10.6) to 22.80 (SD 7.2), while the scores of class B maintained almost the same; 20.05 (SD 9.2) in the pre-test and 20.19 (SD 11.1) in the post-test.

3.2. Think-aloud protocols

Table 4 contains the frequencies of the criteria used performing the history and biology tasks in the pre- and post-test.

As can be seen from Table 4 students evaluated search results using the title or the summary of the hit. They evaluate information by questioning if the information is useable for solving the task. Information is hardly evaluated on reliability or verifiability. The website (source) is also not evaluated often. Appearance or lay-out is decisive.

Examining the differences in total number of evaluations between the pre- and post-test in both conditions Wilcoxon signed-rank tests were performed. Only a significant difference was found in the high road condition ($N = 6$). The median in the pre-test was 14.5 and in the post-test 8.0, $T = 0.00$, $p = .03$, $r = -.64$. The number of evaluations decreased between pre- and post-test.

To gain more insight in how students use the knowledge concerning the criteria during their search for information further qualitative analyses were performed. Table 5 presents the results of the qualitative analyses.

Utterances containing knowledge on evaluation criteria for evaluating search results, information and source were not found in all 11 protocols. In total five students explicitly used knowledge on criteria while thinking aloud. Two students during the history pre-test, four students during the history post-test, one student during the biology pre-test and two students during the biology

Table 4

Frequencies of used criteria per domain and during pre- and post-test.

Sub skill	Criteria	History		Biology	
		Pre-test	Post-test	Pre-test	Post-test
Evaluate search results	1. Title/summary	54	41	58	42
	2. Kind (site/PDF/)	5	2	3	4
	3. Address	2	2	0	2
	4. Rank in hit list	0	0	0	2
	5. Known to user	3	4	3	1
	6. Language	0	0	0	1
Evaluate information	A. Usability				
	1. Language	3	2	1	2
	2. Connection to task	35	15	33	26
	3. Audience	0	0	0	0
	4. Topicality	0	1	0	0
	5. Amount	4	2	0	1
	B. Verifiability				
	1. Author	0	1	0	1
	2. References	0	1	0	1
	3. Information agrees with more sites	3	3	0	3
	4. Information agrees with prior knowledge	1	0	0	0
	5. Organization	0	0	0	0
	C. Reliability				
	1. Kind of information	7	2	2	5
	2. Objectivity	2	0	0	0
Evaluate source	3. Primary/secondary	0	0	0	0
	4. Goal	0	0	0	0
	A. Technical				
	1. Appearance	7	4	6	0
	2. Speed	1	1	0	0
	B. Usability				
	1. Language	0	0	0	0
	2. Connection to task	1	0	0	0
	3. Audience	0	0	0	0
	4. Topicality	0	0	0	1
	C. Verifiability				
	1. Reputation	1	2	0	1
	D. Reliability				
	1. Kind (site/PDF)	2	1	1	1
	2. Primary/secondary	0	0	0	0

Table 5
Number of utterances containing knowledge of criteria.

	History		Biology	
	Pre-test (n = 2)	Post-test (n = 4)	Pre-test (n = 1)	Post-test (n = 2)
Number of utterances with goal justifying an action based on an evaluation criterion	1	2	1	0
Number of utterances with goal adjusting information-problem solving strategy based on an evaluation criterion	3	4	0	1
Number of utterances with goal using an evaluation criterion without drawing conclusions	1	4	0	3

post-test. In total 20 utterances were labelled as containing knowledge on evaluation criteria for evaluating search results, information and source. The utterances were divided in three categories: (1) justifying an action based on an evaluation criterion, (2) adjusting information-problem solving strategy based on an evaluation criterion, and (3) utterances with the goal of using an evaluation criterion without drawing conclusions.

Table 6 shows some examples of utterances from each category.

3.3. Task performance

A task performance scores was calculated for students who performed the thinking aloud tasks. Table 7 provides the means and standard deviations of these scores for both conditions.

A Wilcoxon signed-rank test was performed to test if there was a difference between both conditions in product scores between pre- and post-test on the history tasks. For students in the rich representation program the total product scores between the pre-test

Table 6
Examples of utterances containing knowledge of criteria.

Categories	Examples
Utterances with goal justifying an action based on an evaluation criterion	"I'm am scrolling to the top of the hit list, checking to see if I skipped useful hits. I usually start at the top and click my way down, because the best links are on top" "I am going to check if this is reliable"
Number of utterances with goal adjusting information-problem solving strategy based on an evaluation criterion	"I have found this information on Wikipedia, and that is not always reliable, so I am going to look for another site with the same information" "Everything I find has to do with nature, so I have to adjust my query and add human body"
Number of utterances with goal using an evaluation criterion without drawing conclusions	"This is convenient, the author and date are mentioned" "Last updated September 2004"

Table 7
Means and standard deviations of the product scores (maximum = 9).

	High road program (N = 6)				Rich representation program (N = 5)			
	Mean (SD)	Pre-test	Post-test	Pro test	Mean (SD)	Pre-test	Post-test	Post-test
		history	history	biology		history	history	biology
Total product	6.0 (1.1)	5.1 (1.0)	5.2 (2.1)	5.4 (1.6)	3.4 (2.1)	4.5 (2.5)	4.2 (1.7)	3.7 (2.0)
Quality of Sources (maximum = 4)	2.3 (0.6)	2.5 (0.5)	2.4 (1.3)	1.9 (1.0)	2.6 (1.5)	2.3 (1.6)	3.0 (1.2)	1.7 (0.8)
Triangulation of information (maximum = 1)	1.0 (0.0)	0.5 (0.5)	0.8 (0.4)	0.8 (0.4)	0.0 (0.0)	0.4 (0.5)	0.4 (0.5)	0.4 (0.5)
Coverage of Information (Maximum = 4)	2.7 (0.8)	2.2 (0.7)	2.0 (1.3)	2.7 (0.5)	0.8 (0.8)	1.8 (0.8)	0.8 (0.8)	1.6 (1.5)

(Mdn. = 4.0) and the post-test (Mdn. = 5.0), $T = 3.00$, ns , $r = -.39$ in the history task did not significantly differ. With regard to quality of the sources, triangulation of information and coverage of information, there was only a marginal difference between pre (Mdn. = 1.0) and post-test (Mdn. = 2.0) on coverage of information, $T = 0.00$, $p = .06$, $r = -.19$. For students in the high road program there was a significant difference in product scores between the pre-test (Mdn. = 5.5) and the post-test (Mdn. = 5.3), $T = 0.00$, $p = .04$, $r = -.59$. The total product score of students in the high road program significantly decreased between pre- and post-test. With regard to quality of the sources, triangulation of information and coverage of information, there were no significant differences between pre- and post-test.

To test if there was a transfer effect, differences in product scores between pre- and post-test on the biology tasks was determined. For students in the rich representation program there was no significant difference in the total product scores between the pre-test (Mdn. = 4.0) and the post-test (Mdn. = 4.0), $T = 4.50$, ns , $r = -.06$, neither were differences found on the quality of the sources, triangulation of information and coverage of information. For students in the high road program there was no significant difference in the number of evaluations between the pre-test (Mdn. = 4.9) and the post-test (Mdn. = 5.4), $T = 9.00$, ns , $r = -.09$, neither were differences found on the sub categories.

3.4. Field notes

3.4.1. Rich representation program: the students

During the first observations in the rich representation classes it seemed that students were mostly seeking the one, right answer to the information problem. A student made the remark: "I can't find the answer to question two; I'm going to skip it". The entire information problem was used as a query in a search engine. For instance: What were the outcomes of the treaty of Versailles? Websites with essays made by peers were mostly used to find an answer. Observations further along the training showed that students more often used more than one website for their answer, more often checked it they could contact the author of the site and more often checked if their were references on the site.

3.4.2. Rich representation program: the teachers

In the first lesson that was observed teacher A looked back on a previous lesson and commented on the assignment of some students. Than, he explained the upcoming lesson and let students work on the assignment. After 15–20 min and after asking several students some questions about the websites they visited he asked students to stop working and started a discussion about criteria for evaluating websites. In further observations it became clear that the teacher A adapted the lessons more to the needs of the students. His students needed more structure and more concluding remarks after the lessons.

Teacher B started the lesson with a recapitulation of evaluation criteria educated thus far. To introduce the new criterion central in the upcoming lesson, he asked students to visit two websites and

explain which site they would prefer and why. After a short discussion the teacher explained which criterion will be focused on that day and introduced the assignment. While students worked, the teacher walked around, answering questions of students. The students in class B seemed to accept the new lessons, but the teacher was worried that the knowledge on criteria would not sink in properly with the students.

3.4.3. High road program: the students

During the first observations in the high road program it became clear that students had difficulties filling out the process worksheets. It took them a lot of time, and most students would rather finish the assignment first and tried to fill out the process worksheet afterwards. In one high road program class (C) students became used to the worksheets and filling them out became part of the assignments. Students who filled out the sheets systematically and evaluated websites based on the criteria in the reader on information-problem solving, often finished the assignments as first of their class. In the other class (D) most students would not accept the worksheets and did not write down how they evaluated websites. Even when the process worksheets were shortened by the researcher and students only had to write down their evaluations, most students did not fill out the sheets.

3.4.4. High road program: the teachers

During the observations it was obvious that teacher C was well prepared for each lesson and followed the lessons exactly as designed. She was very strict in having students fill out the worksheets. She listened to her students when they complained about how much work it was to fill out the worksheets and kept explaining to them why the worksheets were important. Students in her class were hardworking students.

Teacher D had a more difficult group of students, who did not really want to change in their approach to the WWW. Students complained about the process worksheet. He mentioned to the students that it was important to fill out the sheets, but he did not actually check if students did so; he was not as strict in having students fill out the sheets as teacher C. Teacher and students were uncomfortable with the assignments.

4. Discussion

Two educational programs based on two different transfer theories have been designed and tested. Effects of the two programs on increase in knowledge and use of evaluation criteria and degree of transfer were determined. Two questions central in this article were (1) what are the effects of instruction on students' evaluation behaviour, that is use of criteria for evaluating results, information and source and on task performance and, (2) do the effects of instruction based on two transfer theories (high road versus rich representation) differ in terms of transfer achieved?

Concerning the first research question we can conclude that students in both programs benefited from the lessons in terms of improved evaluation behaviour regarding the evaluation of websites and information. Students in both conditions improved but no interaction with program was found. This means that there were no differences in the gain in scores between the students in the high road program and the rich representation program. However, within the rich representation condition the classes differed significantly. One class markedly improved while scores in the other class slightly decreased. An explanation can be found in how the lessons were implemented. Field notes of the systematic class observations showed that both teachers in the rich representation program followed the program, but that teacher A was more capable of adjusting the lessons to the needs of the students.

So, one can conclude that for stimulating students to become better evaluators of information and sources found on the WWW the instruction based on principles from both transfer theories had an effect.

No overall improvement was found for the evaluation of hit lists. However, the students in the rich representation program performed slightly better than the students in the high road program. Again we found that one class in the rich representation condition outperformed the other, presumably due to the same difference in adaptation of instruction to class needs.

To gain more insight in the process of evaluation when searching the WWW for information, 11 students also solved an information problem on the WWW while thinking aloud. What can be concluded is that students evaluate hit lists by the title and summary and that they evaluate information and source by the connection to the task. So, usability criteria are used for a first evaluation. After an initial positive evaluation on usability, a critical websearcher would take the next step and evaluate the information and source based on reliability and verifiability. However, our students hardly use more sophisticated criteria, like author or information agrees with more sites for evaluating the reliability and verifiability, before and after the intervention. They do not take the next step, after evaluating information on usability. Moreover, results concerning task performance of these 11 students did not reveal a large improvement either after the intervention. In the rich representation condition a marginal effect was found on coverage of information; this suggests students' answers were more complete after the program. In a nutshell we can conclude that instruction in evaluation of results information and source has a positive effect on students' web searching behaviour.

The second research question addressed the effects of the programs in terms of transfer. The findings regarding the evaluation of websites and information the results on the transfer task (biology) are in line with the results of the tasks performed in the domain of instruction (history). Students in both programs improved their evaluations. One class in the rich presentation condition performed better than the other class, similar to the history tasks. The evaluation of the hit list in the biology context gave the same results as in the history context. The students in the rich representation condition outperformed the students in the high road condition, scores of students in the rich representation condition improved, while score in the high road condition decreased. The thinking aloud protocols collected from a small group of students revealed the same pattern as the history task. Students evaluated, but did not use much sophisticated criteria. Furthermore, their task performance score did not improve in both conditions.

It can be concluded that both training programs, one based on the high road principles of Perkins and Salomon (1989) and one based on the rich representation theory of Simons et al. (2000), can make a difference in stimulating students to become more critical websearchers. One can even say that the rich representation program realized a stronger effect. However, the class observations made us conclude that the implementation of the lessons was not always satisfactory. One teacher and his students were uncomfortable with the new teaching approaches. The changes due to new working formats they are facing may look trivial, but are not. Students have to become used to work with assignments in which they have to search for information themselves and the teachers or textbook no longer provides them with information. This makes that students must become aware of the different kinds of information, and that information is not always true and reliable. This may be very difficult for a subgroup of the students. Research has shown that students' epistemological beliefs about information on the WWW can range from a view that the WWW contains true and specific facts and that information can be accepted without critical evaluation, to a view that doubts that the WWW is a good

source of true factual knowledge and information should be checked against other sources, reason and prior knowledge (Bråten, Strømsø, & Samuelstuen, 2005). This epistemological beliefs are activated when students search the WWW and have to evaluate information (Mason & Boldrin, 2008). In order to become a good web searcher, a shift in epistemological beliefs has to be made from the view that information can be trusted, towards a view that information has to be evaluated. This shift in beliefs takes time. Although students improved their evaluations of websites, the improvement was not optimal. Fifteen lessons may not have been enough to achieve the shift in beliefs and make students really critical websearchers.

This study tried to shed light on the scientific discussion of how to design instruction to foster transfer of complex cognitive skills. The high-road transfer theory of Perkins and Salomon (1989), Salomon and Perkins (1989) states that students have to be stimulated to pay explicit attention to the various steps that have to be taken in a process and to the way these steps can be used flexible in different situations. Research has shown that process worksheets and modelling examples are effective in stimulating the use of a systematic approach (Brand-Gruwel & Wopereis, 2006; van Merriënboer, 1997). Our study has demonstrated the positive effect of this method on students' evaluation skills. At the same time, it also showed that its impact is moderated by the correct use of the process worksheets. While most students in one class did not fill out the worksheets while working on the assignments and the teacher did not put much effort into helping or convincing them to fill out the worksheets correctly, transfer was still achieved. Perhaps the transfer effect would have been bigger, if both teachers had been strict in the use of the process worksheets. A weak point of the worksheets was that students did not like to fill them out and that it asks a lot of the teacher to implement them in the correct way. The solution to this problem from one teacher to be strict and keep letting students fill out the sheets was effective, but perhaps not the most motivating for students. So, it is important to design process worksheets in such a way that the most important steps are provided and that students experience the sheets as an aid in their learning process and not just as a burden that must be filled out to please the teacher. It would also be advisable to design sheets that can be adapted to the students' needs. A good websearcher does not need a highly-structured sheet, whereas the novice may benefit from such a sheet. This study also suggests that the length of the process worksheets should not be too long and that process worksheets should be alternated with other tools to stimulate students to pay attention to the steps in the process. Most students complained about the fact that they had to fill in so many of the same sheets.

A strong point of the high road program is the systematic approach to the whole problem solving process. Students not only had to evaluate hit lists and websites when performing a task, they were also asked to define the problem and to organize and present information. This whole task approach provided students with a 'way to work' instead of just some tips and tricks. This whole task approach requires students to perform all the constituent skills that make up the whole complex skill during task performance (van Merriënboer, 1997). The coherence between the skills and the way the skills are interlinked make students more aware of their own search process.

The rich representation theory of Simons et al. (2000) emphasizes the importance of a good, extensive and well organized knowledge base and the domain specific interpretation of the skills. Research has shown that making knowledge structures or mindmaps is an effective way to obtain a knowledge base (Ausubel, 1963, 1968; Ausubel, et al., 1978; Bransford, et al., 1999; Bu-

zan, 1995; Novak, 1990). This study showed that not only building individual knowledge structures is effective, but group discussions and visualizing a knowledge structure for the entire group is also an effective way to obtain a knowledge base and eventually achieve transfer.

An important strong point of this program was the use of group discussions and that the teacher could decide on the amount and the length of discussions. They could be adapted to the students' needs. Also important is that building the mind map was a cooperative process of the students and the teacher. It was not just the teacher providing the constructs and criteria, but also the students. This can have a positive effect on students' motivation to use the constructs or criteria when searching and evaluating information on the WWW. This strong point is at the same time a pitfall. If a teacher is not capable of leading these discussions, the discussion would probably be less effective and the resulting mind map would not include the important evaluation criteria.

The rich representation theory focuses on the knowledge that is at the heart of the cognitive skill that students have to learn. This leads to instruction that does not pay much attention to knowledge surrounding that skill or the use of that skill. As a consequence, a weak point of instruction based on the rich representation theory is that students miss an overview of the entire process and of the steps that can be taken. If students do not learn *how* to define an information problem correctly and choose the right search terms for instance, results on a task may still be low, despite the fact that students are able to evaluate results and information. A hit list based on the wrong query, still results in information of less quality even though the hit list is evaluated. If the student does not know which queries to use to retrieve the best possible hits, his task performance will still not be optimal.

It must be noted that besides the way the two transfer theories were put into practice, the teachers also caused variation in learning results. In the high road program, the two teachers used the process worksheets differently and in the rich representation program the skill to adapt the group discussion to the needs of the students differed between teachers. An explanation for the difference in executing the programs between teachers could lie in the design process. During the design of the programs teachers worked in teams on the programs using one transfer theory as point of departure. Perhaps, the focus of the teachers was less on the important principles from the theory and more on the practical aspects of the lessons (e.g., Deketelaere & Kelchtermans, 1996). So when for example in the case of the high road program students rebelled against the process worksheets, the teacher forgot about the importance of the sheets with regards to the transfer theory and was more focused on making sure students went back to work. This kind of problem solutions might have been prevented if more attention had been paid in the design phase to which parts of the theory are essential and how they are expressed in the programme. A limitation of this study could therefore be the amount of emphasize on the transfer theories during the design of the programs.

Another limitation in the study is the way data were collected regarding how students evaluate result, information and source online. The method used, thinking aloud, did not consistently reveal the thought processes underlying actions, as some students discarded sites without explaining why. Research by van Gog, Paas, Van Merriënboer and Witte (2005) showed that cued retrospective reporting (the original task and a record of eye movements is used as a cue for retrospection) revealed more complete information than concurrent and retrospective reporting while trouble shooting with electrical circuits. Participants provided more action information, 'how' information and meta-

cognitive information. Perhaps this method could also work for investigating information-problem solving processes and the use of criteria. The search could be discussed in retrospect with the students and students could see the eye movement data to remember what they were looking at while taking specific actions. Perhaps this would result in a better explanation of why they discarded some sites in only a few seconds.

The goal of the present study was to identify success and failure factors of both theories. The way the rich representation theory was put into practice was successful in terms of transfer. A success factor was the group discussions on evaluation criteria. However, these group discussions could also be a failure factor. When teachers are not able to structure these discussions, they would probably not be so effective. A weak point in the design of the program was the lack of knowledge building on the entire problem solving process. In the high road program transfer was also achieved. The success factor was the systematic approach

to the whole process. A failure factor was that the students did not like the process worksheets *and* that a strict way of dealing with them was required, a combination that can pose huge problems for the teacher. The strong points of instruction based on the two theories can largely compensate for the weak points of the other. Hence, it would be wise to combine the two theories and design a new instructional program to achieve a higher effect of the instruction and a higher amount of transfer. Van Merriënboer (1997) also advocates improving mental models or cognitive schemata as well as learning a systematic approach and cognitive strategies. This also pleads for a combination of both transfer theories.

Acknowledgement

This research was supported by The Netherlands Organisation for Scientific Research (NWO) under Project No. 411-03-106.

Appendix. Evaluation criteria

Sub skill	Criteria	
Evaluate search results	<ol style="list-style-type: none"> 1. Title/summary 2. Kind (site/PDF/) 3. Address 4. Rank in hit list 5. Known to user 6. Language 	<p>What is the title given by the search engine and what is the contents of the summary?</p> <p>What kind of source is it, a website, a word document, a PDF file?</p> <p>What is the address? Is it a .com or .org address?</p> <p>How many results are there in total and what is the rank of the result I am evaluating?</p> <p>Have I used this site before, or have I heard good or bad things about it?</p> <p>Is the site in a language I prefer and/or understand?</p>
Evaluate information	<ol style="list-style-type: none"> A. Usability <ol style="list-style-type: none"> 1. Language 2. Connection to task 3. Audience 4. Topicality 5. Amount B. Verifiability <ol style="list-style-type: none"> 1. Author 2. References 3. Information agrees with more sites 4. Information agrees with prior knowledge 5. Organization C. Reliability <ol style="list-style-type: none"> 1. Kind of information 2. Objectivity 3. Primary/secondary 4. Goal 	<p>In what language is the information written? Are there many grammatical or type errors? Is it filled with domain specific language?</p> <p>Does the information answer (part of) the information problem?</p> <p>Is the information aimed at a specific group of readers?</p> <p>Is the information up to date?</p> <p>Is there enough information on the page? Or only a part of the information I'm looking for?</p> <p>Who has written the information? Can I contact him/her?</p> <p>Are there references on the page to used sources? Or links to more websites on the same subject?</p> <p>Can I find the same information on more pages or is this information only available on this site?</p> <p>Does the information confirm what I already know?</p> <p>Which organization is behind the information. A governmental organization, or a health organization? Can I find their logo on this site?</p> <p>What kind of information is it? A newspaper article or a forum? Is it an opinion or results from research?</p> <p>Is the information objective or coloured by a certain point of view? Are there a lot of advertisements on the page?</p> <p>Is the information first hand or is it someone telling about someone who did something?</p> <p>What does the (author of) information want to achieve. Sell something? Convince me of something or just inform me?</p>
Evaluate source	<ol style="list-style-type: none"> A. Technical <ol style="list-style-type: none"> 1. Appearance 2. Speed B. Usability <ol style="list-style-type: none"> 1. Language 2. Connection to task 3. Audience 4. Topicality C. Verifiability <ol style="list-style-type: none"> 1. Reputation D. Reliability <ol style="list-style-type: none"> 1. Kind (site/PDF) 2. Primary/secondary 	<p>Does the site appeal to me? Does it have pictures or only text?</p> <p>Does it take a lot of time to load the page?</p> <p>In what language is the site written? Are there many grammatical or type errors?</p> <p>Does the site have a connection to (part of) my information problem?</p> <p>For whom is the site meant? Who are its visitors?</p> <p>Is the site updated regularly?</p> <p>Is this site famous or infamous? Does it have a good/bad reputation?</p> <p>What kind of source is it, a website, a word document, a PDF file?</p> <p>Is the site an original source or a site telling about what is written somewhere else?</p>

References

- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune and Stratton.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York: Holt, Rinehart and Winston.
- Ausubel, D. P., Novak, J. D., & Hanesian, H. (1978). *Educational psychology: A cognitive view* (2nd ed.). New York: Holt, Rinehart and Winston.
- Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637.
- Beljaarts, M. (2006). *Internet, een populair medium voor het zoeken van informatie bij schoolopdrachten* [Internet, a popular medium for searching information for school assignments]. Malmberg: 's-Hertogenbosch.
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21, 487–508.
- Brand-Gruwel, S., & Wopereis, I. (2006). Integration of the information problem-solving skill in an educational programme: The effects of learning with authentic tasks. *Technology, Instruction, Cognition and Learning*, 4, 243–263.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2005). The relationship between Internet-specific epistemological beliefs and learning within Internet technologies. *Journal of Educational Computing Research*, 33, 141–171.
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20, 485–522.
- Buzan, T. (1995). *The mind map book* (2nd ed.). London: BBC Books.
- Carraher, D., & Schliemann, A. D. (2002). The transfer dilemma. *The Journal of the Learning Sciences*, 11(1), 1–24.
- Deketelaere, A., & Kelchtermans, G. (1996). Collaborative curriculum development: An encounter of different professional knowledge systems. *Teachers and Teaching: Theory and Practice*, 2, 71–86.
- Duijkers, H. M., Gulikers-Dinjens, M. T. H., & Boshuizen, H. P. A. (2001). Begeleiden van leerlingen bij het zoeken, selecteren en beoordelen van informatie. [Supporting students when searching, selecting and judging information]. In J. Ahlers, T. Hoogbergen, P. Leenheer, & J. de Vos (Eds.), *Handboek Studiehuis Tweede Fase*. Samsom: Alphen a/d Rijn.
- Fidel, R., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J., et al. (1999). A visit to the information mall: Web searching behavior of high school students. *Journal of the American Society of Information Science*, 50(1), 24–37.
- Gerjets, P., & Hellenthal-Schorr, T. (2008). Competent information search in the World Wide Web: Development and evaluation of a web training for pupils. *Computers in Human Behavior*, 24, 693–715.
- Gick, M., & Holyoak, K. (1983). Schema-induction and analogical transfer. *Cognitive Psychology*, 12, 1–38.
- Graesser, A. C., Wiley, J., Goldman, S. R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition Learning*, 2, 89–105.
- Grimes, D. J., & Boening, C. H. (2001). Worries with the web: A look at student use of web resources. *College and Research Libraries*, 62(1), 11–23.
- Hirsch, S. G. (1999). Children's relevance criteria and information seeking on electronic resources. *Journal of the American Society for Information Science*, 50, 1265–1283.
- Jones, S. (2002). *The internet goes to college. How students are living in the future with today's technology*. Washington, DC: Pew Internet and American Life Project.
- Kafai, Y., & Bates, M. J. (1997). Internet web-searching instruction in the elementary classroom: Building a foundation for information literacy. *School Library Media Quarterly*, 25, 103–111.
- Koot, U., & Hoveijn, M. (2005). Google, de magische antwoordmachine. [Google, the magical machine with answers]. *Pedagogiek in Praktijk*, 11(28), 18–22.
- Kuiper, E., Volman, M., & Terwel, J. (2008). Integrating critical web skills and content knowledge: Development and evaluation of a 5th grade educational program. *Computers in Human Behavior*, 24, 666–692.
- Larkin, J., McDermott, J., Simoa, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335–1342.
- Lorenzen, M. (2002). The land of confusion? – High school students and their use of the World Wide Web for research. *Research Strategies*, 18, 151–163. Available from: <<http://www.libraryinstruction.com/confusion.html>>.
- Lyons, D., Hoffman, J., Krajcik, J., & Soloway, E. (1997). *An investigation of the use of the World Wide Web for on-line inquiry in a science classroom*. Chicago, IL: Paper presented at the meeting of the National Association for Research in Science Teaching.
- 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, 155–172.
- Mason, L., & Boldrin, A. (2008). Epistemic metacognition in the context of information searching on the web. In M. S. Khine (Ed.), *Knowing, knowledge and beliefs: Epistemological studies across diverse cultures* (pp. 377–404). New York: Springer.
- Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student web use, perceptions of information credibility, and verification behavior. *Computers & Education*, 41, 271–290.
- Nokes, T. J. (2009). Mechanisms of knowledge transfer. *Thinking & Reasoning*, 15(1), 1–36.
- Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27, 937–949.
- Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18, 16–25.
- Rothenberg, D. (1998). How the web destroys student research papers. *Education Digest*, 63, 59–61.
- Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, 24, 113–142.
- Simons, R. J., Van der Linden, J., & Duffy, T. (2000). New learning: Three ways to learn in a new balance. In R. J. Simons, J. Van der Linden, & T. Duffy (Eds.), *New learning*. Dordrecht: Kluwer. pp. 1–20.
- Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. London: Harvard University Press.
- Stadler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool metaaware on the web search of laypersons. *Computers in Human Behavior*, 24, 716–737.
- Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 247–261.
- Van Gog, T., Paas, F., Van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11, 237–244.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills*. Englewood Cliffs, NJ: Educational Technology.
- Wagner, J. F. (2006). Transfer in pieces. *Cognition and Instruction*, 24(1), 1–71.
- Wallace, R. M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences*, 9(1), 75–104.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2008). Information-problem solving: A review of problems students encounter and instructional solutions. *Computers in Human Behavior*, 24, 623–648.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers and Education*, 25, 234–246.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (submitted for publication). Teachers and the World Wide Web: How teachers evaluate search results, information and source.
- Wertheimer, M. (1961). *Productive thinking*. London: Tavistock.