Becoming a critical websearcher

Effects of instruction to foster transfer

OpenUniversiteitNederland

The research reported here was carried out at the

OpenUniversiteitNederland

was funded by:

NWO
Netherlands Organisation for Scientific Research

Project no. 411-03-106

In the context of the research school:

ICO
Interuniversity Centre for Educational Research

ISBN: 978-90-79447-13-8
© Amber Walraven, Arnhem, the Netherlands, 2008
Cover design: Guus Paul Mast, GPM Design
Printed by VDA-groep
All rights reserved

Becoming a critical websearcher

Effects of instruction to foster transfer

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit Nederland
op gezag van de rector magnificus
prof. dr. ir. F. Mulder
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 19 december 2008 te Heerlen
om 15:30 precies

door

Amber Walraven

geboren op 9 maart 1980
te Schiedam

Promotor:
Prof. dr. H.P.A. Boshuizen

Co-promotor:
Dr. F.L.J.M. Brand-Gruwel

Overige leden beoordelingscommissie:
Prof. dr. J.J. Beishuizen, Vrije Universiteit Amsterdam
Prof. dr. J.H. van Driel, Universiteit Leiden/ ICLON
Prof. dr. B.H.A.M. van Hout-Wolters, Universiteit van Amsterdam
Prof. dr. A.J. Mooij, Open Universiteit Nederland
Dr. A.W. Lazonder, Universiteit Twente

Voorwoord

Van alles wat ik in de afgelopen 4 jaar heb moeten schrijven, vind ik dit voorwoord toch wel een van de moeilijkste. In dit voorwoord kan ik iedereen die mij de afgelopen tijd heeft gesteund bij de tot standkoming van dit proefschrift, en iedereen die me dierbaar is, bedanken. En nu ben ik zo bang dat ik mensen ga vergeten. Daarom nu alvast mijn welgemeende excuses als ik jou, lezer van dit voorwoord, niet bij naam noem en je toch meent dat je genoemd had moeten worden!

Als eerste wil ik heel graag Saskia Brand-Gruwel bedanken, die als mijn dagelijkse begeleider er voor heeft gezorgd dat ik met veel plezier werkte aan mijn proefschrift, dat ik scherp en gemotiveerd bleef, en die me het gevoel gaf alles bij haar kwijt te kunnen. Saskia, ik heb de samenwerking met jou enorm gewaardeerd. Je liet me mijn eigen gang gaan, maar was streng genoeg op de momenten dat ik dat nodig had. Je maakte altijd tijd vrij en hield je altijd aan je woord. Ik hoop dat we nog regelmatig met elkaar zullen samenwerken. Ook mijn promotor Els Boshuizen wil ik graag op deze plaats bedanken. Tijdens onze gesprekken was je in staat om mijn blik eens op andere kanten van de zaak te richten. En hoewel ik daar destijds niet altijd blij mee was, besef ik nu hoe waardevol die momenten waren. Het heeft er voor gezorgd dat de hoofdstukken van het proefschrift een sterke samenhang hebben en het heeft mij geleerd met een bredere blik naar mijn eigen onderzoek en functioneren te kijken. Ik heb het ook gewaardeerd dat je enkele wijze woorden tot me richtte toen ik vertrok uit Heerlen. Ik denk dat we even aan elkaar hebben moeten wennen, maar dat we uiteindelijk een goed team vormden.

Mijn onderzoek had nooit plaats kunnen vinden zonder de tomeloze inzet van 'mijn' docenten: Jannie Lina, Erwin Reinders, Willem Ubaghs en Nico Zijlstra. Ik wil jullie nogmaals van harte bedanken voor de tijd en energie die jullie in het project hebben gestoken. Ik heb onze samenwerking altijd leerzaam gevonden, en de etentjes natuurlijk erg gezellig. Jullie zijn het soort docenten waar elke school een heel leger van moet hebben! Naast de docenten, wil ik ook de directie en leerlingen van het Sintermeertencollege te Heerlen, het Stella Mariscollege te Meerssen en het Grotiuscollege te Heerlen bedanken voor hun medewerking. In het bijzonder dhr. Meens, dhr van Loo en mevr. Sieben.

Naast werk was er gelukkig ook een hoop tijd voor gezelligheid. Mijn mede aio's (sommigen zijn inmiddels geen aio's meer, zo gaat dat) Fleurie, Danny, Helen, Wendy, Monique, Ludo, Femke, Sandra, Marjo, Greet, Gemma, Karen, Tamara, Ingrid, Bettine, Chantal, Pieter, Judith en Liesbeth: jullie waren altijd in voor een gesprekje, sinterklaasavondje, etentje of een workshop theatersport. Ik heb me als niet-Limburger soms (of vaak...) schuldig gemaakt aan opmerkingen over Limburgers die ik hier niet zal herhalen (Fleurie kan ze stuk voor stuk vast nog wel opnoemen), maar ik moet toegeven: Limburgers en bewoners van Limburg zijn gezellig en ik zal jullie missen! Naast de aio's wil ik nog Olga bedanken voor de fijne samenwerking, Ellen voor het buurvrouwschap en Iwan voor gedeelde humor. Het bestuur van het VPO, in het

bijzonder Roeland, Reyn, Ellen en Femke dank ik voor de gezellige tijd en goede samenwerking. En ik noem ze niet allemaal bij naam, maar alle medewerkers van de MAL en van CELSTEC (voorheen OTEC) als mede mijn nieuwe collega's bij de vakgroep Curriculumontwerp en Onderwijsinnovatie bij de Universiteit Twente, bedankt voor jullie getoonde interesse en goed collegaschap. Op de Universiteit Twente is mijn fascinatie voor onderzoek doen ook begonnen en in dat kader wil ik Sarah Manlove, Ard Lazonder en Ton de Jong nogmaals bedanken.

Dan zijn er natuurlijk ook nog enkele mensen in mijn privéleven die ik moet bedanken. Als eerste Johan, Hennie, Truus, Richard, Joke en Ruben. Jullie zijn altijd belangstellend geweest en ik vermoed ook wel een beetje trots. Ik wil jullie bedanken voor de belangstelling en het aanhoren van mijn verhalen. Ik durf niet te garanderen dat die eindeloze verhalen met het tot stand komen van mijn proefschrift nu voorbij zijn...ze zullen wellicht ergens anders over gaan.....

Ook mijn eigen familie, opa, oma, ooms en tantes wil ik bedanken voor belangstelling en steun. Guus wil ik bedanken voor zijn mooie ontwerp voor de kaft van het proefschrift.

Rogier, je bent mijn broer en ik hou van je, wat er ook gebeurt.

Lieve papa en mama, ik hoefde niet lang te denken wie ik als paranimfen achter me wilde hebben staan tijdens de promotie. Jullie staan heel mijn leven al achter me en steunen me in alles. Ik lijk op jullie alletwee en daar ben ik heel erg trots op. Jullie hebben me de kans gegeven te doen wat ik wilde doen en stonden altijd met raad en daad klaar. Het is niet voor niets dat ik nog graag bij jullie langskom en regelmatig jullie mening vraag. Ik ben er trots op jullie dochter te zijn!

Als laatste wil ik Chris bedanken. Dankzij jou heb ik het kunnen volhouden om te wonen in Arnhem en te werken in Heerlen. Jij zorgde ervoor dat ik 's avonds moe maar blij naar huis ging, omdat ik wist dat jij daar zou zijn. (Het hielp natuurlijk ook een klein beetje dat je bijna elke avond kookte...). We kunnen over alles samen praten en jij bent de enige die door mijn ochtendhumeur heen kan kijken, en op mijn beurt hou ik van jou precies zoals je bent; ontzettend lief, erg enthousiast en slechts af en toe een tikkie eigenwijs! We hebben het ontzettend goed samen en ik hoop dat we nog heel lang samen zullen zijn.

Amber Walraven
Arnhem, 2008

Contents

General Introduction 15

Information-Problem Solving: A Review of Problems Students Encounter and Instructional Solutions 19

Introduction 20

Method 23

Selection 23

Analysis system 24

Results 24

Problems people encounter when solving information problems 24

Define information problem 24

Search information 25

Scan information 27

Process information 28

Organize and present information 29

Regulation 29

Summary 30

Instructional solutions 31

Instruction and support for young children 31

Instruction and support for teenagers 34

Instruction and support for adults 36

Summary 38

Discussion 40

How students evaluate sources and information when searching the World Wide Web for information 43

Introduction 44

Method 46

Participants 46

Material 47

Tasks 47

WWW knowledge Questionnaire 49

Post hoc group interviews 49

Procedure 49

Data analysis 51

Results 53

Students' information problem solving processes 54

Students' evaluating processes 56

Evaluation criteria 57
Criteria for evaluating: domain specific 58
Students' products 59
Students' awareness of evaluation criteria 60
Discussion 63
Teachers and the World Wide Web: How teachers evaluate search results, information and source 67
Introduction 68
Method 73
Participants 73
Material 73
Information-problems 73
WWW Questionnaire 74
Post hoc group interviews 74
Procedure 75
Data analysis 75
Results 78
Teachers' evaluations of search results, information and sources 78
Criteria for evaluating: domain specificity 80
Qualitative analysis 81
Teachers' task performance 82
Teachers' awareness of evaluation criteria 83
Discussion 85
Fostering transfer of websearchers' evaluation skills: a field test of two transfer theories 89
Introduction 90
Transfer of training 90
Evaluation of search results, information and source on the Internet 91
Instruction in Web searching skills 92
Research questions 96
Method 96
Participants 96
Materials 97
The educational programs 97
High road program 97
Rich representation program 100
Measurements 101
Design and procedure 102
Data analysis 103
Results 104
Evaluation tasks hit list and websites 104
Effects of the instruction 105

Transfer effects of instruction *105*
Think aloud protocols 106
Task performance 108
Field notes 109
Discussion 110
Fostering students' evaluation behaviour while searching the Internet: Using a Design-Based Research methodology 115
Introduction 116
Instruction in evaluating information while searching on Internet 116
Design-Based Research 118
Design Research: the 'evaluation of Internet information'-program 119
Design team *119*
Analysing practical problems *119*
Designing instruction using design principles *120*
Testing the design *121*
Redesign of the program *121*
Testing the redesigned program *122*
Method 122
Participants 122
Program 123
Goal and overview of the lessons *123*
Reader *124*
Process worksheets *124*
Discussions *126*
Measurements 126
Evaluation hit list *126*
Evaluation of websites and information *127*
Learning results *127*
Field notes *127*
Reflective stories *128*
Design and Procedure 128
Data analysis 128
Results 129
Evaluation tasks hit list and websites 129
Effects of the instruction *129*
Transfer effect of instruction *129*
Learning results 130
Field notes 130
Reflective stories 132
Discussion 134
General Discussion 137
Main findings and conclusion 138
Practical implications 142

Future research 143
References 145
Summary 151
Samenvatting 157
Appendix: Evaluation criteria 165
ICO dissertation series 167

CHAPTER 1

GENERAL INTRODUCTION

General Introduction

Teenagers nowadays spend more time online than they watch television. The World Wide Web (WWW) is their playground. They use it to communicate with friends, watch movies, download music and play games. Since they use the Web so frequently in everyday life, it is not surprisingly that students rely on the Web for educational tasks, such as writing essays and preparing for presentations, as well. Using the Web for educational purposes requires that they identify their information needs, locate information sources, extract and organize information from each source, and synthesize information from a variety of sources. This set of activities is frequently defined as Information-Problem Solving (IPS) (Brand-Gruwel, Wopereis, & Vermetten, 2005).

In the process of solving information-problems the student's ability to evaluate the hit list and the information on websites including the website itself plays a critical role. However, the way students evaluate is far from ideal, because it is not always done based on clear criteria but on intuition (Koot & Hoveijn, 2005). Young children tend to believe that everything on the Web is true and teenagers use information that can solve their information-problem without thinking about the purpose of a site (Fidel, Davies, Douglass, Holder, Hopkins, Kushner, et al., 1999). They also find it hard to express how they evaluate and select information (Lorenzen, 2002), while some students admit that they do not check information at all. A non-critical attitude towards information on the WWW can seduce students to cut and paste information for accomplishing assignments without evaluating it, which results in reports and learning that lack quality (Britt & Aglinskas, 2002).

The importance of instruction in information-problem solving skills and in the critical use of the WWW has been recognized by teachers. This instruction should not only address the evaluation skills to evaluate search results (hit list), information on a website and source (the website itself), but should also focus on the adaptation of these skills to new situations and changed tasks (i.e., transfer), because failing this last quality will sooner or later lead to skills obsolescence. Different theories could be used to design instruction that fosters transfer of evaluation skills. One of these theories is the high road theory of Salomon and Perkins (1989), which 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 flexibly in different situations. Another transfer theory is the rich representation theory of Simons, van der Linden and Duffy (2000), which emphasizes the importance of a good, extensive and well organised knowledge base and the domain specific interpretation of the skills.

The role of the teacher is of importance when implementing instruction to foster students' evaluation skills. Teachers should function as a role model in the correct use of the WWW when accomplishing learning assignments and should stimulate students to use evaluation skills. But before teachers can help their students to become critical websearchers, they should have knowledge and skills on how to use the WWW and evaluate results, information and source themselves. They should be aware of the different criteria when evaluating and should have insight in which criteria can be used in different situations. Unfortunately, not all teachers are used to work with the WWW.

The aims of this research are gaining insight in the evaluation skills used by students and teachers and to design, implement and evaluate instruction to foster the transfer of these skills to various settings. Research questions were: 1) which evaluation criteria do

students and teachers of secondary education use for evaluating search results, information and source and do the criteria used differ when solving tasks in different domains, 2) can instruction designed according to the high road transfer theory and the rich representation theory foster the transfer of students' evaluation skills?

These questions were addressed in five studies. One review study, two experimental studies to determine the criteria students and teachers use to evaluate results, information and source on the WWW in different domains, and two experimental studies to examine the (transfer) effect of designed instruction were conducted.

Chapter 2 describes the review study. The goal of this study was to determine what kinds of problems children, teenagers and adults experience when solving information-problems on the WWW, and what kind of instructional support can help to solve these problems. Fifteen articles concerning problems children, teenagers and adults experience when searching for information on the WWW were found after a systematic search in PsycINFO and ERIC, and using references of the articles found. The articles were analyzed using the decomposition of the information-problem solving skill developed by Brand-Gruwel et al. (2005) as a framework. Furthermore, twelve studies regarding instructional support to foster students' information-problem solving skills were found. They were analysed on the effectiveness of instructional measures and categorised by the age of the participants.

Chapter 3 presents the results of the first empirical study, aiming at determining the evaluation criteria 9th grade students in secondary education use to accomplish assignments. Twenty-three students solved two tasks from different domains while thinking aloud. They were asked to answer the question posed in the task by selecting information from the WWW and pasting this into a word document. The protocols were transcribed verbatim and using a coding scheme based on a scheme used by Brand-Gruwel et al. (2005) evaluation criteria were scored. Task performance was measured by three criteria: answer contains information of more than one source, quality of the sources and suitability of information used. A day after the students individually completed the tasks, they returned in groups of three. These focus groups were interviewed to obtain students' knowledge and conceptions on criteria for evaluating information and source.

The goal of the study described in Chapter 4 was to uncover criteria teachers use when searching the WWW and the possible influence of domain knowledge on their evaluation skills. Eleven teachers solved two information problems while thinking aloud, one in the domain they taught, and one in a different domain. This means that they were domain experts while accomplishing one assignment and non-experts while solving the other assignment. All think-aloud protocols were transcribed verbatim. Again, the criteria used to evaluate results, information and source were coded. Furthermore, to gain more insight in the use of prior knowledge when evaluating information, sites or hit lists, utterances showing use of domain knowledge were categorized based on the goal of the utterance: 1) activate prior knowledge, 2) evaluate results, information or source, or 3) to make a decision upon the search strategy. The performance was measured in the same way as in the previous study. A day after teachers individually completed the tasks, they returned in groups of three. These focus groups revealed teachers' knowledge and conceptions on criteria for

evaluating information and its source in the two different domains. Results shed light on how teachers evaluate and how they think domain knowledge influences the search process. Results of the studies described in Chapter 3 and 4 are used to design instruction to foster students' evaluation skills and use of evaluation criteria.

In Chapter 5 a study is reported in which two instructional programs were designed to foster 9th grade students' evaluation skills, as well as the transfer of these skills to another domain. The first program was designed using the transfer theory of Salomon and Perkins (high road) as a starting point, and for the design of the second program the transfer theory of Simons, van der Linden and Duffy (rich representation) was used. Both programs were developed together with four secondary education history teachers, using a design-based research approach. The same teachers also implemented the programs. Two 9th grade classes received the high road program; two other classes participated in the rich representation program. Three lessons were observed in every class. Effects of the programs on students' use of criteria for evaluating search results, information and source were measured with two paper and pencil tests, consisting of a hit list and a website evaluation task. One test was in the domain of instruction (history) and one in the transfer domain (biology). Students received a manufactured hit list on paper and had to select three sites they would open and three sites they would not open. Students also received a booklet of eight printed out websites and were asked which sites and what information they would or would not use. These paper and pencil tests came close to reality while enabling us to test a large group of students at the same time. Furthermore, a small group of students thought aloud while accomplishing two tasks; one task in the domain of instruction (history) and one in the transfer domain (biology). Results regarding the effects of the programs and the added value of designing instruction according to different transfer theories are discussed.

In Chapter 6 a study is described in which the best of both programs, as studied in the previous experiment, is combined in the design of a new program. The same four history teachers designed and implemented this new program. The goal of this program was to teach students how to evaluate results, information and source and fostering transfer of these evaluation skills to other domains. This new program was tested with four experimental classes and results were compared to a control class. Effects of the program on students' use of criteria while evaluating search results, information and source were again measured using two paper and pencil tests (history and biology). Furthermore, because the instruction of the evaluation skills was embedded in history lessons, also students' knowledge on the subject matter (Second World War) involved was measured after the intervention. Not only the results of this experiment are described in this chapter, but also the teachers' experiences with Design-Based Research (DBR) will be reported.

Main conclusions and findings are presented in Chapter 7, as well as some practical implications and suggestions for future research.

CHAPTER 2

INFORMATION-PROBLEM SOLVING

A REVIEW OF PROBLEMS STUDENTS ENCOUNTER AND INSTRUCTIONAL SOLUTIONS

Information-Problem Solving: A Review of Problems Students Encounter and Instructional Solutions

Searching and processing information is a complex cognitive process that requires students to identify information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources. This process is called information-problem solving (IPS). IPS can be characterized as a complex cognitive skill, which may need direct instruction to reach high levels of performance. However, IPS has been given little attention in schools, and instruction in this skill is rarely embedded in curricula. And yet, by giving students assignments in which students have to solve an information-problem, teachers assume that their pupils have developed this skill naturally. A literature study was done to determine what kinds of problems students experience when solving information problems using the WWW for searching information, and what kind of instructional support can help to solve these problems. Results show that children, teenagers and adults have trouble with specifying search terms, evaluating search results and evaluating information and source. Regulating the search process is also problematic. Instruction designed specifically for IPS using the WWW for searching information is rare but indeed addresses the problematic skills. However, there are differences between various methods and it is unclear which method is most effective for specific age groups.

This chapter is based on 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

Introduction

Our current society is an information society. Recent technological developments such as mobile communication, GPS, and the Internet provide us with large bodies of information every day. It is up to us to decide what to do with all this information. One could decide to ignore it altogether, but this will almost certainly result in alienation from society. A more fruitful approach might be to regularly gain access to new information. This approach requires people to identify their information needs, locate corresponding information sources, extract and organize relevant information from each source, and synthesize information from a variety of sources into cogent, productive uses (Bawden, 2001; Brand-Gruwel et al., 2005; Eisenberg & Berkowitz, 1990, 1992; Marchionini, 1999; Moore, 1995, 1997; Shapiro & Hughes, 1996; Spitzer, 2000). Together these activities constitute a process that we refer to as information-problem solving (IPS).

IPS is not only important in everyday life, but also in education. In the last decade educational systems have undergone changes. Instead of a system aiming at the reproduction of knowledge, new learning is aiming at learning outcomes that are durable, flexible, functional, meaningful, and applicable. Active pedagogical methods, in which students learn by doing instead of listening and in which the teacher has a guiding role, fit this new learning (Simons et al., 2000). Students are given assignments, such as writing an essay on Tibet, that require them to search for information. And although there are many other sources to search for information about Tibet (e.g., the library or an encyclopaedia) the most probable source a student would use nowadays is the World Wide Web (WWW) (Beljaarts, 2006). Assignments like the one mentioned above are common; children in the early elementary grades are already confronted with it. However, little is known about how children, teenagers, and adults become proficient in solving information-problems. It seems to be assumed that the IPS-skill develops spontaneously, that means, the skill receives little attention in schools and instruction is rarely embedded in curricula. It is, however, rather unlikely that students spontaneously develop a full-fletched IPS-skill. Research has shown that children, adolescents and adults have problems with IPS (Brand-Gruwel et al., 2005; Duijkers, Gulikers-Dinjens, & Boshuizen, 2001; Hirsch, 1999; Kafai & Bates, 1997; MaKinster, Beghetto, & Plucker, 2002; Monereo, Fuentes, & Sànchez, 2000). Different age groups may, however, have different problems with IPS and hence benefit from different kinds of support. The goal of the study in this chapter is to give an overview of the problems people of different ages encounter with solving information-problems. After specifying these problems, a review of research addressing instructional methods for IPS is presented. Especially we want to answer the question whether and how these instructional methods foster problems students experience.

In this study the 'Information Problem Solving while using Internet'-skill decomposition developed by Brand-Gruwel and Wopereis (2006) is used as an analytical framework. This decomposition, visualized in Figure 1, is based on empirical research findings regarding the IPS- process of students who searched for information on the WWW and has been developed to support instructional designers. The skill

decomposition defines IPS as consisting of five constituent skills: define information problem, search information, scan information, process information, and organize and present information. As can be seen, these constituent skills can be divided into sub skills. Furthermore, regulation is considered an important aspect in the entire search process. As the WWW is a comprehensive source of information, searching the WWW relies on peoples' regulative abilities. Regulatory aspects such as orientation, monitoring and steering play a key role in this process (Boekhorst, 2003; Brand-Gruwel et al., 2005; Hill, 1999; Lazonder, 2003).

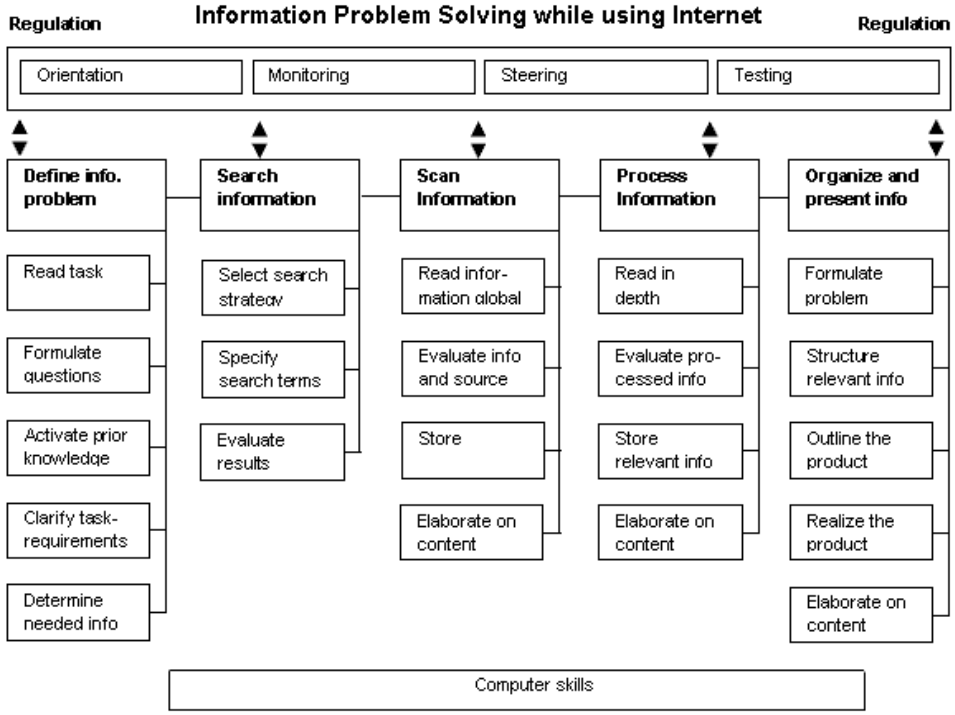


Figure 1 The information-problem solving skill decomposition (based on Brand-Gruwel et al., 2005)

Imagine a 16-year old student, Rita. Rita has been given a very open assignment of writing an essay on Tibet. After reading this task, Rita does not have a fully defined information-problem yet. *Defining the information problem* is important in order to get a clear insight into the problem (e.g., Hill, 1999; Land & Greene, 2000; Moore, 1995). Rita asks herself the following questions: What should be the focus of the essay (e.g., government, nature, population or religion)? What questions must be answered? What knowledge do I already have on one (or all) of these subjects? This activation of prior knowledge helps Rita to integrate the new information found with old, known, information (Brand-Gruwel et al., 2005; Hill, 1999; Moore, 1995). Next, Rita should also pay attention to the task requirements, for instance is there a minimum or maximum number of pages specified for the essay, and what is the target audience? Once Rita

has established all this, she can derive which information is needed and can start her search.

Rita has decided to write an essay on the government of Tibet, and enters the terms 'government + Tibet' in Google™. In the first three results she comes across a website by the Chinese government (http://www.chinese-embassy.org.uk/eng/zt/zgxz/default.htm) and two sites claiming to be the official website of the government in exile (http://www.tibet.net and http://www.tibet.com). Based on these results, she decides to open the second site, http://www.tibet.net. From previous experience she has learned that sites with a .com address are often commercial sites. She therefore expects the quality and reliability of the .net site to be higher. Quality and reliability are in this case criteria used to evaluate the found sources. During this *search for information* Rita has selected a search strategy (using a search engine), specified search terms and evaluated results. Computer skills like using a mouse and keyboard are also important in this part of the process (Brand-Gruwel et al., 2005; Marchionini, 1995; Sutcliffe & Ennis, 1998). After opening the site http://www.tibet.net Rita discovers that this site is owned by the government in exile and the information is up to date. This site is useful for the essay and she copies some information in her own file. However, she decides that she needs more information on the Chinese view of the facts (Tibet has been occupied by China). She can take a look at her first result (http://www.chinese-embassy.org.uk/eng/zt/zgxz/default.htm) or do another search with keywords like Tibet + China. Rita has used the sub skills read information global, evaluate information and source, store relevant information and elaborate on content during this *scanning of information.*

After viewing the website http://www.chinese-embassy.org.uk/eng/zt/zgxz/default.htm for information, Rita finds that this is the website of the Chinese embassy in the United Kingdom. She decides that this is not very useful and searches Google with Tibet + China and opens the website www.tibet-china.org/indexE.html. This site gives a historic overview of Tibet and its relation with China, from Chinese perspective. Rita reads this site very carefully, and selects information that she can use in her essay. Reading the site carefully is part of the constituent skill *process information.* The goal is to reach a deep understanding of the information (Dochy, 1993; Schmeck & Geisler-Brenstein, 1989) and reaching an integration of the different pieces of information found and relevant prior knowledge so that the information-problem can be solved (Wopereis, Brand-Gruwel, & Vermetten, 2008).

Making the product as required in the task is the goal of the constituent skill *organize and present information.* Several products are possible: a presentation or a poster, or, as in our example, a text document such as an essay. For every type of product, it is important to formulate the problem. The layout must be determined and the components defined in this outline further structured and filled in. While organizing and presenting information elaboration remains important (Wopereis et al., 2008). Rita has found sufficient information to write her essay. First, she determines what the line of reasoning will be and structures the information found according to this line of reasoning.

As can be seen in the skill decomposition *regulation* activities will be carried out during the entire IPS-process. Rita for example, was regulating when she decided she needed more information on the Chinese view. She compared the information found with her problem definition and decided that it was not enough to solve her information-problem. Regulation is related to the effectiveness and efficiency of the entire process (Hill, 1999; Hill & Hannafin, 1997; Land & Greene, 2000; Lazonder, 2003; Marchionini, 1995).

Rita's IPS behaviour we described here is very sophisticated. She has learned to execute all constituent and sub skills. By using Rita as an example we described an ideal rather than a real student. Research suggests that at least some skills are problematic for real students (e.g., Brand-Gruwel et al., 2005; Duijkers et al., 2001; Hirsch, 1999; Kafai & Bates, 1997; MaKinster et al., 2002; Monereo et al., 2000), but some years ago Rita herself might have had trouble with some sub skills too.

The skill decomposition will be used to categorize the problems people have with IPS. These problems will be categorized for young children (age 6-12), teenagers (13-18) and adults (18 and older). Then, instructional solutions will be described in terms of their focus (i.e., the skills involved) and also the underlying didactical principles. The research questions addressed in this study are:

1) When people (of three age groups, 6-12, 13-18, 18+) experience problems with information-problem solving, under which constituent or sub skill of the complex cognitive IPS-skill can these problems be placed?
2) What is the effect of different kinds of IPS instruction or support offered to these age groups and can instructional guidelines be deducted?

Method

Selection

In order to find information on the difficulties people experience when solving information- problems, PsycINFO and ERIC were searched with combinations of the following keywords: information (problems, skills, seeking, searching, literacy), WWW and Internet. The references of the articles found were used to search for new articles and books. Only articles in which an overview was given of problem areas and skills mastered by students were included in the overview.

Keywords used for gathering information about training, instruction or interventions concerning information-problem solving were combinations of: information skills, instruction, education, information-problem solving, WWW, Internet. References were used for searching additional literature on this topic.

Articles dated before 1995 were excluded from the analysis because the rise of the World Wide Web started in 1995.

Analysis system

The IPS-skill decomposition by Brand-Gruwel and Wopereis (2006) was used as analytical framework. Studies concerning problems students experience while solving information-problems were categorized according to the constituent skills these problems pertain to. The studies on instructional methods were categorized by the addressed age group.

Results

Problems people encounter when solving information problems

The literature search resulted in fifteen studies. Table 1 provides an overview of the studies found.

Define information problem

The skill 'define information problem' is rarely included in information-problem solving research. Only one study explicitly addressed this constituent skill (Brand-Gruwel et al., 2005), and three studies did not (Koot & Hoveijn, 2005; Lorenzen, 2002; Rosell-Aguilar, 2004). Other studies mentioned the task students had to solve and made some comments on the problem definition, but in these studies the focus was on the search itself (Bilal, 2000; Duijkers et al., 2001; Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Large & Beheshti, 2000; Lyons, Hoffman, Krajcik & Soloway, 1997; Makinster et al., 2002; Monereo et al., 2000; Schacter, Chung & Dorr, 1998; Wallace, Kupperman, Krajcik & Soloway, 2000). Although defining the information problem is not the focus of these studies, some conclusions regarding this constituent skill can be drawn from their results. The sub skill 'read task' does not cause problems in any of the three age groups. Students in all of the studies seemed to understand the task and knew what was expected of them.

Brand-Gruwel et al. (2005) found that adults are capable of 'formulating questions' to define the problem. Teenagers on the other hand find formulating questions difficult (Lyons et al., 1997; Wallace, Kupperman et al., 2000). When teenagers had to search for information on the World Wide Web about a subject matter to accomplish a task, they had trouble with formulating useful inquiry questions. They often asked questions with a single correct answer instead of questions that required them to synthesize information from multiple sources. They asked a somewhat general question and tried to find information on it. When they could not find information to answer their question they simply changed the question. They adapted the question to available information found online and had troubles with posing good and rich questions.

'Activating prior knowledge', 'clarifying task requirements' and 'determining needed information' is also difficult for teenagers. Most teenagers start searching immediately without exploring the topic, planning the search or thinking about the task (Duijkers et al., 2001; Fidel et al., 1999; Lyons et al., 1997). Young children do not focus on

the task either (Bilal, 2000), while adults do seem to activate prior knowledge (Brand-Gruwel et al., 2005; Monereo et al., 2000).

Table 1 Studies on problems people encounter while solving information-problems

Bilal (2000)
Brand-Gruwel et al. (2005)
Duijkers et al. (2001)
Fidel et al. (1999)
Hirsch (1999)
Kafai and Bates (1997)
Koot and Hoveijn (2005)
Large and Beheshti (2000)
Lorenzen (2000)
Lyons et al. (1997)
MaKinster et al. (2002)
Monereo et al. (2000)
Rosell-Aguilar (2004)
Schacter et al. (1998)
Wallace, Kupperman, et al. (2000)

With regard to 'clarifying task requirements' something remarkable happened in the study by Wallace, Kupperman et al. (2000). Children seemed to entertain extra goals during their search that were not specified in the task. For instance, they tried to find the perfect webpage, to get a limited number of search results and find a ready-made answer to their question.

To conclude, it appears that adults do not have trouble with the constituent skill 'defining the information problem'. Teenagers have trouble with 'formulating questions', 'activating prior knowledge', 'clarifying task requirements' and 'determining needed info'. Little is known about young children and their problems with this constituent skill, but based on the problems teenagers have, we assume that the same problems occur with younger children.

Search information

All studies in this review address this part of the IPS-process. 'Searching for information' on the Web can be done in several ways. The three most common strategies are using search engines, entering URLs, and browsing subject categories. Young children are capable of browsing and following bookmarks; the other strategies are too difficult for them. Entering URLs becomes less problematic from the age of eight (Kafai & Bates, 1997). From the age of ten people are capable of using all strategies (Bilal, 2000; Brand-Gruwel, et al., 2005; Fidel et al., 1999; Kafai & Bates, 1997; Schacter et al., 1998).

The choice for a specific strategy depends on the problem at hand. Young children browse when the task is ill-defined (e.g., 'What should be done to reduce crime in California?'), and use a search engine with well-defined tasks (e.g., 'What are the three types of crime that happen most in California?'), although searching is difficult for them (Schacter et al., 1998). This results in a trial and error strategy without a systematic approach (Koot & Hoveijn, 2005).

'Specifying search terms' is difficult for all age groups (Bilal, 2000; Kafai & Bates, 1997; Large & Beheshti, 2000; Lyons et al., 1997; Makinster et al., 2002; Schacter et al.,

1998; Wallace, Kupperman et al., 2000). Young children often use full sentences instead of keywords (Bilal, 2000; Koot & Hoveijn, 2005; Schacter et al., 1998). Teenagers do not always know which search terms to employ, especially when multiple keywords are involved (Large & Beheshti, 2000). And when they do use multiple keywords, they often make their searches too broad, resulting in an overload of results (Duijkers et al., 2001).

In some studies involving adult searchers the groups investigated were subdivided into successful and unsuccessful searchers. Successful searchers used "well-composed keywords phrases and often put their keywords in quotes" (MaKinster et al., 2002, p.161). Unsuccessful searchers had more trouble finding the right keywords. Search success appeared to be strongly related to domain expertise. Students with more domain knowledge were more successful. Their domain knowledge helped them to specify better search terms and they were able to distinguish better between usable and non-usable sites. Novices had more trouble with understanding the structure of the information and did not know where to start the search (MaKinster et al., 2002). Monereo et al. (2000) also concluded that adults who were subject matter experts were better searchers than domain novices.

Another important sub skill is 'evaluate search results'. Some young children are reluctant to read or scan the list of results (hit list). They base their choice for opening a site on titles only (Kafai & Bates, 1997), while others rely heavily on the summaries describing the results (Hirsch, 1999; Koot & Hoveijn, 2005) or the rank in the hit list (Koot & Hoveijn, 2005). In the studies of Lyons et al. (1997) and Wallace, Kupperman et al. (2000), young children evaluated the results based on the number of results their search produced. When only a few hits were generated by the search engine, they took this as a clue that the right answer was on one of those websites. Koot and Hoveijn (2005) also found that young children use a search engine as a magical machine, they expect the machine to provide them with the complete answer. Teenagers view every result without a clear evaluation of the results (Duijkers et al., 2001; Fidel et al., 1999).

Unsuccessful adult searchers do not evaluate results and the summaries, mostly because of a lack of domain knowledge. Like teenagers, the strategy chosen by unsuccesful adults was to inspect the search results in the order they were presented. Successful adult searchers with considerable background knowledge evaluate results by looking at the title, the origin of the source, the description and useful information or identifiers in the URL such as ".edu" or ".com" (MaKinster et al., 2002).

In this phase of the process computer skills are determinative for the result of a search in young children (Kafai & Bates, 1997). This is different with adults. Adults with domain expertise but without computer skills solved the given problem in less time than adults without domain expertise but with computer skills (Monereo et al., 2000). Brand-Gruwel et al. (2005) also revealed that the way adults searched the WWW is more influenced by domain knowledge than by computer expertise. In their study, both experienced and novice adult webusers searched the Web in a similar way on a task none of them was familiar with.

To conclude, most problems in the constituent skill 'search information' occur with sub skills 'specify search terms' and 'evaluate search results'. Young children, teenagers and adults do not always know which search terms to use. Young children tend to use natural language or long sentences. The use of keywords improves with age, but only if

domain knowledge is high. Adults with low domain knowledge lack the knowledge to come up with useful keywords and make their search too broad, resulting in an overload of hits. Moreover, evaluating the search results is not done systematically. People of all ages do not always open websites based on a valid evaluation of the results. The source is not always questioned and the choice for opening a site is mostly guided by the title or summary of the site.

Scan information

After opening a website, the site will be scanned. When reading the information globally, young children and teenagers seem to be looking for exact matches to the answer they have in mind (Hirsch, 1999) and to be trying to find the perfect webpage and a ready-made answer (Fidel et al., 1999; Lyons et al., 1997; Wallace, Kupperman et al., 2000). They scanned pages for the presence of pictures or read the first paragraph of a site to determine if it was worthwhile (Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997). Young children also tended to believe that everything that is posted on the Web is true (Hirsch, 1999; Schacter et al., 1998). Koot and Hoveijn (2005) found that young children say they trust the information they find, even if this information does not agree with their own experience. Relevance criteria mentioned most by young children are topicality, novelty and interest. Language (own versus foreign), authority and recency were hardly mentioned and young children did not actively consider the truthfulness, accuracy or validity of the information they found (Hirsch, 1999). Koot and Hoveijn (2005) found that young children are aware of the fact that not all information on the WWW is true. However, they rarely check information from one site with another site, especially when the information agrees with prior knowledge. Evaluating is mostly done based on appearance, the length of the text and use of language (i.e., difficult words). When children are equipped with more knowledge of the Internet and WWW, they become more critical. They evaluate the owner of the site, look for up to date information and read more sites. However, the source and owner has to be recognizable or easy to spot. Children rarely actively search for the owner. Sources mentioned on the site are not checked either, and if there are more sites with the same information, the information is accepted as correct without further research.

Teenagers also have trouble separating reputable and questionable materials, and have problems with selecting and evaluating information (Duijkers et al., 2001; Lorenzen, 2002; Lyons et al., 1997). They use information that could answer their question, even if the site was from a commercial source and not intended for science assignments (Fidel et al., 1999). There is one study that exclusively focused on the sub skill of evaluating and selecting information of teenagers. Lorenzen (2002) interviewed 10th and 12th grade high school students to reveal how students are using the WWW to find information and how they evaluate the information. Results showed that students relied heavily on the search engine to distinguish good from bad sites. The criteria used by the students to evaluate a website and the information are the organization behind a page, the extension of the URL (.edu and .gov), the author and bibliography, whether the information was believable, spelling and grammar and the elaborateness of a site. These criteria seem rather advanced, but the students had trouble to formulate and apply them. It took the students much time to come up with criteria and they found it

hard to express how they distinguished between good and bad sites. The criteria they mentioned were used too rigorously. For instance, they believed that the domain extensions guaranteed quality and they gave too much credence to the layout and elaborateness of a page. One of the students also rejected a good website because it had a spelling error. In fact, the spelling "error" was a British instead of American spelled word (honour versus honor). Furthermore, students do not seem to realize that the author of a site can be biased or that the authorship of a page is not always as advertised. So, teenagers use some criteria to evaluate webpages, but do not know how to use these criteria and how they can tell the difference between good and bad information (Lorenzen, 2002).

When asked which sources they use, adults in the study by Rosell-Aguilar (2004) said that they consult reliable sources like the university page, local newspapers and so on. They scanned a page thoroughly and followed links, using multiple sources of information. Monereo et al. (2000) reported that the majority of adult respondents to their questionnaire had great faith in the credibility of the information they had found. Results of Brand-Gruwel et al. (2005) revealed that adult expert searchers evaluated the quality and relevance of the information and the reliability of the sources more often than novice searchers did.

After evaluating information and source, relevant information should be stored. Young children do not record useful URLs or websites, resulting in trying to recreate good searches to return to previous sites (Hirsch, 1999; Large & Beheshti, 2000; Wallace, Kupperman et al., 2000). Schacter et al. (1998) found that young children did not bookmark many documents spontaneously. When they were explicitly asked to find at least three sources, they bookmarked more sites.

Children have the tendency to use the "Back" button to return to useful sites, instead of bookmarking (Fidel et al., 1999). It looks like young children and children do not store information and do not elaborate on content, but use the relevant information the first time they see it and integrate the scanning and processing phase. Furthermore, the expert adult searchers in the Brand-Gruwel et al. (2005) study spend more time on elaboration on content than the novices.

To conclude, the biggest problem while 'scanning information' is that evaluating is done based on expected information and not on aspects like validity, authority and recency. Most young searchers do not store relevant information. If a source seems useful after initial scanning, the site is read in depth and information is processed. They do not elaborate on content. Adult searchers seem to take the time to first scan and then process the information. In terms of the skill decomposition students of all age groups have problems with 'evaluating information and source'. Young children and teenagers also have problems with 'storing relevant information' and 'elaborating on content'.

Process information

Only five studies included results that concern the constituent skill 'process information'. Young children rarely take the time to read a site in-depth (Kafai & Bates, 1997; Schacter et al., 1998; Wallace, Kupperman, et al., 2000). They also tend to evaluate processed information by looking for words they expected to find. "They accepted the

source as valuable if it contained those words, in some cases irrespective of the actual meaning of the page" (Wallace, Kupperman et al., 2000, p. 93). Teenagers tended to do the same. In the study by Lyons et al. (1997) children used a commercial website to answer their question; they were unaware that the page only "applied to a specific product and was not necessarily the norm" (p. 21).

Young children do not store relevant information but modify text from the site in their own words and add it to their final product (Large & Beheshti, 2000). Wallace, Kupperman et al. (2000) state that some young children "never read enough of the page to understand that its content had nothing to do with their question, and they used it as evidence that they had finished their assignment" (p. 94).

From these few studies it can be concluded that young children do not read to understand the text in depth. 'Evaluating processed information' seems to be a problem for young children and teenagers. Furthermore young children seem to have trouble with 'storing relevant information'.

Moreover one can question if the difficulty with processing information springs from the fact that the Internet is made up with HyperText Markup Language (HTML), a language that allows documents to integrate references to other documents. Rouet and Levonen (1996) conclude that reading hypertext has benefits: extra information becomes easier to access in a hypertext environment. However, the risk of disorientation is higher and processing information in hypertext imposes a higher cognitive load on the users. Yet, providing users with structure and coherence cues can help overcome these problems (Rouet & Levonen, 1996).

Organize and present information

This constituent skill and its sub skills formulate problem, structure relevant information, outline the product, realize product and elaborate on content, is only mentioned in one of the 15 studies. Brand-Gruwel et al. (2005) mentioned that experts and novice adults spent an equal amount of time on this phase, but experts paid more attention to the formulation and reformulation of the problem. In general, adults do not seem to have problems with this skill. In conclusion: organize and present information has not been described enough to point out which problems children, teenagers and adults may encounter. It can be stated that the way the information must be organized and presented in itself can be a complex cognitive skill. For instance, writing a scientific article is not an easy job. Research focusing on students' writing skills will probably give more insight in the problems students encounter with this skill.

Regulation

In six articles comments are made on regulation. Hirsch (1999) stated that young children "did not keep track of how they searched for information. They did not record useful URLs or keep a record of search queries" (p. 1271). Teenagers did not feel the need to plan a search or to check whether their planning was adequate (Fidel et al., 1999; Lyons et al., 1997). However, they did check their spelling in URLs and search terms and were aware of the fact that spelling can influence the results of a search (Fidel et al.)

Adults who can be categorized as strategic or successful searchers show signs of orientation, monitoring, steering and evaluating. Non-strategic searchers are less successful and do not regulate their search process (MaKinster et al., 2002; Monereo et al., 2000). Brand-Gruwel et al. (2005) stated that adult experts monitored and steered their process more often than novices.

To conclude: there is evidence that students in all age groups have problems with regulation. From the results it can be inferred that the quality of the IPS-process is influenced by regulation. Children, teenagers and adults become better searchers when they orientate, test, monitor, steer and evaluate during the ongoing process.

Summary

This review of research focuses on problems people have with the different constituent and sub skills involved in the IPS-process. Table 2 gives a summary of the results.

The skills 'searching' and 'scanning information', have been mostly addressed. The results show that the sub skills 'specify search terms' and 'evaluate search results' of the constituent skill 'search information' and the sub skill 'evaluate source' and 'evaluate information' of the constituent skill 'scan information' are a problem for all age groups.

It would be logical that instruction to foster students' information-problem solving ability should address the skills students have difficulties with. In the next part of this chapter we will discuss several instructional methods. Only empirically tested instruction and support is included in the overview.

Table 2 Problematic sub skills per age group

	Constituent skills					
	Define information problem	**Search information**	**Scan information**	**Process information**	**Organize & present information**	**Regulation**
Children (6-12 year)	Formulate questions Activate prior knowledge Clarify task requirements Determine needed info	Specify search terms Evaluate search results	Evaluate information and source Store relevant information Elaborate on content	Read in depth Evaluate processed information Store relevant information	Undetermined	Orientation Testing Monitoring Steering Evaluation
Teenagers (13-18 year)	Formulate questions Activate prior knowledge Clarify task requirements Determine needed info	Specify search terms Evaluate search results	Evaluate information and source	Evaluate processed information	Undetermined	Orientation Testing Monitoring Steering Evaluation
Adults	No problematic skills	Specify search terms Evaluate search results	Evaluate information and source	No problematic skills	Undetermined	Orientation Testing Monitoring Steering Evaluation

Instructional solutions

The constituent and sub skills of the IPS-process can cause problems for students of all age groups. The next question is: how can instructional support foster students to become more proficient in information-problem solving? There are general instructional methods that focus on information-problem solving with (electronic) library systems (e.g., Berner, McGowan, Hardin, Spooner, Raszka Jr., & Berkow, 2002; Eskola, 2005; Larkin & Pines, 2004; Todd, 1995; Wallace, Shorten & Crookes, 2000). The focus of these methods was mostly on the constituent skill search information, target groups were mostly children or adults. Although results of experimental groups were better than those of most control groups (e.g., Larkin & Pines; Todd, 1995; Wallace, Shorten, et al., 2000), we did not use these studies in our review. As mentioned these instructional settings addressed searching within a specific system and not on the Web. Searching a library database and searching the Web appeal on different skills. For instance, the Web does not have an index or table of contents, and selecting the right keyword is therefore more important. Furthermore, the Web is much more extensive than a library database. The risk of wandering off is high and processing information is much more difficult.

A quote from Larkin and Pines (2004) points to another important difference, selecting and evaluating information is harder on the Web: "To ensure that they selected quality studies, the instructions required that they use the library databases (e.g., EBSCOhost, PsychInfo, etc.) and not Google or Yahoo" (p. 43). In our review therefore only empirically tested instructional methods for searching on the Web are included.

The results section will be organized by instruction for young children, teenagers and adults. In total 12 studies were found and will be analyzed (see Table 3).

Table 3 Studies on support and instructional methods for IPS

Britt and Aglinskas (2002)
Colaric (2003)
De Vries et al. (2008)
Duijkers et al. (2001)
Feddes et al. (2003)
Hoffman et al. (2003)
Kuiper et al. (2008)
Lazonder (2001)
Pritchard and Cartwright (2004)
Gerjets and Hellenthal-Schorr (2008)
Stadtler and Bromme (2008)
Wopereis et al. (2008)

Instruction and support for young children

De Vries, van der Meij, and Lazonder (2008) created a task-related portal to support reflective web searching by elementary school children (fifth and sixth grade) while working on a collaborative task in the domain of biology. This portal was embedded in biology lessons. In the first design experiment, four elementary classrooms of different schools participated. Children worked in groups on a biology assignment for six lessons.

They were asked to activate their prior knowledge. They used the portal (a webpage with task-related categories and hyperlinks with meaningful names, indicating the content of the page) to answer their research questions. The children also received a worksheet on which they wrote down their own research questions, and, after they had completed their search, their answer. Results showed that this portal provided the children with too little structure.

The second experiment was conducted with two classrooms, with an adjusted portal. A hierarchy of main topics was added and a sitemap was provided. The hyperlinks were enriched with indications of the amount and sort of information that could be found. The worksheet was also slightly adjusted: children wrote down their research questions, their provisional answers and their final answers. The children worked in groups and formulated their answers together. This stimulated them to express their thoughts, reflect on findings on the Web, and relate new information to prior experiences by talking about it. The results of this design-based research show that the portal helped them to find relevant websites and select useful information.

Hoffman, Wu, Krajcik and Soloway (2003) used a software program called Artemis to unravel the information seeking strategies of middle school students. Artemis provides students with a digital library to search and sort science information related to project-based investigations. Artemis only offers websites appropriate for school age children, selected and screened by librarians. "It helps students focus on the content of the on-line resource, evaluate its usefulness, and synthesize information rather than spending the majority of time simply locating appropriate sites on the WWW" (Hoffman et al., 2003, p. 324). In this study, the authors "developed on-line and off-line learning materials to provide scaffolding, to support students' information-seeking activities as they asked questions of interest, searched for information, assessed their findings, and created rich representations of their newly constructed understandings" (p. 324).

This post-test only study investigated the depth and accuracy of 16 sixth-grade students' content understandings as well as their use of search and assesses strategies as they used on-line resources via Artemis. Results showed that the depth of students' understanding after working with Artemis varied. Most participants were able to articulate explanations and relations during an interview but these were only partially accurate. Some students could provide accurate understandings, but these were not very deep and often limited to recalling information. The results of the interviews were better than the products students delivered, the products "communicated a simple recall of factual information" (p. 336). The students who adequately engaged in inquiry strategies obtained more accurate understandings. These students thought about a number of possible search topics and were careful in the use of queries. They also showed selectivity in sources, deep navigation into sites, browsed the contents, and paused to read information related to their on-line inquiry. Students with better content understandings also used more complicated strategies to assess on-line resources. They judged whether information was relevant to their driving question before investing time on a site. Decisions were based on a site's content rather than appearance or title. The majority of time was spent with worthwhile and understandable information; however, trustworthiness of the source was often based solely on the URL (e.g., .org, .com, .gov, .edu). Students were able to provide a limited critique of a site's appearance and

content. Students with less content understandings were more likely to trust information, and judged relevancy based on appearance. Results show that students may benefit from the scaffolding features in Artemis and the off-line materials, but this is not true for all students and does not occur automatically.

In a part of a pilot study for a larger project concerning ways to improve the use of Internet for information location Pritchard and Cartwright (2004) asked 54 children (ages 10 and 11) to produce an information sheet about the history of bikes for children of their own age. Participants received a list of things they had to take into account when creating the sheet and a list of ten relevant websites. Before they were allowed on the Internet, they had to activate prior knowledge through brainstorming with the teacher. The instruction consisted of a set of rules and examples of the use of the rules. Children could work on the assignment for two lessons. The rules were: (1) Keep any extract from the Internet short. (2) Make a comment about any extract you include. (3) Say where the information came from. The first two rules encouraged children to engage with the text, think about the extract and give it a context. They necessitated reading and making decisions about which part to select. The third rule helped to avoid unintentional plagiarism. There was no control group in this study. Results revealed that the end products were not optimal, "some children took extracts directly from a website and gave the impression of not having read the words which they were using" (p. 28). Children had not engaged with the content in a meaningful way, although there was evidence that some children had composed their own text. Some children were able to use the rules and make comments on the sites, indicating "that they had considered the information and had gone beyond the information given" (p. 30). However, the support had little impact on the children's learning: children were not able to recall what they had learnt about bikes a week after they made the sheet.

Kuiper, Volman, and Terwel (2008) designed a curriculum for fifth graders to acquire Web skills. It was a multiple case study design in which four different schools participated. The knowledge domain of the curriculum was healthy food. The implemented curriculum consisted of eight weekly lessons of 1,5 to 2 h each. The first five lessons were aimed at developing websearching, reading and evaluation skills. In the three last lessons, students received assignments and used the Web to search for information and composed their own texts based on that information. Results showed that students' knowledge about webskills improved. Students appeared to be inconsistent webusers, who did not always act upon their knowledge of web searching skills. Students showed unexpected, inconsistent or inflexible web behaviour and little planning and reflection.

To summarize, instruction for young children often combines a project on a certain topic with instruction on IPS, and can thus be categorized as embedded instruction. Collaborative instruction and discussions between students (De Vries et al., 2008; Hoffman et al., 2003; Kuiper et al., 2008) helps children become more engaged with the subject and information than individual instruction (Pritchard & Cartwright, 2004). Furthermore, three studies used computer based instruction (De Vries et al., 2008; Hoffman et al., 2003; Kuiper et al., 2008), one study used paper materials (Pritchard & Cartwright, 2004).

Table 2 shows the problematic skills of young children. All instructional methods at least addressed the problematic skill 'evaluate information and source'. 'Formulate questions' is addressed by De Vries et al. (2008). 'Activate prior knowledge' is addressed by De Vries et al. (2008), and Pritchard and Cartwright (2004). All methods paid attention to the beginning of the search process by either addressing the sub skill 'formulate questions' (De Vries et al., 2008) or addressing the sub skill 'specify search terms' (Hoffman et al., 2003). Pritchard and Cartwright address 'store relevant information' and 'elaborate on content'. The latter is also addressed by Hoffman et al. (2003). 'Read in depth' and 'evaluate processed information' is addressed by Kuiper et al. (2008).

Problematic skills not addressed by these methods were: 'clarify task requirements', 'determine needed info', 'evaluate search results' and the constituent skill 'regulation'.

These four studies do not use a pre-test and control group in their designs. It cannot be excluded that improvement of IPS skills, knowledge and rules is also caused by natural development and not only by instruction. Most studies have a large *N*, only the study by Hoffman et al. (2003) has an *N* lower than 20.

Instruction and support for teenagers

Britt and Aglinskas (2002) developed The Sourcer's Apprentice, a computer application for teaching sourcing (identifying critical features of the source like author, author's position, date, document type, etc.), contextualization ("identifying relevant features of a source that can be useful in creating a context for historical information", p. 489) and corroboration (checking facts or interpretations from one source against other sources) in the context of researching a historical controversy. The Sourcer's Apprentice provides students with several documents about a controversy and information about the documents such as author's credentials and possible motives. After reading the documents students fill in note cards. The note cards allow students to fill in information about six source and three content features like author (who, position, how know and author motives) and document (when, type). After filling in the note cards, students receive a series of questions about the sources and contents of the documents and are asked to write an essay on the controversy.

A pre-test-post-test control group design with one experimental and one control group was used to test the Sourcer's Apprentice. The experiment was conducted twice, with different populations. Eleventh grade students of two American history classes (*N* = 15) and eleventh grade students of two economics classes (*N* = 29). During the pre-test, all participants were asked to read six documents centered on a controversy while taking notes. Then they received a question booklet, containing sourcing questions (e.g., "Which document was written earliest") and two essay questions. Next, the experimental group received a 2-day exposure to the Sourcer's Apprentice and a control group received 2 days of regular classroom activities on the module topic. Post-test was the same as the pre-test, but centered on a new controversy. Results showed that the experimental group outperformed the control group on the post-test; their sourcing skills had improved.

Duijkers et al. (2001) provided 28 teenagers (age 14) with a step-by-step plan to stimulate a critical look at sources and information. Participants in this case study worked in pairs. They had to choose four out of 20 sources to answer a research question. The step-by-step plan guided the children through the steps of thinking about criteria for sources and evaluating the sources on applicability to answer the research question. The sources had to be divided in three groups, usable, may be usable and not usable. Students had to explain why they put a source in a certain group. Four sources had to be chosen from the group with usable sources to answer the question. Participants stated that working with the step-by-step plan helped them to work more effectively and defend a choice for a specific source.

Lazonder (2001) instructed teenagers (mean age 14.2) in basic procedural skills and self-regulatory skills while searching the WWW. There were three instructional groups: a memory aid group, a timesharing group and a control group. The total number of participants was 168. All groups received materials on procedural skills (e.g., entering an URL, following hyperlinks). The materials differed with regard to the instructional strategy to learn self-regulatory skills. The memory aid version included a diagram of the search process to introduce self-regulatory skills. The diagram was explained in the first chapter of the materials, prior to the procedural skills. Subsequent chapters only contained procedural skills instruction, though students were encouraged to use the diagram when following the instruction. Students could consult the diagram at will. In the timesharing version, the same regulatory skills were addressed. The skills were introduced in conformance with appearance in the search process. The instruction on regulatory skills was integrated with the procedural skills instruction. There was no diagram of the search process provided. The control group only received the procedural skills material. These procedural materials were identical to the material in the other groups. Each group attended four sessions of 50 min each. It was expected that students in memory aid and timesharing groups would outperform the students in the control group on Web search tasks and search tasks in an electronic database (OPAC task). The study used a 3 x 3 factorial design with three levels of instructional condition (memory aid, timesharing, control) and three levels of Web expertise (novice, beginner, intermediate). Results showed no performance gains of self-regulatory instruction. Students in all groups performed the same on the search tasks.

Gerjets and Hellenthal-Schorr (2008) designed and empirically evaluated a training program called CIS-WEB (Competent Information Search in the World Wide WEB). The program was developed to improve pupils' (ages 12-13) processing of information in order to foster their ability to competently search for information on the WWW. CIS-WEB consists of six modules and is designed as an in-class training. The six modules aimed at basic knowledge about the Internet, the WWW and search systems in the Web, information-problems, structure of websites and use of web tools, evaluation with regard to credibility and actuality, segmentation of information-problems and processing of the resulting sub tasks. Students listened to presentations, worked collaboratively in a hypermedia environment and worked individually with paper and pencil materials like worksheets.

It was assumed that following the CIS-WEB program would result in an improvement of pupils' declarative knowledge of the Web and in better search

performance. Furthermore, a stronger improvement was expected for pupils with higher engagement in the web training compared to those with lower engagement. 61 students participated. Data was gathered four times during the training. Declarative knowledge was measured with a multiple-choice test, search performance was measured by the way students solved sets of information-problems. Results showed that CIS-WEB enhanced pupils´ declarative knowledge about the Web and their search performance compared to the control group.

To summarize, instruction for teenagers is offered embedded (Britt & Aglinskas, 2002) as well as stand alone (Duijkers et al., 2001; Gerjets & Hellenthal-Schorr, 2008; Lazonder, 2001). Student motivation and engagement seem to be important factors for improving IPS-skills (Gerjets & Hellenthal-Schorr, 2008). Only one instructional method for teenagers used a collaborative method (Duijkers et al., 2001); the other methods were given to individual students (Britt & Aglinskas, 2002; Lazonder, 2001) or use collaborative and individual methods (Gerjets & Hellenthal-Schorr, 2008). Two methods were 'paper en pencil' based (Duijkers et al. 2001; Lazonder, 2001), the other two used computer based instruction materials (Britt & Aglinskas, 2002; Gerjets & Hellenthal-Schorr, 2008).

Table 2 shows the problematic skills of teenagers. Except for Lazonder (2001) all methods at least addressed the problematic sub skill 'evaluate information and source'. The instruction by Lazonder (ibid) is the only method addressing the constituent skill 'regulation'. 'Specify search terms' is addressed by Lazonder (ibid), and Gerjets and Hellenthal-Schorr (2008). The latter also addressed 'formulate questions'.

Problematic skills not addressed by these methods are 'activate prior knowledge', 'clarify task requirements', 'determine needed information' and 'evaluate search results'.

Three of four studies (Britt & Aglinskas, 2002; Gerjets & Hellenthal-Schorr, 2008; Lazonder, 2001) use a design with a control group. The study by Duijkers et al. (2000) has a smaller *N* (28) than the other studies. Only Britt and Aglinskas (2002), and Gerjets and Hellenthal-Schorr (2008) use a pre-test.

Instruction and support for adults

Colaric (2003) examined three instructional treatments to support adults in using a search engine, including specifying search terms. The three treatments, instruction by example (*N* = 59), conceptual models without illustrations (*N* = 61), and conceptual models with illustrations (*N* = 56), were compared on differences in knowledge acquisition: declarative (understanding of factual information about a search engine), syntactic (understanding of the appropriate formulation of a search query) and semantic knowledge (understanding of the major objects and actions of a search engine). Pre-test and post-test were the same. Declarative knowledge was measured with questions on factual knowledge of search engines. Syntactic knowledge was measured by the elements of a search query with regard to a provided search problem. Semantic knowledge was measured by the participant's explanation of how a search engine works. The three treatments were each given during one class period. Results reveal that all instructional treatments were effective for increasing the three types of knowledge. However, syntactic knowledge (which can be compared with specifying search terms, a sub skill all age groups have problems with) increased most

with instruction by example. This study was done with written material and did not involve actual searches on the Web.

In a study by Feddes, Vermetten, Brand-Gruwel, and Wopereis (2003) adults received an IPS training. The training was based on the skills defined by Eisenberg and Berkowitz (1990). During the pre- and post-test the participants (*N* = 4) were given an information-problem and were interviewed about how they would solve the problem. Results show changes in 'problem definition', 'searching', and 'processing'. Participants gave more elaborate descriptions of the latter two after the training. 'Problem definition' was almost absent in the pre-test and was mentioned by most participants after the training.

Stadtler and Bromme (2008) provided adults with little medical knowledge (*N* = 118) with evaluation and monitoring prompts while searching the WWW on a medical topic. For this purpose, the metacognitive tool met.a.ware was developed. This tool enabled users to store information they have found systematically. It provides them with different labelled tabs (ontological classification) under which they can store information. To test met.a.ware, participants received preselected websites on the topic of cholesterol. Participants (aged 19 to 38, mean age 23.81) received prompts to evaluate sources on credibility or assess how well they comprehended information and how much they still needed to search for more information. There were four experimental conditions. The difference between the experimental conditions was the type of prompts participants received. The evaluation group received evaluation prompts, the monitoring group received monitoring prompts, the evaluation and monitoring group received two types of prompts and the no prompts group did not receive prompts. There were also two groups who did not work with met.a.ware. One of these control groups took notes with paper and pencil; the other used a text window to copy and paste information from the WWW into text slots. A pre-test was administered to measure computer and Internet experience, as well as factual knowledge on cholesterol. After 40 min of searching, participants repeated the test on factual knowledge and answered four questions on subject matter. The post-test also included an assessment on knowledge about sources and a rating of the credibility of websites.

Results showed that prompts for monitoring and evaluation increased knowledge on content and sources, and ontological classification helped to structure notes and focused participants on important ontological categories.

Wopereis et al. (2008) compared a control group that did not receive instruction with an experimental group that received additional instruction on IPS integrated in professional distance education skill training (*N* = 16). The aim of their study was "to find out what the effect of the integrated IPS instruction was on the way distance education students solve information problems" (p. 9). Adults in the experimental group were taught how to seek information efficiently. The emphasis of the training was on the regulation of the process. A pre-test post-test control group design was used. On the pre-test no differences were found with regard to prior knowledge on IPS and regulation ability. Results on the post-test differ somewhat between the two groups. The constituent skill 'scan information' and the sub skill 'evaluate information' were performed more by students of the experimental group. Participants in the experimental groups also

monitored and steered their process more often. Time spent on defining the problem was low in both groups.

Almost all instructional methods for adults were offered as separate courses. Only the instruction provided by Wopereis et al. (2008) was embedded in a curriculum. All described instructional methods for adults are individual. Adults seem to benefit from instruction that focuses on the process of IPS. When they receive instruction or examples on how to search effectively, their results on a task improve. Only Stadtler and Bromme (2008) use computer based instruction.

Table 2 shows the problematic skills of adults. These are 'specify search terms', 'evaluate search results', 'evaluate information and source', and the constituent skill 'regulation'. Colaric (2003) only addressed 'specify search terms'. Feddes et al. (2003) and Wopereis et al. (2008) addressed all constituent skills and sub skills. Stadler and Bromme (2008) addressed the problematic sub skill 'evaluate information and source' and the constituent skill 'regulation'.

Two studies use a control group (Stadtler & Bromme, 2008; Wopereis et al., 2008) and all studies use a pre- and post-test. The study by Feddes et al. (2003) has a very small number of participants.

Summary

Table 4 gives an overview of the IPS instruction and support per age group. It shows that five instructional methods are given embedded (Britt & Aglinskas, 2002; De Vries et al., 2008; Hoffman et al., 2003; Pritchard & Cartwright, 2004; Wopereis et al., 2008) and one partly embedded (Kuiper et al., 2008). Six methods are not embedded in a course or project (Colaric, 2003; Duikers et al., 2001; Feddes et al., 2003; Gerjets & Hellenthal-Schorr, 2008; Lazonder, 2001; Stadtler & Bromme, 2008).

Four of the 12 discussed methods used collaborative instruction or assignments (De Vries et al., 2008; Duijkers et al., 2001; Hoffman et al., 2003; Kuiper et al., 2008). The other instructional methods were individual methods (Britt & Aglinskas, 2002; Colaric, 2003; Feddes et al., 2003; Gerjets & Hellenthal-Schorr, 2008; Lazonder, 2001; Pritchard & Cartwright, 2004; Stadtler & Bromme, 2008; Wopereis et al., 2008).

Five out of 12 studies use a computer program for their instruction (Britt & Aglinskas, 2002; De Vries et al., 2008; Hoffman et al., 2003; Kuiper et al., 2008; Stadtler & Bromme, 2008), six studies use paper materials (Colaric, 2003; Duijkers et al., 2001; Feddes et al., 2003; Lazonder, 2001; Pritchard & Cartwright, 2004; Wopereis et al., 2008). One study combines computer materials with paper materials (Gerjets & Hellenthal-Schorr, 2008).

The problematic sub skills for every age group are 'specify search terms', 'evaluate search results', 'evaluate information and source' and the constituent skill 'regulation'. One or more of these skills are addressed in every instructional method.

Table 4 Support and instructional methods for IPS

Age group	Authors	Instruction/ support	Embedded/ stand alone	Individual / collaborative	Constituent skill(s) addressed	Sub skill(s) specially addressed	Effective
Child (6-12 year)	De Vries et al. (2008)	Support	Embedded	Collaborative	Define information problem	Formulate questions Activate prior knowledge Evaluate information and source	Yes
	Pritchard & Cartwright (2004)	Support	Embedded	Individual	Define information problem Scan information Process information	Activate prior knowledge Evaluate information and source Store relevant information Eleborate on content	Partly
	Hoffman et al. (2003)	Support	Embedded	Collaborative	Search information Scan information Process information Organize and present information Regulation	Specify search terms Evaluate information and source Realize product	Partly
	Kuiper et al. (2008)	Support	Partly embedded	Collaborative	Search information Scan information Process information	Select search strategy Specify search terms Evaluate search results Read information global Evaluate information and source Elaborate on content Read in depth Evaluate processed information	Partly
Teens (13-18 year)	Britt & Aglinskas (2002)	Support	Embedded	Individual	Scan information Process information	Evaluate information and source	Yes
	Duijkers et al. (2001)	Support	Stand alone	Collaborative	Scan information Process information	Evaluate information and source	Yes
	Lazonder (2001)	Instruction	Stand alone	Individual	Search information Regulation	Specify search terms Orientation Testing Monitoring Steering Evaluation	No
	Gerjets & Hellenthal-Schorr (2008)	Instruction	Stand alone	Individual	Define information problem Search information Scan information Process information	Formulate questions Specify search terms Evaluate information and source	Yes
Adults	Colaric (2003)	Instruction	Stand alone	Individual	Search information	Specify search terms	Yes
	Feddes et al. (2001)	Instruction	Stand alone	Individual	All constituent skills	All sub skills	Yes
	Stadtler & Bromme (2008)	Support	Stand alone	Individual	Scan information Process information Regulation	Evaluate information and source Orientation Testing Monitoring Steering Evaluation	Yes
	Wopereis et al. (2008)	Instruction	Embedded	Individual	All constituent skills, emphasis on regulation	All sub skills	Yes

Most of the instructional methods are effective. However, the evidence for this conclusion is not very strong due to methodological shortcomings. Only very few methods used a control group (Britt & Aglinskas, 2002; Gerjets & Hellenthal-Schorr, 2008; Lazonder, 2001; Stadtler & Bromme, 2008; Wopereis et al., 2008); and in only one occasion have the results of the instructional support been tested for transfer (Lazonder, 2001).

Some studies are only partly effective. In the studies by Pritchard and Cartwright (2004) children did not engage with content because they were focusing more on the mechanical aspects like navigating and cutting and pasting information. And although they did follow the instructed rules, their final products lacked quality. Pritchard and Cartwright state: "simple exposure to information or simple copying from one place to another does not imply learning" (p. 30). The rules provided by Pritchard and Cartwright were not enough, children should also be encouraged to actively engage with the information. Differences in the active engagement in the inquiry process are also the reason for differences in acquired knowledge in the Hoffman et al. (2003) study.

The instructional method by Lazonder (2001) was not effective. One of the possible reasons is the time factor. Instruction time may have been too brief for regulation skills to develop.

Discussion

The goal of this literature study was to give an overview of the problems children, teenagers and adults encountered while solving information-problems using the Web for searching information and of the effects of different kinds of instructional support to foster students' information-problem solving ability.

It can be concluded that people in every age group experience some problems with IPS. Some constituent and sub skills are mastered during the process of growing up, other skills remain problematic throughout life. Children, teenagers and adults have problems with specifying search terms, evaluating search results, evaluating information and source and regulating their search process. Children and teenagers also have trouble with the constituent skills define information problem and process information. Instruction in IPS should therefore take into account the age of the target group and adjust the instruction accordingly.

The instructional and support methods reviewed in this article can be grouped based on several features: the way the instruction is offered (either embedded in the curriculum or as a separate course); the way the instruction is followed by participants (individually or collaboratively), tools used during the instruction, and the skills addressed in the instruction. The review shows that there are only a few empirically tested instructional or support methods for IPS. Most of the methods found were stand-alone courses for individual use. Tools used in these methods differ from a web-based portal or a computer application (Britt & Aglinskas, 2002; De Vries et al., 2008; Hoffman et al., 2003; Kuiper et al., 2008; Stadtler & Bromme, 2008), to worked-out examples and visualizations (Gerjets & Hellenthal-Schorr, 2008), to worksheets (Lazonder, 2001), to paper material only (Colaric, 2003; Duijkers et al., 2001, Feddes et al., 2003; Pritchard &

Cartwright, 2004; Wopereis et al., 2008). It is promising that all methods aim at (some of) the problematic skills of their target group and that most of them are effective. However, the effectiveness of the methods has not been established without doubt. The first, and perhaps one of the biggest, question marks that can be placed by the instructional methods is the fact that only one of them (Lazonder, 2001) tested for transfer. Yes, most of the instructional methods were effective, but none of them were tested again after a certain amount of time and only one within a different context. It is not certain that the knowledge and skills participants gained during the instruction were embedded in long-term memory and can be called upon while solving new information-problems.

Next, it remains unclear whether or not instruction in IPS should best be given embedded or stand alone. Although results from library research point towards embedded instruction as being most effective (Larkin & Pines, 2004; Todd, 1995; Wallace, Shorten et al., 2000), stand alone methods in this review also have positive results. However, these results have not been compared to an embedded version of the instruction, and again, have not been tested for transfer.

Some researchers argue that skills that are highly regulative, as IPS, can be learned in specially designed courses, because the skills do not vary across disciplines (e.g., Paul, 1992). On the other hand, Brown (1997) stated that highly regulative skills must be taught embedded in a context of a specific subject matter, in such a way that transfer to other domains is possible. Brown points out the importance of using real-life problems, because it motivates and stimulates active involvement. Also research of Ten Dam and Volman (2004) reveals that stand alone programs stimulating skills that are highly regulative and make an appeal to students' critical thinking ability are not effective.

Another question concerning the design of IPS instruction is whether they should be given collaboratively or individually. Again, results are inconclusive. Most methods are for individual use. In the study by De Vries et al. (2008) the collaborative nature of the instruction was one of the reasons why sub skills 'formulate questions' and 'activate prior knowledge' improved. Collaboration also has a positive influence on regulation (Lazonder, 2005), one of the problematic areas in IPS. So a combination of individual and collaborative assignments seems a good instructional strategy.

Another issue concerns the tools used in instruction. Different kinds of tools are used in the reviewed research. It is hard to say which tools (paper and pencil, worksheet, computer tools, etc.) are most effective, because the setting, the characteristics of the students, and the objectives of the instructions must be taken into consideration.

An important issue when designing IPS-instruction is the focus of the instruction. IPS-instruction should strive to encourage students to actively engage in the process and not only focus on the 'mechanical' aspects. This also implies that the whole process should be taken into account. This finding is confirmed in the literature on problem-solving and the development of complex cognitive skills (van Merriënboer, 1997). Students should work on whole tasks, which are authentic and comprehensive. These tasks require students to perform all the constituent skills that make up the whole complex skill during task performance.

Moreover, scaffolding students to improve regulation should be part of the instructional setting. To improve regulation, cognitive apprenticeship (Collins, Brown, & Newman, 1989) is an appealing approach. This approach focuses on specific methods for carrying out complex cognitive tasks in which regulation is important. Apprentices learn these methods through the combination of observation, guidance and practice, or, from the teacher's point of view, through modelling, coaching and fading. The student repeatedly observes 'the expert' explicitly executing (modelling) the target process. The 'model' hereby externalizes the usually implicit cognitive thinking processes. After observing, the student attempts to execute the process with guidance and help. A key aspect in the coaching process is the provision of scaffolding: support, in the form of reminders and help so that the student can approximate the execution of the entire cognitive task. Once students have grasped the skill to be learned, the teacher reduces support (fading), providing only limited hints. So, cognitive apprenticeship intends to bring out internal cognitive processes in the open. Students are taught to act in the same way as modelled by the teacher.

A form of scaffolding and fading is providing the novice students with a process-oriented worked example. A study by Van Gog, Paas and van Merriënboer (2006) showed that novices who received a process-oriented worked example (the problem state, the end state, the solution steps that are to be taken to reach the end state and the strategic "how" and principled "why" information used in selecting the steps), performed better on a transfer task than novices who only received a conventional problem. However, after some time, the examples can cause a cognitive overload and should no longer be offered. It is therefore also necessary to adapt the instruction at the level of the students.

To conclude, students of all ages encounter problems with solving information-problems. Aspects as deriving search terms and evaluating information and source are often problematic. Instructional support to foster students' IPS skill is therefore essential. Research about instructional support does give ideas and guidelines for designing this kind of IPS support, like working with whole tasks and a focus on the whole process. However, further research should aim at the mentioned issues and should especially include transfer of the IPS skill.

CHAPTER 3

HOW STUDENTS EVALUATE SOURCES AND INFORMATION WHEN SEARCHING THE WORLD WIDE WEB FOR INFORMATION

How students evaluate sources and information when searching the World Wide Web for information

The World Wide Web (WWW) has become the biggest information source for students while solving information-problems for school projects. Since anyone can post anything on the WWW, information is often unreliable or incomplete, and it is important to evaluate information and source before using them. Earlier research has shown that students have difficulties with evaluating information and source. This study investigates the criteria secondary educational students use while searching the Web for information. Twenty-three students solved two information problems while thinking aloud. After completing the tasks they were interviewed in groups on their use of criteria. Results show that students do not evaluate results, information and source very often. The criteria students mention when asked which criteria are important for evaluating information are not always the same criteria they mention while solving the information-problems. They mentioned more criteria but also admitted not always using these criteria while searching the Web.

This Chapter is based on Walraven, A., Brand-Gruwel, S, & Boshuizen, H. P. A. (in press). How students evaluate sources and information when searching the World Wide Web for information. *Computers and Education*.

Introduction

In the last decade, the Dutch secondary educational system has undergone major changes. Instead of a system aimed at the reproduction of knowledge, the new learning environment aims at learning outcomes that should be durable, flexible, functional, meaningful and applicable. Active pedagogical methods, in which students learn by doing instead of listening, and teachers have a guiding role, fit this new learning (Simons et al., 2000). The teacher no longer provides students with information by standing in front of the classroom lecturing, but more and more, students actively and independently have to construct knowledge, collect information, evaluate information and integrate this with prior knowledge. This educational approach requires that students have a different type of skills than in traditional education. One of these skills is problem solving: a complex higher-order cognitive process that requires the modulation and control of more routine or fundamental skills (Goldstein & Levin, 1987). These routine or fundamental skills are clarifying the description of the problem, analyzing causes, identifying alternatives, assessing each alternative, choosing one, implementing it, and evaluating whether the problem was solved or not. Problem solving, a form of active participation in learning, is more satisfying than passive transfer of information from the teacher to the student and leads to enhanced retention and recall (Bransford, Brown, Cocking, 2000). Problem solving skills call upon regulation- and planning skills, also known as metacognitive skills (Perkins & Salomon, 1989; Van Merriënboer, 1997, 1999).

A specific type of problem that can be distinguished is an information-problem. Solving these kinds of problems is called the process of 'information-problem solving' (IPS). Being able to solve information-problems means that students can identify information needs, locate information sources, extract and organize information from each source, and synthesise information from a variety of sources (Brand-Gruwel et al., 2005; Eisenberg & Berkowitz, 1990; Moore, 1995). Within this IPS process evaluation of information and source is an important sub skill, especially since students often search on the Internet for information while accomplishing learning tasks. The Web is easily accessible and students are seduced to cut and paste the information without evaluating it (Grimes & Boening, 2001; Rothenberg, 1998), resulting in reports that lack quality. A lack of evaluation of results, information and source can affect students' learning and schoolwork.

The aim of this study is to examine how students in secondary education solve information-problems and especially to determine the criteria students use while evaluating search results, information and source using Internet for finding information while accomplishing an assignment. A description of what an information-problem is will be given first.

An information-problem arises when a discrepancy occurs between information needed to answer a question and information already known. An example of a simple information-problem is: you have to travel from London to Harwich on September 8. You need to arrive at 18.00. At what time does your train leave? This problem can be solved easily by visiting a site of the UK railways. There is probably only one correct answer to

this question. The assignment: 'write an essay on the treatment of cancer' is more complex. Solving this problem requires more than visiting only one webpage. There are several forms of cancer for instance and several treatment methods. Hence, more information has to be dealt with in this complex information-problem and writing an essay is more difficult than answering one question.

Solving information-problems requires several activities and these activities together constitute a process that we refer to as information-problem solving (IPS). The process consists of the constituent skills *defining the information problem* (i.e., reading the task, activating prior knowledge), *searching information* (i.e., choose search strategy, specify search terms, evaluate search results), *scanning information* (i.e., read information global, evaluate source and information, elaborate on content), *processing information* (i.e., read in depth, evaluate information, store relevant information, elaborate on content) and *organizing and presenting the information* (i.e., structure relevant information, realize the product).

Within the IPS process several problem solving processes can be distinguished. In the second phase, search information, one has to choose a search strategy and specify search terms. Choosing a search strategy and specifying search terms can be seen as problem solving processes on their own. What is the best strategy and best search term for the information problem at hand? With regard to the search strategy one could decide to use a depth first or breadth first strategy. Choosing a search term could be solved with the problem solving strategies brainstorming or trial and error, for instance. Another problem solving process within the IPS process, and the process that will be the focus of this study is evaluating information and source. In the next section the importance of evaluating will be discussed.

Information-problems can be solved with information from various sources. In the last decade, the World Wide Web (WWW) has become one of the largest and best accessible sources of information. And although there are many (more reliable) sources to collect information (e.g., the library or an encyclopaedia) the most probable source a student would use is the WWW (Beljaarts, 2006). Searching the Web for information differs from searching a library database or a table of contents. The Web does not have an index, the amount of information is enormous and there are no gatekeepers that filter information. Research has shown that young children, university students and adults encounter difficulties when searching for information on the Web (Brand-Gruwel et al., 2005; Duijkers et al., 2001; Hirsch, 1999; Kafai & Bates, 1997; MaKinster et al., 2002; Monereo et al., 2000). A problem for every age group is evaluating search results, and evaluating and selecting information and source (Duijkers et al., 2001; Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Koot & Hoveijn, 2005; Lorenzen, 2002; Lyons et al., 1997; MaKinster et al., 2002; Wallace, Kupperman et al., 2000).

Evaluating results (which site am I going to open) is mostly done based on titles and summaries provided by the search engine, the number of results and the order of results (Duijkers et al., 2001; Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Koot & Hoveijn, 2005; Lyons et al., 1997; Wallace, Kupperman et al., 2000). Students' prior knowledge influences the criteria used to evaluate results. Students with a lot of prior knowledge evaluate results by title, origin, summary, and identifiers in the URL (.edu or .gov) (Makinster et al., 2002).

Evaluating and selecting information and source (which information from which site am I going to use?) is not always done based on clear criteria (e.g., I see this is the official website of the National Health Organisation), but on intuition (e.g., I guess it is ok) (Koot & Hoveijn, 2005). Young children tend to believe that everything on the Web is true (Hirsch, 1999; Schacter et al., 1998), especially when they find the same information on more sites (Koot & Hoveijn, 2005). Teenagers use information that can solve their information-problem without thinking about the purpose of a site (Fidel et al., 1999). They also find it hard to express how they evaluate and select information (Lorenzen, 2002), and 82 % of the students admit that they rarely check information (Beljaarts, 2006).

An explanation for this lack of evaluating could be that students do not have enough metacognitive skills. As said before, problem solving skills call upon metacognitive skills (Perkins & Salomon, 1989; Van Merriënboer, 1997, 1999). Research has indeed shown that students rarely use metacognitive strategies in the correct way (Bannert, 2004; Gerjets & Scheiter, 2003; Rouet & Eme, 2002; Stadtler & Bromme, 2004).

So, it can be concluded that students hardly evaluate results, information and source. However, little research has been done on criteria students mention while solving information-problems and if these criteria depend on the domain of the problem or if they are general. This study tries to fill this gap.

The research question is as follows:

How do students solve information-problems and what kind of criteria do students use when evaluating information and source while searching the WWW for information?

This general question can be divided into several sub-questions:

- How much time do students spend on the constituent skills search, scan, process and organize while searching the World Wide Web and how often to they use these skills?
- How often do students evaluate their search results, information on a website, and the source, the website itself?
- Which criteria do students use for evaluating search results, information and source?
- Are the criteria general or do they depend on a certain domain?
- Does students' evaluation behaviour influence task performance?
- Are students aware of the criteria they use or can use?

Method

Participants

Twenty-three students of two schools for secondary education participated in this study. All students (8 boys and 15 girls; mean age 14.22, SD 0.422) were in the 9th grade of secondary education (Dutch: VWO, 6 year program).

Material

Tasks

Teachers of the two participating schools created twelve information-problems (4 physics, 4 geographic and 4 language and culture tasks). The problems consisted of a question, which had to be answered with information found on the Web. To answer this question, students had to choose which information to use and evaluate if the information could help solve the problem. To ensure comparability between problems, teachers received instruction on how to design them. For instance, it should not be a fact-finding question, the keywords leading to usable sites should not be mentioned, and solving the problem should require a search on the Web for preferably 30 min or more.

Furthermore, to ensure comparability even more, constructs that define the difficulty levels of prose-task processing - type of information requested, type of match, and plausibility of distracters (Mosenthal, 1998) - were controlled for. *Type of information requested* refers to the kind of information that readers must identify. Examples are persons, amounts, goals, cause, effect (or outcome, result), evidence, opinion, explanation, equivalence and difference. Tasks that require the identification of cause, effect, equivalence and difference are more difficult to answer. *Type of match* refers to 'the processes used to relate information in a question or directive to corresponding information in a text and to the processes used to select an answer from a range of response options' (Mosenthal, 1998, p. 285). Type of match has several levels. The first level (locate tasks) requires students to match one or more features in a question to one or more features in a text. An example of a locate task is 'how many democrats have been president of the United States of America between 1980 and 2007?' This question gives the student information about the answer that is needed, the number of democratic presidents between 1980 and 2007. With tasks of the fourth level (generate tasks) students are required to make inferences, match given information in a question to corresponding information in a text and identify the correct requested information. Not all information is in the question. An example of a fourth level task is 'Discuss whether the title of the article represents a good summary of the article's main point'. Students do not know how many points they need to discuss and they have to decide for themselves what the main point of the article is. The tasks used in this study are fourth level tasks: usable keywords are not mentioned in the tasks and the tasks do not specify the number of responses required. *Plausibility of distracters* refers to possible different interpretations or definitions of features from a question or information. An example: when looking for information on the painter Francis Bacon, you come across a very informative website with information about his life and death. After reading it, you discover that this is a site about Francis Bacon, the philosopher. With a high plausibility of distracters a task becomes more difficult to solve. The tasks in this study were tested by using several keywords. With each keyword a number of sites came up that had the keyword either in the title or the summary but did not include usable information.

After the tasks were designed, a panel of four persons tested them. They solved the problems and filled out a questionnaire after each task, concerning time on task, keywords and websites, and difficulty level. Also the construct 'type of information

requested' defined by Mosenthal (1998) was tested by the panel. Panel member 1 solved tasks 1-7, member 2 solved 1-12, member 3 solved 1-6 and member 4 solved 7-12. Table 1 summarizes the results for difficulty level (5-point Likert scale: simple to difficult), type of information and time on task and Table 2 shows the mean difficulty level and SD for all tasks and per domain, and the mean and SD of time on task of all tasks and per domain.

Table 1 Results of the test panel concerning difficulty level, type of information requested and time on task

Task (domain)	**Panel member**	**Difficulty level**	**Type of information requested**	**Time on task** (min)
1 Kyoto (geography)	1	Difficult	D	30
	2	Fairly difficult	D	15
	3	Fairly difficult	D	9
2 French dialects (language and culture)	1	Fairly simple	B/D	21
	2	Fairly simple	D	20
	3	Fairly simple	A/D	12
3 Great Britain colonial history (language and culture)	1	Difficult	D	30
	2	Difficult	D	30
	3	Fairly difficult	D	11
4 Road reflection (Physics)	1	Easy	D	30
	2	Easy	D	5
	3	Fairly Difficult	D	14
5 Coal (geography)	1	Difficult	D/E	30
	2	Difficult	D	25
	3	Easy	D	10
6 Resemblance English and French language (language and culture)	1	Fairly difficult	D	30
	2	Fairly difficult	D	20
	3	Easy	D	5
7 Hail (Physics)	1	Simple	D/B	30
	2	Easy	D	15
	4	Fairly difficult	D	10
8 Missing children (language and culture)	2	Difficult	D	30
	4	Difficult	D/C	20
9 Electron (Physics)	2	Easy	D	10
	4	Easy	D	11
10 Pollution and water (geography)	2	Fairly simple	D	15
	4	Fairly difficult	D	28
11 City patterns (geography)	2	Fairly difficult	D	30
	4	Difficult	D	17
12 Snow (Physics)	2	Fairly difficult	D	25
	4	Difficult	D	10

Note: Type of information requested:
A = Identification of person, group, location or object
B = Identification of amount, time, attribute, species, actions, locations or procedures
C = Identification of goal, function, alternative, attempt, condition, order, verification or problem
D = Identification of cause, effect, result, evidence, opinion or explanation
E = Identification of equivalence, difference, theme

Table 2 Means and standard deviations of difficulty level and time on task

Tasks	N	Difficulty level Mean (SD)	Time on task Mean (SD)
All	12	3.8 (0.9)	19.3 (4.0)
Geography	4	4.0 (0.7)	21.2 (2.3)
Physics	4	3.4 (0.8)	15.7 (3.5)
Language and culture	4	3.8 (1.3)	21.2 (3.7)

Table 1 shows some small differences between panel members on time on task and difficulty level, but all panel members agree on type of information requested. Difference in time on task can be explained by differences in prior knowledge of panel members. A univariate analysis of variance on difficulty level and time on task showed no significant differences between the domains. Based on the remarks of the panel, some tasks were adjusted; the task on Britain's colonial history was removed, because it was found unclear, and substituted by a task on MSN and SMS.

Table 3 presents the 12 tasks as used in the study. 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, try to answer the question in a few sentences, using the information you found. You have 30 min for this task. Think aloud during your search'.

WWW knowledge Questionnaire

A questionnaire was used to obtain students' knowledge and conceptions about the WWW. The questionnaire consisted of three parts. The first part obtained additional information on name, age, sex, nationality, grades on Dutch, English, French, physics and geography. In the second part, students were asked how much time they spent on the WWW per day, what they used the WWW for, and what their three favourite websites are. The third part consisted of 13 7-point Likert-scale items. Six items asked about the importance and usability of the WWW (e.g., I think it is useful to learn how to work with the Internet), $\alpha = .84$. Seven items asked about the expectations about their own knowledge on the WWW (e.g., I am sure I can learn how to work with the Internet), $\alpha = .87$.

Post hoc group interviews

A day after students completed the tasks, they returned in groups of three. In total there were eight focus groups. These interviews were used to obtain students' knowledge and conceptions on criteria for evaluating information and source.

Procedure

During the individual sessions participants first filled out the WWW knowledge questionnaire, and then received an instruction on how to think aloud. After this instruction, each participant conducted two tasks. With 23 students and 12 tasks, this means that -except for the dialect task- all tasks were made four times. The dialect task

was solved only twice. After reading the first task, participants 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?'). Participants returned in focus groups the next day for approximately 1 h. In every group, three tasks were discussed (one geography, one physics, one language and culture). This means that two students per group made each task. They received a booklet with screen shots of websites they had opened and had or had not used for their answer. They were asked to write down why they had or had not used the sites. This was followed by a group discussion. First, students were asked what they had written down and what they paid attention to while searching for information. This was done per task. Then a general discussion followed, where students were asked to come up with criteria they found important while searching. If students could not think of more criteria they were given a few criteria and asked whether they thought they could be important. They were also asked if they had paid attention to these criteria while searching the day before or if they ever paid attention to these criteria. Finally, students were asked if the criteria were the same for the courses language and culture, geography and physics. These sessions were also recorded.

Table 3 The tasks used in the study

Number	Title	Course	Task (translated from Dutch)
1	Kyoto	Geography	Ecological changes have become clearly visible during the last years. The snowline of the Kilimanjaro is moving and the ice of the polar caps is melting. Can all this be prevented if the Kyoto protocol is followed?
2	Dialect	Language and Culture	There are several dialects within the French language. Which dialects exist and what are the differences between these dialects?
3	MSN & SMS	Language and Culture	Young people used msn and sms a lot. Does this affect their Dutch?
4	Road reflection	Physics	Under which circumstances does the road reflect like a mirror in the distance?
5	Coal	Geography	Why does the Netherlands have charcoal layers of 1-2 metres thick, situated far below the ground, while the USA has layers 40 metres thick situated near the surface of the earth?
6	French and English	Language and Culture	How can the equivalences between the English and French vocabulary be explained?
7	Hail	Physics	Why are hailstones sometimes small and sometimes big?
8	Missing children	Language and Culture	How effective is searching for missing children by placing their pictures on milk cartons?
9	Electron	Physics	An electron has a charge, but what is this charge exactly?
10	Pollution	Geography	What is the effect of pollution on the quality of tap water?
11	City patterns	Geography	You have probably never thought about it, but many cities have certain city patterns. European cities are built different than cities in the USA. What are the differences in city patterns?
12	Snow	Physics	Why is there so much air in a snowflake?

Data analysis

All think-aloud protocols were transcribed verbatim. Then, all utterances and actions were coded. The coding scheme to analyse the thinking aloud protocols was developed in earlier studies (Brand-Gruwel et al., 2005; Wopereis, Brand-Gruwel, & Vermetten, 2005). An inductive-deductive method was used to develop this system. This means that the coding system has an empirical and theoretical grounding; it has been tested and re-adjusted in a few iterations. For scoring the protocols two kinds of codes were used: descriptive codes and interpretative codes (Miles & Huberman, 1994). Descriptive codes entail little interpretation and can be linked to segments of the text in a straightforward way. Interpretative codes require more interpretation by the rater. The scoring system itself consisted of three types of categories. The first category, the constituent skills, was scored in an exclusive and exhaustive way. Also the time invested in the constituent skills was recorded. The second category consisted of the sub skills of each constituent skill. The third category entailed regulation activities. Overall, the variables measured with this coding system are the constituent and sub skills, and regulation activities of the information-problem solving using Internet model (IPS-I model) based on Brand-Gruwel et al. (2005). The sub skills evaluate search results, evaluate information and evaluate source were enriched for this study with criteria derived from literature on evaluating websites and information (Barker, 2005; Beck, 1997; Boekhorst, 2000; Kirk, 1996; Ormondroyd, 2004). Table 4 shows the constituent and sub skills and Table 5 shows the evaluation criteria of the coding system. Table 4 mentions all sub skills of the entire IPS-process, in this study only the sub skills evaluate search results, evaluate information and evaluate source are discussed.

Table 4 Thinking aloud protocols: constituent and sub skills scored

Constituent skill	Sub skills
Define information problem	Read task
	Formulate questions
	Activate prior knowledge
	Clarify task requirements
	Determine needed info
Search information	Select search strategy
	Specify search terms
	Evaluate search results
Scan information	Read information global
	Evaluate scanned information and source
	Store relevant information
	Elaborate on content
Process information	Read in depth
	Evaluate processed information
	Store relevant information
	Elaborate on content
Organize and present information	Formulate problem
	Structure relevant information
	Outline the product
	Realize the product
	Elaborate on content

Two raters individually scored 6 of the 46 protocols. Interrater reliability was calculated. The similarity between the two raters, expressed in Cohen's Kappa was for all the protocols higher than .80.

The task performance (answer on the question) of the participants was judged based on three criteria: (1) Answer contains information of more than one source. Zero points for only one site, one point for more than one site, (2) Quality of the sources is judged based on the judging criteria in Table 5. Quality can be zero to four points, (3) Suitability of information used was determined by identifying content related elements. Suitability can be zero to four points. Maximum score for task performance is nine.

The post hoc group interviews were also transcribed verbatim. Criteria mentioned by the students were grouped and labelled, based on the criteria used in the coding system (see Table 5). The criteria were grouped in four categories. The first category consisted of criteria mentioned by students, and the second of students' reactions to criteria mentioned by the researchers. The third category covered students' reactions to the question if the use of criteria differed between domains and the last category consisted of students' reactions to the question about the difference in use of criteria during the tasks and the use of criteria while searching in general.

Table 5 Thinking aloud protocols: scored evaluation criteria

Sub skill	**Criteria**	
Evaluate search results	1.Title/Summary	What is the title given by the search engine and what is the contents of the summary?
	2.Kind (site/PDF/)	What kind of source is it, a website, a word document, a PDF file?
	3.Address	What is the address? Is it a .com or .org address?
	4.Rank in hit list	How many results are there in total and what is the rank of the result I am evaluating?
	5.Known to user	Have I used this site before, or have I heard good or bad things about it?
	6.Language	Is the site in a language I prefer and/or understand?
Evaluate information	**A Usability**	
	1. Language	In what language is the information written? Are there many grammatical or type errors? Is it filled with domain specific language?
	2. Connection to task	Does the information answer (part of) the information problem?
	3. Audience	Is the information aimed at a specific group of readers?
	4. Topicality	Is the information up to date?
	5. Amount	Is there enough information on the page? Or only a part of the information I'm looking for?
	B Verifiability	
	1. Author	Who has written the information? Can I contact him/her?
	2. References	Are there references on the page to used sources? Or links to more websites on the same subject?
	3. Information agrees with more sites	Can I find the same information on more pages or is this information only available on this site?
	4. Information agrees with prior knowledge	Does the information confirm what I already know?
	5. Organization	Which organization is behind the information? A governmental organization or a health organization? Can I find their logo on this site?

	C. Reliability	
	1.Kind of information	What kind of information is it? A newspaper article or a forum? Is it an opinion or results from research?
	2. Objectivity	Is the information objective or coloured by a certain point of view? Are there a lot of advertisements on the page?
	3. Primary/Secondary	Is the information first hand or is it someone telling about someone who did something?
	4. Goal	What does the (author of) information want to achieve. Sell something? Convince me of something or just inform me?
Evaluate source	**A Technical**	
	1. Appearance	Does the site appeal to me? Does it have pictures or only text?
	2. Speed	Does it take a lot of time to load the page?
	B Usability	
	1. Language	In what language is the site written? Are there many grammatical or type errors?
	2. Connection to task	Does the site have a connection to (part of) my information problem?
	3. Audience	For whom is the site meant? Who are its visitors?
	4. Topicality	Is the site updated regularly?
	C Verifiability	
	1. Reputation	Is this site famous or infamous for something? Does it have a good/bad reputation?
	D. Reliability	
	1. Kind (site/PDF)	What kind of source is it, a website, a word document, a PDF file?
	2. Primary/secondary	Is the site an original source or a site telling about what is written somewhere else?

To answer the research questions one-way ANOVAs were performed to determine if there was a difference in the amount of time spent on constituent skills, the frequencies of the used constituent skills, the frequencies of the sub skills evaluate results, evaluate information and evaluate source between the domains (language and culture, geography and physics). One-way ANOVAs were performed on the evaluation criteria, the total product score, the use of more than one site, the quality of sources, and suitability of information used to test if there was a difference between domains. To control for task order, one-way ANOVAs on the amount of time spent on constituent skills, the frequencies of the used constituent skills, the frequencies of the sub skills evaluate results, evaluate information and evaluate source were also performed to determine the difference between the first and second task. These were all not significant. All ANOVAs were performed with an a of .05. The ANOVAs are performed to test separate hypotheses, therefore a Bonferroni correction is not necessary (Tutzauer, 2003).

Results

First analyses were performed to determine whether students were comparable in terms of Internet experience. The majority of the participants (82.6%) spent 1-5 h a day on the Internet. The other 17.4% spent less than 1 h a day on the Internet. The mean score on

the importance and usability of the WWW was 6.3, SD 0.70 and the mean score on the expectations about their own knowledge on the WWW was 5.5, SD 0.80. This means students find the WWW very usable and important to work with and expect that they know enough about the WWW to be able to work with it.

Students' information problem solving processes

Figure 1 presents the average amount of time (in percentages of total time spent) students spent on searching, scanning, processing and organizing while performing the tasks. The constituent skill 'define information-problem' is not included since all students were asked to read the task and write down prior knowledge before starting their search. This prescribed action that can be seen as defining the problem is not scored.

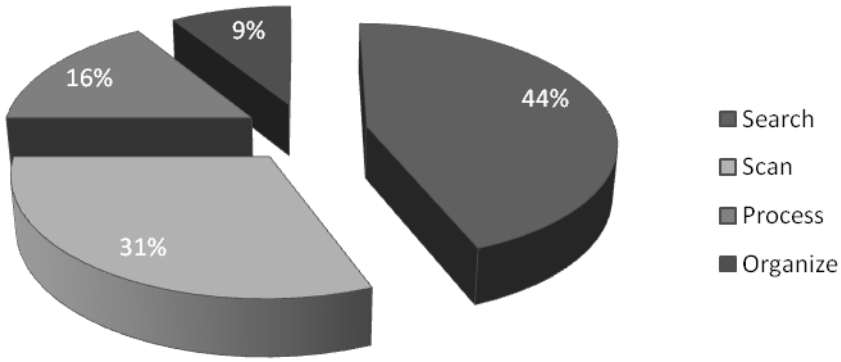


Figure 1 Time (%) spent per constituent skill

Figure 1 shows that students spent 44 % of the time on searching, 31 % on scanning, 16% on processing and 9 % on organizing information. Table 6 shows the means and standard deviations per constituent skill for all tasks and for tasks per domain.

Table 6 Means and standard deviations of time spent on constituent skills

Tasks	N	Search Mean (SD)	Scan Mean (SD)	Process Mean (SD)	Organize Mean (SD)	Total Mean (SD)
All	46	10.7 (6.7)	7.0 (4.1)	3.9 (4.8)	1.7 (2.1)	23.9 (7.1)
Geography	16	11.3 (6.1)	8.2 (3.9)	4.2 (5.4)	2.2 (2.5)	26.5 (4.2)
Physics	16	9.4 (7.4)	4.8 (2.8)	4.3 (4.9)	1.4 (2.0)	20.4 (8.9)
Language and culture	14	11.6 (7.0)	8.4 (4.7)	3.1 (4.1)	1.4 (1.6)	25.1 (6.1)

Two-way ANOVAs with domain and task order as between factors and time as a within factor with four levels, were performed to test if the amount of time spent on each constituent skill differed within subjects and between geographic, physics and language and culture tasks and between the first and second task. Results showed no

significant interaction between task order and domain. There was a significant difference within subjects on the amount of time spent on each constituent skill, $F(3,120)=23.51$, MSE=10,796.50, $p=.00$, $\eta^2= .37$. Students spent significantly more time on searching than on scanning, $F(1,40)=5.30$, MSE=6438.03, $p=.03$, $\eta^2= .12$. They spent significantly more time on searching than on processing $F(1,40)=26.05$, MSE=34,596.37, $p=.00$, $\eta^2= .39$, and organizing $F(1,40)=57.30$, MSE=52,939.53, $p=.00$, $\eta^2= .59$. Students also spent significantly more time on scanning than on processing $F(1,40)=10.77$, MSE=11185.95, $p=.00$, $\eta^2= .21$, and organizing $F(1,40)=46.56$, MSE=22454.59, $p=.00$, $\eta^2= .54$. There was no significant difference between the amounts of time spent on the skills process and organize.

The IPS-process turned out to be iterative and students switched frequently between the constituent skills. After an initial search, they either selected a site from the hit list, or they adjusted their search term and started another search. This can be seen in Table 7. If participants would follow the cycle search-scan-process-organize-search the frequencies would be equal. Since the frequencies of search and scan are high, it can be assumed that many iterations are made between these processes. Table 7 shows the means and standard deviations of the used constituent skills (frequencies) calculated over all 46 tasks (23 students * 2 tasks) and on the four tasks per domain.

Table 7 Frequencies of constituent skills

Tasks	N	Search Mean (SD)	Scan Mean (SD)	Process Mean (SD)	Organize Mean (SD)
All	46	14.5 (6.8)	12.8 (6.9)	2.9(2.9)	1.6 (1.8)
Geography	16	16.6 (7.5)	15.1 (8.7)	2.9 (3.0)	1.4 (1.4)
Physics	16	14.4 (7.4)	12.4 (6.3)	3.6 (3.4)	1.5 (2.1)
Language and culture	14	12.2 (4.8)	10.5 (4.2)	2.1 (2.1)	2.0 (1.9)

Two-way ANOVAs with domain and task order as between factors and frequencies as a within factor with four levels, were performed to test if the use of constituent skills differed within subjects and between domain and task order. Results showed no significant interaction between task order and domain. There was a significant difference within subjects on the use of each constituent skill, $F(3,120)=97.39$, MSE=1921.38, $p=.00$, $\eta^2= .71$. Students searched significantly more than scanned $F(1,40)=19.60$, MSE=132.07, $p=.00$, $\eta^2= .33$, processed $F(1,40)=116.32$, MSE=5954.78, $p=.00$, $\eta^2= .75$, or organized $F(1,40)=135.69$, MSE=7208.01, $p=.00$, $\eta^2= .77$. Students also scanned significantly more than processed $F(1,40)=69.04$, MSE=4313.21, $p=.00$, $\eta^2= .63$ or organized $F(1,40)=99.81$, MSE=5388.70, $p=.00$, $\eta^2= .71$. Finally, students significantly processed more than they organized $F(1,40)=6.32$, MSE=59.80, $p=.02$, $\eta^2= .14$.

So, it can be concluded that the process is iterative, especially between the search and scan processes. The fact that so many iterations are made between these processes makes it even more interesting to determine the criteria students use while evaluating. How do students decide to select a certain site or to click it away? This process of evaluating occurs quite often, because of the reasonable number of iterations. In the next section results on students' evaluating processes are presented.

Students' evaluating processes

To answer the research questions how often students evaluate during the IPS-process, what criteria they use, and whether these criteria are domain specific, the protocols were analysed in depth. Table 8 shows the means and standard deviations of the frequencies of the sub skills 'evaluate search results' (i.e., evaluate the hit list), 'evaluate information' and 'evaluate source'.

Table 8 Frequencies of evaluating search results, information and source per task per constituent skill

Tasks	N	Evaluate results during search Mean (SD)	Evaluate information during scan Mean (SD)	Evaluate source during scan Mean (SD)	Evaluate information during process Mean (SD)	Evaluate source during process Mean (SD)
All	46	9.5 (6.6)	3.5 (3.2)	0.2 (0.5)	1.1 (1.7)	0.0 (0.2)
Geography	16	10.3 (6.8)	4.3 (3.7)	0.2 (0.4)	1.3 (1.7)	0.0 (0.0)
Physics	16	9.6 (7.3)	3.3 (3.1)	0.2 (0.5)	1.4 (2.0)	0.1 (0.3)
Language and culture	14	8.6 (6.0)	2.9 (2.8)	0.2 (0.4)	0.5 (0.9)	0.0 (0.0)

It should be noted that only the evaluations that were explicitly mentioned were scored. Students also clicked pages away without telling why. In those cases, it would seem they did evaluate the site, but it is not clear on what grounds. These evaluations are not included in Table 8.

It seems that students hardly evaluate information and source explicitly (neither during scanning the site nor during processing). Especially, sources are hardly evaluated. Evaluating the hit list to select a site that may give an answer to the problem is done most frequently. Furthermore, standard deviations are large, especially on the skills 'evaluate search results' and 'evaluate information while scanning sites'. Differences between students are large concerning the execution of these skills.

Two-way ANOVAs with domain and task order as between factors and kind of evaluations as a within factor with five levels were performed to test if the frequencies and kind of evaluations (evaluate search results, evaluate information during scan, evaluate information during process, evaluate source during scan, evaluate source during process) differed within subjects and between domain and task order. Results showed no significant interaction between task order and domain. There was a significant difference within subjects on the use of evaluations, $F(4,160)=69.54$, $MSE=726.83$, $p=.00$, $\eta^2= .64$. Participants significantly more often evaluated search results than information during scanning, $F(1,40)=34.92$, $MSE=1627.21$, $p=.00$, $\eta^2= .47$ and than source during scanning, $F(1,40)=89.51$, $MSE=4000.79$, $p=.00$, $\eta^2= .69$. Search results were also significantly evaluated more than the information during processing, $F(1,40)=77.58$, $MSE=3293.60$, $p=.00$, $\eta^2= .66$, and the source during processing, $F(1,40)=91.89$, $MSE=4139.47$, $p=.00$, $\eta^2= .70$. Participants significantly more often evaluated information during scanning than the source during scanning, $F(1,40)=65.93$, $MSE=525.02$, $p=.00$, $\eta^2= .62$, and information during processing, $F(1,40)=35.74$, $MSE=290.75$, $p=.00$, $\eta^2= .47$ and the source during processing $F(1,40)=65.46$, $MSE=576.00$, $p=.00$, $\eta^2= .62$.

Participants significantly more often evaluated source during scanning than the information during processing, $F(1,40)=13.32$, MSE=34.36, $p=.00$, $\eta^2= .25$, and source during processing, $F(1,40)=6.74$, MSE=1.18, $p=.01$, $\eta^2= .14$. Finally, the information during processing was significantly evaluated more than the source during processing $F(1,40)=18.57$, MSE=48.29, $p=.00$, $\eta^2= .32$.

Evaluation criteria

Although students hardly gave expression to their evaluation, it is important to determine which criteria they used when they evaluated results, information and sources. Table 9 shows the criteria as defined in the method section and the frequencies on how often students used these criteria for evaluating results, information and source.

Table 9 Frequencies of criteria used for evaluating search results, information and source

Sub skill	Criteria	Frequencies
Evaluate search results	1.Title/Summary	217
	2.Kind (site/PDF/)	0
	3.Address	0
	4.Rank in hit list	3
	5.Known to user	4
	6.Language	2
Evaluate information	**A Usability**	
	1. Language	10
	2. Connection to task	89
	3. Audience	0
	4. Topicality	0
	5. Amount	7
	B Verifiability	
	1. Author	0
	2. References	0
	3. Information agrees with more sites	0
	4. Information agrees with prior knowledge	2
	5. Organization	1
	C. Reliability	
	1.Kind of information	8
	2. Objectivity	0
	3. Primary/Secondary	0
	4. Goal	0
Evaluate source	**A Technical**	
	1. Appearance	0
	2. Speed	3
	B Usability	
	1. Language	0
	2. Connection to task	0
	3. Audience	0
	4. Topicality	0
	C Verifiability	
	1. Reputation	0
	D. Reliability	
	1. Kind (site/PDF)	1
	2. Primary/secondary	0

The students appeared to use only a few of the possible 29 criteria during information problem solving. They mainly evaluated search results or selected sites in the hit list based on title and summary. They read the title and the summary and decided if the information on the site could help them find an answer on the question and then decided to open the site or not. Only a few times was the rank in the hit list or language used as criterion. In total (46 tasks) the results were viewed 949 times and they were explicitly evaluated 217 times, a percentage of 24 %. The following description gives an impression of students' evaluation behaviour.

> *A student has just performed a search with Google, and is looking at the results. He points his mouse to the summary, reads it aloud: "French influence on Dutch dialects" and opens this site.*

The usability of the information found on a site was often evaluated with the criterion '*connection to task'*. This means that students sought for information that could help answer the question, so the criterion was the content. Sometimes the language (foreign, slang, spelling errors) was a reason for determining if information was usable. Criteria of verifiability were hardly used to evaluate information. On two occurrences a student connected the information to prior knowledge. Reliability of information was checked eight times with the criterion '*kind of information'*.

Information was evaluated 117 times out of 780 page views (15 %). The following description gives an impression of students' evaluation behaviour.

> *A student is trying to answer a task on French dialects. He opens a page, scans it, and while closing the page, says: "It doesn't say anything here about the dialects".*

The source was evaluated on speed and one time on the kind of source. Only 4 times out of 780 page views the source was judged, a percentage of only 0.5%. The following description gives an impression of students' evaluation behaviour.

> *A student has opened a page and it is loading. After a minute, only the coloured background has appeared. The student says: "This takes too much time" and closes the page.*

Students also evaluated results, information and source without mentioning criteria. Utterances like 'these are good results', 'this is a nice website', are examples of these kinds of evaluations. In total there were 316 of these undefined evaluations. Since the focus in this study is on criteria mentioned by students, these undefined evaluations are not included in the rest of the result section. We will come back to them in the discussion.

Criteria for evaluating: domain specific

Table 10 presents the means en standard deviations of the used criteria per domain.

Table 10 Means and standard deviations of criteria use

Sub skill	Criteria	All (N=46) Mean (SD)	Geography (N=16) Mean (SD)	Physics (N=16) Mean (SD)	Language and culture (N = 14) Mean (SD)
Evaluate search results	Title/Summary	4.7 (5.1)	5.7 (5.4)	4.2 (4.6)	4.2 (5.5)
	Rank in hit list	0.1 (0.2)	0.2 (0.4)	0.0 (0.0)	0.0 (0.0)
	Known to user	0.1 (0.3)	0.1 (0.3)	0.1 (0.3)	0.0 (0.0)
	Language	0.0 (0.2)	0.0 (0.0)	0.1 (0.3)	0.0 (0.0)
Evaluate information	Language	0.2 (0.6)	0.2 (0.4)	0.3 (0.8)	0.1 (0.5)
	Connection to task	1.9 (2.0)	2.4 (2.7)	2.1 (1.8)	1.2 (1.2)
	Amount	0.2 (0.4)	0.3 (0.4)	0.1 (0.3)	0.1 (0.3)
	Information agrees with prior knowledge	0.0 (0.2)	0.1 (0.3)	0.1 (0.3)	0.0 (0.0)
	Organization	0.0 (0.1)	0.0 (0.0)	0.1 (0.3)	0.0 (0.0)
	Kind of information	0.2 (0.7)	0.1 (0.5)	0.1 (0.5)	0.3 (1.1)
Evaluate source	Speed	0.1 (0.2)	0.1 (0.3)	0.0 (0.0)	0.1 (0.4)
	Kind	0.0 (0.1)	0.0 (0.0)	0.1 (0.3)	0.0 (0.0)

Two-way ANOVAs with domain and task order as between factors and used criteria search results as a within factor with four levels, two-way ANOVAs with domain and task order as between factors and used criteria information as a within factor with six levels and two-way ANOVAs with domain and task order as between factors and used criteria source as a within factor with two levels were performed, to test if the use of criteria differed within subjects and between domain and task order. Results showed no significant interaction between task order and domain. There was a significant difference within subjects on the use of search results criteria, $F(3,120)=33.67$, MSE=242.88, $p=.00$, $\eta^2= .46$. The search results were significantly evaluated more on "*title/summary*" than on "*rank in hit list*", $F(1,40)=33.04$, MSE=967.78, $p=.00$, $\eta^2= .45$, "*known to user*", $F(1,40)=34.39$, MSE=966.68, $p=.00$, $\eta^2= .46$, and "*language*" $F(1,40)=33.99$, MSE=980.02, $p=.00$, $\eta^2= .46$.

There was a significant difference within subjects on the use of information criteria, $F(5,200)=34.82$, MSE=25,87, $p=.00$, $\eta^2= .47$. The information was evaluated significantly more on "*connection to task*" than on "*language*" $F(1,40)37.33$, MSE=137.03, $p=.00$, $\eta^2= .48$, "*amount*" $F(1,40)=41.10$, MSE=148.53, $p=.00$, $\eta^2= .51$, "*information agrees with prior knowledge*", $F(1,40)=43.55$, MSE=167.79, $p=.00$, $\eta^2= .52$, "*organization*", $F(1,40)=43.45$, MSE=171.03, $p=.00$, $\eta^2= .52$ and "*kind of information*", $F(1,40)=45.00$, MSE=144.64, $p=.00$, $\eta^2= .53$. Information was also evaluated significantly more on "*amount*" than on "*information agrees with prior knowledge*", $F(1,40)=6.79$, MSE=0.59, $p=.01$, $\eta^2= .16$, and "*organization*", $F(1,40)=5.00$, MSE=0.79, $p=.03$, $\eta^2= .11$. Information was evaluated significantly more on "*language*" than on "*organization*", $F(1,40)=5.25$, MSE=1.88, $p=.03$, $\eta^2= .12$.

There was no significant difference in the use of criteria to evaluate sources.

Students' products

Evaluating is crucial when selecting information, since the World Wide Web contains so much information that is incorrect, subjective or biased. If one uses information for an

essay or answer to an information-problem without questioning the source or considering if the information is correct it is possible that the end product is not optimal. In this section, the quality of the students' answers to the questions asked in the tasks are analysed and related to used criteria.

To answer the questions about the quality of the solved problems and the relation to the used criteria, students' answers were analysed. Table 11 shows the means and standard deviations of the total product score and the scores on quality of sources and suitability of the information.

Table 11 Means and standard deviations of the product score (Maximum = 9)

Tasks	N	Total Product Mean (SD)	Using more than one site (maximum = 1) Mean (SD)	Quality of Sources (maximum = 4) Mean(SD)	Suitability of Information (maximum = 4) Mean(SD)
All	46	4.1 (1.7)	0.8 (0.4)	1.8 (0.8)	1.5(1.1)
Geography	16	3.6 (1.4)	0.8 (0.4)	1.9 (0.9)	0.9 (0.6)
Physics	16	5.1 (1.8)	0.8 (0.4)	2.0 (0.8)	2.3 (1.2)
Language and culture	14	3.5 (1.3)	0.9 (0.4)	1.6 (0.6)	1.1 (0.9)

The mean scores on all the aspects are low. The maximum possible total product score is 9 and the mean is 4.1. The maximum achieved total score is 7. The maximum (possible and achieved) score on quality of the answer is 4 and the mean is 1.6. Thus, student scores are far from optimal.

A one-way ANOVA shows that the subject area (geography, physics and language and culture) has a significant influence on the total product score, $F(2,45)=5.23$, MSE=12.22, $p=.01$, $\eta= 0.20$ and the suitability of information, $F(2,45)=10.23$, MSE=9.26, $p=.00$, $\eta= 0.32$. The four physics tasks resulted in significantly higher scores than the task from other domains. Posthoc analysis showed that there was no significant difference between the four physics tasks. One-way ANOVAs on differences between the quality of the sources and the score on the use of more than one source did not yield significant results.

Regression analyses with the number of evaluations of search results, information and sources as predictors and the total product score and suitability of information as dependent variables were not significant. Regression analyses with the used criteria as predictors and total product score and suitability of information as dependent variables were not significant either.

Students' awareness of evaluation criteria

As shown, students do not use a lot of criteria for evaluating results, information and source. Is that because they are not aware of these criteria or because they do not feel the need to use or explicitly express them? In this paragraph the result of the group interviews will be discussed.

Table 12 shows the criteria and the number of groups that mentioned these criteria during the group interviews. Only two criteria are mentioned by all eight groups;

connection to task and *information agrees with more sites*. More than half of the groups mentioned criteria like *title/summary, language* and *appearance*. Four or fewer groups mentioned criteria like *audience*, *author* and *goal*. Comparing Table 12 to Table 9 (the criteria actually used during the tasks) makes it clear that students mentioned more criteria during the group interviews than they explicitly used during the tasks. None of the students mentioned *information agrees with more sites* during the tasks but this criterion is mentioned during all interviews. Criteria like *audience*, *author*, *references* and *goal* are not mentioned while solving the tasks, but are mentioned by some groups. Criteria like *topicality*, *objectivity* and *primary/secondary* are neither used nor mentioned.

The results on students' use of criteria are nuancated by what the students seem to know about criteria when asked about criteria. They seem to know more criteria than they explicitly use. This difference could be explained by the fact that students do not only use criteria they know (and explicitly mention when evaluating) but are sometimes triggered to use a criterion. When triggered to use a criterion students are not always aware of using the criterion, and do not mention it explicitly, but when asked about the criterion later, they do have knowledge about this criterion. For instance the criterion *author* is only important if the author is mentioned on the website. When the author is not mentioned, they do not try to find out who the author is and it is not a reason to doubt the information. This could imply that the criterion author is not a criterion that is used systematically by students, but its use is triggered. However, this triggered use, could lead to use of the criterion based on misconceptions. One student said he did not check for the author because the author was not mentioned and he did not know the website. This implies, that he would only check the author if the author was mentioned. In fact, correct use of the criterion would be to doubt a website that does not mention an author. To some students, the author is not important, as long as they understand the information and the information agrees with more sites.

Their view on the use of the criterion *information agrees with more sites* is rigid. If two sites contain the same information, this information is regarded as true. Of course, this way of using the criterion could lead to the use of false or biased information, since some websites copy content of other sites. If the information is not checked by criteria like author, references and goal, the information could agree with more sites, but may still be false or biased. Students do not always realize this, as shown by the following remark: "They made the effort to build a page, why would they put information on it that is not true?"

Furthermore, students were asked if the criteria they did not mention could be important. The criterion *topicality* was an eye-opener for most students. They admitted they hardly looked at the date a page was last updated. They agreed that *topicality* could be important, depending on the information problem. "If you have to find out where New York is located, that will be the same for a couple of years." The suitability of the other criteria also depended on the task and on the students' perceptions of teachers' expectations. As for the criterion *kind of information*, students admitted that they knew that information from a forum or an opinion of somebody is not always reliable, but if they cannot find anything else they will use the information anyway.

All students felt that the basic criteria (*connection to task* and *information agrees with more sites)* were important no matter what the domain was. The other criteria depended on the task, and not really on the domain, in their view.

Table 12 Criteria mentioned without prompting during group interviews

Sub skill	Criteria	Number of groups (maximum = 8)
Evaluate search results	1.Title/Summary	5
	2.Kind (site/PDF/)	0
	3.Address	2
	4.Rank in hit list	5
	5.Known to user	3
	6.Language	0
Evaluate information	**A Usability**	
	1. Language	5
	2. Connection to task	8
	3. Audience	2
	4. Topicality	0
	5. Amount	4
	B Verifiability	
	1. Author	4
	2. References	1
	3. Information agrees with more sites	8
	4. Information agrees with prior knowledge	1
	5. Organization	2
	C. Reliability	
	1.Kind of information	4
	2. Objectivity	0
	3. Primary/Secondary	0
	4. Goal	2
Evaluate source	**A Technical**	
	1. Appearance	6
	2. Speed	0
	B Usability	
	1. Language	0
	2. Connection to task	0
	3. Audience	0
	4. Topicality	0
	C Verifiability	
	1. Reputation	4
	D. Reliability	
	1. Kind (site/PDF)	2
	2. Primary/secondary	0

The overall image created by these group interviews was that the students opened websites based on the summary given by the search engine and checked the information on connection to task. The information had to be easily traceable and had to agree with more sites. The author was not really important; the organization behind a

site could be a clue for the reliability. But if the organization was not mentioned, that was no reason to discard the information, especially when the information was found on more sites. Dutch sites were preferred and the information should be understandable.

Discussion

The central question of the study in the Chapter is how students solve information problems and what kind of criteria they use when evaluating results, information and source while searching the WWW for information.

While solving information problems students spent most of their time on searching and scanning and only a small amount of time on processing and organizing information. This finding is in line with results Brand-Gruwel et al. (2005) found. Ph-D-students and freshman also use more time on searching and scanning than on processing. Subjects in their study however, spent a substantial amount of time organizing and presenting the information, but this was due to the fact that they had to deliver an outline for an article.

During the IPS-process, students do not very often evaluate results, information and source. Only 24% of the times that search results are viewed they are evaluated explicitly, information is evaluated in only 15 % of the page views and sources are evaluated only 0.5 % of the views. When students evaluate the results of a hit list, they mostly use the title and the given summary. When evaluating information the criterion used most often is if the information is connected to the task. The criteria students mention, when asked which are important for evaluating information, are not always the same criteria they mention while solving the information problems. During the group discussion the criterion if the information can be found on more sites was mentioned as important. They mentioned more criteria but also admitted that they did not always use them while searching the Web. The reasons were time pressure, motivation and convenience.

Furthermore, the students did not mention differences between the domains. All criteria are applicable in the different domains, although type of task can make criteria less or more important.

If students do not evaluate search results, information or source, one may expect a less optimal product. Results show that students' task performance was far from optimal, and that the quality was rather low. However no relation between evaluation behaviour and task performance could be found. Students who use more criteria do not achieve significantly higher results. However, this could be due to the fact that although they used more criteria, they did not always use the more sophisticated criteria or used the criteria in a correct way.

One of the limitations of this study was the time pressure students felt during the experiment. Some students admitted they did not check information because of the time limit of 30 minutes. They felt they had to come up with an answer and were glad when they found information. They said they would pay more attention to the criteria

when they had to write an essay or when they were searching for something they were really interested in. This could mean they would achieve better results in those cases.

However, there is certainly room for improvement, because there are still criteria they had never thought about and it is important to make them aware of their use of criteria. It is not only important for schoolwork, but in every day life as well. Most striking was the fact that students expected to find an answer on one single website, preferably in the first couple of sentences. They did not bother to combine information or check on things. If the author of the site was not mentioned they did not try to find out who the author was. Students wanted the information served on a silver platter and did not want to do a lot of work themselves.

Another factor that may have influenced the results is the data collection procedure. Thinking aloud is not always easy for students. They discarded sites without explaining why and as mentioned before there were 316 utterances with undefined criteria. This could be an explanation for the difference between the used criteria and the criteria mentioned in de focus groups. 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) worked better than concurrent and retrospective reporting while trouble shooting with electrical circuits. Participants provided more action information, 'how' information and metacognitive 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. Perhaps this would result in a better explanation of why they discarded some sites in only a few seconds.

The results of this study agree with earlier studies in this field. Students have trouble evaluating information and do not have a critical attitude towards information on the WWW. Koot and Hoveijn (2005) also used think aloud sessions and interviews in their study of 11-year olds. Their focus was also on the children's view on the reliability of information. These young children approached the Web the same way as the high school students in our study. They also hardly evaluated results, information and source. They were aware that not all information on the Web is reliable, but seemed to forget this while searching. And just like our students, there was a discrepancy between what they did and what they reported. When asked, they were capable of mentioning rational arguments why information was reliable, but when searching they based their evaluations more on intuition. Lorenzen (2002) interviewed high school students about how they used the World Wide Web to find information for school assignments, how they went about finding information and how they knew if the information they found on the Web was scholarly or factual. Students admitted that they were not really sure how they could distinguish between good and bad information. They tended to trust institutional pages, knew to check spelling, bibliography and web pages author. However, they seemed to forget that the authors name on a site does not mean this person is real or has actually written the information on a site. It is also possible that a site may look like an institutional page, but it is not. A big problem was that students evaluated a site based on how elaborate it looks: 'if a web site looks good, appears to be professional, and has a lot of detail on it, many of the students will accept it as a

good web site for information. Clearly, this method of evaluation is weak and not really helpful' (Lorenzen, 2002, p. 161). The students in the study by Lorenzen (2002) seemed to approach the web in the same way our students did.

Finally, a poll by a research company (Beljaarts, 2006) also showed that students felt they could trust the Internet and hardly checked their information. The students in our study were not different from students in other age groups or students of their own age.

This attitude (not evaluating information critically) can cause problems in school and everyday life. It is important to make students aware of their evaluation behaviour, the misconceptions they may have and point them to the criteria they can use to evaluate information. This will help them achieve better results in school. Future research should aim on developing instruction in IPS, focusing on evaluating and selecting. Since IPS is a complex cognitive skill, the students should work on whole tasks, which are authentic and comprehensive. These tasks require from students to perform all the constituent skills that make up the whole complex skill during task performance (van Merriënboer, 1997). And since IPS is important throughout life, instruction should also focus on the transfer of the skill to multiple domains, tasks and situations.

CHAPTER 4

TEACHERS AND THE WORLD WIDE WEB

HOW TEACHERS EVALUATE SEARCH RESULTS, INFORMATION AND SOURCE

Teachers and the World Wide Web: How teachers evaluate search results, information and source

This study investigated the use of evaluation criteria and the influence of domain knowledge on evaluating search results, information, and source when searching the Internet for accomplishing school assignments. Eleven secondary educational teachers solved two information problems while thinking aloud. One problem was within their domain of expertise and the other problem was from a different domain. After task completion they were interviewed in groups concerning the used evaluation criteria. Results show that teachers used the criteria *title/summary* and *connection to task* to evaluate search results, information and source, and do not use the more sophisticated criteria like *author, topicality* and *audience*. Furthermore there is no difference in use of evaluation criteria between a task in their domain of expertise and a task outside their domain of expertise. Teachers use more domain knowledge when solving a task in their domain of expertise, but this domain knowledge does not influence task performance. Teachers do not find better information or sites in their domain of expertise. The group interviews revealed that teachers know more criteria than they use, and feel domain and domain knowledge have an influence in use of evaluation criteria.

This Chapter is based on Walraven, A., Brand-Gruwel, S, & Boshuizen, H. P. A. (submitted). Teachers and the World Wide Web: How teachers evaluate search results, information and source.

Introduction

Ten years ago, owning a cell phone was special. Nowadays, ten year olds are equipped with high-tech cell phones to call their friends for a play date. Fifteen years ago, the Internet was something you could only access at your local library. Presently, you can go online with your laptop or mobile phone wireless on almost every street corner. Society has changed from a knowledge society into an information and technology society. New tools and gadgets are being invented and launched onto the market every day. The best part of the vast growing collection of these new tools is not designed for educational use, but is being used for educational purposes nevertheless. The World Wide Web (WWW) is one of the best examples of such a tool.

The WWW was not originally designed for educational use, and its entry in education was not planned. Moreover, the human mind is not capable of working with tools like the WWW, without our minds being shaped to do so (Saljö, draft). So, using the WWW in education without sufficient knowledge on how to use the WWW in general can lead to problems which have not been analysed sufficiently. For example, teachers should have knowledge and skills on how to use the WWW and how they can help their students use the WWW.

The aim of this study is to determine whether secondary education teachers are equipped with the necessary skill to help their students use the WWW as a source of information for accomplishing educational tasks. First, the necessary information skills will be presented. Second, problems and difficulties concerning these skills will be addressed, and third, the role of the teacher in fostering these skills will be discussed, resulting in the research questions of this study.

Originally, the WWW was designed to make it easier for a small group of nuclear physics researchers to share information and it is now connecting people everywhere around the world. It has become the most used source of information for students (Beljaarts, 2006). In the days before the WWW, students used books as a source for information for their essays, nowadays they use the WWW. However, searching the WWW differs from searching the database of a library. First, the 'collection' of the WWW does not have an index like the library collection, which makes selecting keywords more difficult. Second, the WWW is also bigger than the average library, which makes it easier to 'get lost'. Third, the amount of information on the WWW that can be retrieved by a single keyword is bigger than in a library and is therefore harder to review or process. Fourth, the WWW is made up in hypertext, which makes a dynamic organization of information through links and connections possible. However, this also requires more regulation and cognitive capacity compared to traditional text. Students can become overwhelmed by the amount of information to process (Rouet, Levonen, Dillon & Spiro, 1996). Finally, the WWW does not have a board of editors, which decides whether the information is suitable for the collection. Anyone can post anything on the WWW and information can be incomplete, false or biased. This makes evaluating information an essential skill in the educational use of the WWW.

There are three things that need to be evaluated: 1) the search results, 2) information and 3) the source. There are several criteria that can be used to evaluate search results, information and source (Barker, 2005; Beck, 1997; Boekhorst, 2000; Kirk, 1996; Ormondroyd, 2004).

The *search results* are the websites presented by the search engine (e.g., Yahoo or Google) after a query. The results are often called a hit list. Evaluating these results will answer the question: which site am I going to open? Criteria that can be used for evaluating search results are the title and summary of the site, the kind, the address, the rank in the hit list, whether the result is known to the user and the language. The title and summary can provide an overview of the subject and content of the site. The criterion 'kind' can be used to divide results in sites, pdf-files or PowerPoint presentations. The address can give a clue concerning the owner of the site or file. For instance, an address with a .gov extension is owned by a governmental organization, while an address with a .com extension refers to a commercial organization. The rank in the hit list can provide information on popularity and connection to the query. The criterion 'known to user' could be used to open sites with a good reputation, or discard sites that the user has opened before and were not found very useful. If a site or file is written is a language the user does not speak, this result will probably be ignored.

After evaluating the results and opening a website or file, the *information* on that site or in the file has to be evaluated. This can be done 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 criterion 'language' does not only refer to the language of the site (e.g., English or French) but also to the quality of the language used. The criterion 'connection to task' is used to decide whether the site or file can provide an answer to the task at hand. If, for example the task was about the painter Francis Bacon and the site provides the biography of the philosopher Francis Bacon, the site has no connection to the task. Audience refers to the audience a site or file is meant for. A site can be aimed at a specific group of readers, children or experts for instance. The criterion 'topicality' can be used for checking when a site was last updated, or what the date of publication is. Information on a site can easily be superseded if the site is not regularly updated. The 'amount' of information refers to whether or not there is enough information on the page to answer the question.

The verifiability of information can be evaluated by the criteria author, references, information agrees with more sites, information agrees with prior knowledge and organization. Information on a site becomes more verifiable if the author of the information is known and can be contacted. It is important to check if there are any references provided of the sources used by the author, or if there are any links to other sites on the same subject. The criteria 'information agrees with more sites' and 'information agrees with prior knowledge' are used to check whether the information could be considered as common knowledge or is only the opinion of one author. The information could come from a certain organization, like a governmental or health organization.

The reliability of information can be evaluated by kind of information, objectivity, primary/secondary and goal. The information found could be a newspaper article, a

forum, or results from research. The kind of information could affect the reliability of the information. Information could be coloured by a certain point of view (e.g., a website about Martin Luther King created by a white power organization). A lot of advertisements on a page could mean the website is being sponsored and therefore less objective. The information could be first hand (primary), or it could be someone telling about someone who did something (secondary). The goal of the information could influence the reliability. Some information is used to convince its readers, or used to sell readers a certain product.

The *source* can also be evaluated. This can be done on technical, usability, verifiability or reliability grounds. Criteria for evaluating the technical side of a source are appearance and speed. Appearance refers to the look and feel of the site. The speed refers to the time it takes to load the page or file. The criteria for usability are language, connection to task, audience and topicality. These are the same criteria as for evaluating information. Verifiability has only one criterion: reputation. A site can be famous or infamous for something and have a good or bad reputation. Reliability of a source can be evaluated with the criteria 'kind of source' and 'primary or secondary source'. These are the same as for evaluating information.

It is not always necessary to evaluate search results, information and source with all criteria. Evaluating information based on one criterion can make using the other criteria redundant. For instance, if the information does not have a connection to the task at hand, it is not necessary to check if the information agrees with more sites, or who the author is.

Using these criteria and thus evaluating results, information and source can help avoiding the use of incomplete, false and biased information. However, research has shown that children, teenagers as well as adults are deficient in the process of searching, and especially in evaluating results, information and source (Brand-Gruwel et al., 2005; Britt & Aglinskas, 2002; Duijkers et al., 2001; Hirsch, 1999; Kafai & Bates, 1997; MaKinster et al., 2002; Metzger, Flanagin & Zwarun, 2003; Monereo et al., 2000). The question is what causes these deficiencies? (1) Do people lack the skills to evaluate, (2) do they not know that evaluation is necessary, or (3) do they not evaluate even though they have the necessary skills and are aware of the importance?

Britt and Aglinskas (2002) found that high school and college students did not spontaneously attend to source information when reading multiple documents. Therefore they developed the Sourcer's Apprentice, a computer application. This program taught sourcing, contextualization and corroboration in the context of researching a historical controversy. The Sourcer's Apprentice provided students with several documents about a controversy and information about the documents, such as author's credentials and possible motives. After reading the documents students filled out note cards. The note cards allowed students to write down information about six source and three content features like author (who, position, how know and author motives) and document (when, type). After filling out the note cards, students received a series of questions about the sources and contents of the documents and were asked to write an essay on the controversy. Results showed that the sourcing skills of students using the Sourcer's Apprentice had improved. The essays of the Apprentice group were more integrated, cited more sources and referenced more information from primary

and secondary sources than the essays of the comparison group. The study by Britt and Aglinskas (2002) clearly shows that evaluating information and source is by no means an automatism, even when author and date can be easily derived from the source. Deriving these features from a website can be even harder, and evaluation has to be done much more consciously (Britt & Gabrys, 2000).

The questionnaire used by Metzger et al. (2003), regarding students' use of Web-based information, their perceptions of information credibility across media, and their verification efforts, showed that many students (age 20) trust the Internet to provide accurate information, without taking the requisite steps to ensure the credibility of the information they obtain. However, if they do take the effort to evaluate information, they report that they do more than only checking the 'date stamp' on a Web page. They also check if the information is complete and comprehensive or if the information agrees with more sites.

Walraven, Brand-Gruwel and Boshuizen (in press) also found that high school students know criteria for evaluating search results, information and sources, but that they do not use them. Students use criteria like 'title/summary', 'language' and 'connection to task' when evaluating information. When asked which criteria students knew, criteria mentioned most by students were 'language', 'connection to task', 'amount', 'author', 'information agrees with more sites' and 'kind of information'. Students admitted that they did not use all the criteria they knew, because they were not asked to use them and would not be rewarded extra credits for using the criteria in a school context.

All these studies (Britt & Aglinskas, 2002; Metzger et al., 2003; Walraven et al., in press) suggest that students have at least some skill necessary for evaluating results, information and sources, but do not always see the need to use these skills, and often have to be prompted to use them. Britt and Aglinskas (2002) showed that not using evaluation skills can result in essays of less quality. Furthermore, not evaluating results, information and source can lead to the use of incorrect information and the building of incorrect knowledge.

Teachers are responsible for the development of evaluation skills in their students, as well as prompting them to use these skills, assessing students' work (e.g., essays) and thereby also assessing students' use of evaluation skills. For instance, in the domain of history, students are explicitly taught how to evaluate historical information and sources. But in other courses teachers do not explicitly teach evaluation skills. Furthermore, evaluation of information on the WWW is not taught, even in the domain of history. Teaching students how to evaluate search results, information and source on the WWW is necessary if the WWW will continuously be used by students for educational tasks. Before teachers can help their students develop evaluations skills, it is imperative that teachers have good evaluation skills themselves.

However, a study by Dirkx, Theuns and Timmers (2006) revealed that teachers are not always aware of the criteria they could use to evaluate search results, information and source themselves and how they could grade students' evaluations. Although the study by Dirkx et al. is only a pilot study with a small number of participants, it suggests that teachers themselves may not have the search and evaluation skills to help their students develop such skills. Participants (teachers) in this study were asked to grade an

essay, while paying extra attention to the correct use of information. The subject of the essay was not connected with the expertise of the teachers. A discussion on their experiences and grading made clear that although the teachers were asked to pay attention to information use, most teachers did not grade use of information or source. They also admitted that they did not know how they could grade the use of information and source. Moreover, teachers remarked that grading an essay in their own field of expertise would probably have led to even more content related grading of the essays, and less attention to information use. This might suggest an inverse relation between domain expertise and attention to information use in grading student papers.

Research on differences between novices and experts has indeed shown that experts and novices have different ways of solving information-problems. Lundeberg (1987) and Wineburg (1991) showed that domain experts paid more attention to the source of the document than domain novices. Law professors in the Lundeberg study (1987) paid much attention to the date and the name of the judge involved, when reading legal cases. This kind of source information was often ignored by novices in the legal domain. Wineburg (1991) found comparable results with experts and novices in the domain of history, while Britt and Aglinskas (2002) concluded that novices do not spontaneously attend to source information. Rouet, Britt, Mason, and Perfetti (1996) found that undergraduates mostly used characteristics of the content to evaluate documents, while graduate students in history mainly based their evaluations on document type. The studies by Britt and Aglinskas (2002), Lundeberg (1987), Rouet, Britt, et al. (1996) and Wineburg (1991) all show that experts use more source information when evaluating information than novices. However, in all these studies printed documents like textbooks were used.

Brand-Gruwel et al. (2005) compared the process of novices and experts, while searching the Web for information. Experts in this study were PhD students in the field of educational technology and the novices were psychology freshmen. The participants solved an information-problem while thinking aloud. Contrary to the other studies that investigated expertise effects, Brand-Gruwel et al. used a problem outside the expertise domain. Instead of education or psychology, the problem regarded how to deal with food that is out of date. Results showed that experts in solving information problems spent more time on defining the problem compared to novices. Experts and novices showed no differences in the way they searched the WWW; that is, using a search engine, typing in keywords, following links. Experts evaluated information more often than novices and regulated the process more often than novices. However, Brand-Gruwel et al. did not investigate the kind of criteria experts and novices use when evaluating search results, information and source.

The findings by Brand-Gruwel et al. (2005) are in line with the studies of Britt and Aglinskas (2002), Lundeberg (1987), Rouet, Britt, et al. (1996) and Wineburg (1991). All studies suggest that expertise effects in evaluation are not solely dependent on domain knowledge but also on more generic skills. Having domain knowledge does not improve evaluation of information if other problem solving skills are lacking. This is true for evaluation of printed documents as well as information on the WWW.

In summary, using the Web as a source of information for educational purposes requires the use of criteria to evaluate search results, information and source. Examples

of criteria are 'title and summary', 'connection to task', 'information agrees with prior knowledge' and 'reputation'. Children, teenagers and adults seem to have some knowledge on these criteria, but do not always use these criteria. The use of criteria does not only depend on domain knowledge but also on more generic problem solving skills. Novice searchers like teenagers and students may not posses these problem solving skills, or may not posses these skills on the required level. Teachers are responsible for developing these skills in their students, prompting them to use these skills, assessing students' essays and assessing students' use of evaluation skills. Teachers can only teach students how to solve an information-problem and how to use criteria to evaluate search results, information and source when they are aware of the criteria they use themselves and of the possible influence of their domain knowledge on the use of these criteria.

To uncover criteria teachers use and the possible influence of domain knowledge on the evaluation skills an in-depth study will be performed. The research questions are:

- Which criteria do teachers of secondary education use for evaluating search results, information, and source when searching the Internet for accomplishing school assignments?
- Is there a difference between the use of criteria when solving a task in their own domain of expertise and in a different domain?
- Does the use of criteria and domain knowledge influence the results of the tasks?
- Are teachers aware of the criteria they use and the criteria that can be used?

Method

Participants

Eleven teachers (nine male and two female; mean age = 43.00, *SD* = 13.47) of two schools for secondary education participated in this study: four physics teachers, three geography teachers, and four language teachers.

Material

Information-problems

Nine information-problems were developed regarding teachers' field of expertise. Table 1 presents the tasks as used in the study. The tasks were developed and tested in an earlier study (Walraven et al., in press). These tests revealed that all tasks had the same level of complexity. The tasks were based on the notion that these tasks could be used in the third year of pre-university education (9th grade). The tasks were preceded by the following text: 'In this task, you have to answer a question. You can search the Internet for information needed to answer the question, and you can copy and paste the information into a word-file. Finally, try to answer the question in a few sentences using the information found. You have 30 min for this task. Think aloud during your search'.

Table 1 The tasks used in the study

Number	Title	Domain	Task (translated from Dutch)
1	Kyoto	Geography	Ecological changes have become clearly visible during the last years. The snowline of the Kilimanjaro is moving and the ice of the polar caps is melting. Can all this be prevented if the Kyoto protocol is followed?
2	Dialect	Language and Culture	There are several dialects within the French language. Which dialects exist and what are the differences between these dialects?
3	Pit coal	Geography	Why does the Netherlands have pit coal layers of 1-2 metres thick, situated far below the ground, while the USA has layers 40 metres thick situated near the surface of the earth?
4	French and English	Language and Culture	How can the equivalences between the English and French vocabulary be explained?
5	Missing children	Language and Culture	How effective is searching for missing children by placing their pictures on milk cartons?
6	Electron	Physics	An electron has a charge, but what is this charge exactly?
7	Pollution	Geography	What is the effect of pollution on the quality of tap water?
8	City patterns	Geography	You have probably never thought about it, but many cities have certain city patterns. European cities are built different than cities in the USA. What are the differences in city patterns?
9	Snow	Physics	Why is there so much air in a snowflake?

WWW Questionnaire

A questionnaire was used to obtain teachers' use of and conceptions about the WWW. The questionnaire consisted of two parts. In the first part teachers were asked about the amount of time spent on the WWW per day, what they used the WWW for, and their three favourite websites. The second part consisted of 16 seven-point Likert-scale items. Ten items focused on the importance and usability of the WWW in class (e.g., 'I think the use of Internet in education should be stimulated'), Cronbachs a = .87. Six items focused on their perceived WWW skills (e.g., 'I have enough knowledge of the Internet to be able to work with it'), Cronbachs a = .87. These 16 items were based on the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich, Smith, Garcia & McKeachie, 1993). This adapted version of the MSLQ was used in an earlier study of Lazonder (2005).

Post hoc group interviews

A day after the teachers completed the tasks, they returned in groups for a group interview. These interviews were used to obtain teachers' knowledge and conceptions about criteria for evaluating search results, information and source.

Procedure

During the individual sessions participants first filled out the WWW Questionnaire, and then received an instruction on how to think aloud. After this instruction each participant conducted two tasks; one in his or her own domain of expertise, and one in another domain. After reading the first task participants searched the WWW for a maximum of 30 min to solve the problem while thinking aloud. 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. Participants returned in focus groups the next day for approximately one h. Two groups consisted of two participants, one group of four participants and one group of three participants. Each teacher received a booklet with screen shots of websites they had opened and had or had not used for their answer. They were asked to write down why they had or had not used the sites. This was followed by a group discussion. First, the participants were asked to report what they had written down and what they had paid attention to while searching for information. This was done per task. Next teachers were asked to come up with important evaluation criteria. Furthermore, they were asked if they had paid attention to these criteria while searching the day before, or if they ever paid attention to these criteria. Finally, teachers were asked if the criteria were the same for the courses language and culture, geography and physics. These sessions were recorded.

Data analysis

All *think-aloud protocols* were transcribed verbatim. To analyse all evaluations, a coding scheme, developed by Walraven et al. (in press) was used. This scheme is presented in Table 2.

It should be noted that only the evaluations were scored that were explicitly mentioned. Teachers also discarded pages without telling why. In those cases they did evaluate the results, information or source, but it was not clear on what grounds. These evaluations were not included in the analyses.

Furthermore, to gain more insight in the use of prior knowledge when evaluating information, sites or hit lists, all utterances (a sentence or a group of sentences focused on one subject) containing use of domain knowledge were underlined. An utterance was characterised as a *domain knowledge utterance* when specific knowledge concerning the task was used. For instance when the evaluation was based on names of experts, synonyms or explanations of phenomena. An utterance like: 'That is correct. Ion. Cation. Anion. That is all known, right?' while performing the task on the electron, is coded as a domain knowledge utterance. Furthermore, the domain knowledge utterances were divided based on the goal of the utterance. One can activate prior knowledge important to solve the problem, use domain knowledge to evaluate results, information or source, or one can use domain knowledge to make a decision upon the search strategy.

Table 2 Evaluation criteria

Sub skill	Criteria	Questions
Evaluate search results	1. Title/Summary	What is the title given by the search engine and what is the content of the summary?
	2. Kind (site/PDF/)	What kind of source is it, a website, a word document, a PDF file?
	3. Address	What is the address? Is it a .com or .org address?
	4. Rank in hit list	How many results are there in total and what is the rank of the result I am evaluating?
	5. Known to user	Have I used this site before, or have I heard good or bad things about it?
	6. Language	Is the site in a language I prefer and/or understand?
Evaluate information	**A Usability**	
	1. Language	In what language is the information written? Are there many grammatical or type errors? Is it filled with domain specific language?
	2. Connection to task	Does the information answer (part of) the information problem?
	3. Audience	Is the information aimed at a specific group of readers?
	4. Topicality	Is the information up to date?
	5. Amount	Is there enough information on the page? Or only a part of the information I'm looking for?
	B Verifiability	
	1. Author	Who has written the information? Can I contact him/her?
	2. References	Are there references on the page to used sources? Or links to more websites on the same subject?
	3. Information agrees with more sites	Can I find the same information on more pages or is this information only available on this site?
	4. Information agrees with prior knowledge	Does the information confirm what I already know?
	5. Organization	Which organization is behind the information? A governmental organization or a health organization? Can I find their logo on this site?
	C. Reliability	
	1.Kind of information	What kind of information is it? A newspaper article or a forum? Is it an opinion or results from research?
	2. Objectivity	Is the information objective or coloured by a certain point of view? Are there a lot of advertisements on the page?
	3. Primary/Secondary	Is the information first hand or is it someone telling about someone who did something?
	4. Goal	What does the (author of) information want to achieve. Sell something? Convince me of something or just inform me?
Evaluate source	**A Technical**	
	1. Appearance	Does the site appeal to me? Does it have pictures or only text?
	2. Speed	Does it take a lot of time to load the page?
	B Usability	
	1. Language	In what language is the site written? Are there many grammatical or type errors?
	2. Connection to task	Does the site have a connection to (part of) my information problem?
	3. Audience	For whom is the site meant? Who are it's visitors?
	4. Topicality	Is the site updated regularly?
	C Verifiability	
	1. Reputation	Is this site famous or infamous? Does it have a good/bad reputation?
	D. Reliability	
	1. Kind (site/PDF)	What kind of source is it, a website, a word document, a PDF file?
	2. Primary/secondary	Is the site an original source or a site telling about what is written somewhere else?

Two raters individually scored six of the 22 protocols. Interrater reliability was calculated. The similarity between the two raters, expressed in Cohen's Kappa, was higher than .80. for all protocols. One of the raters scored the remaining protocols.

The *group interviews* were also transcribed verbatim. The evaluation criteria mentioned by teachers were scored and it was counted how many groups mentioned the criteria from Table 2. Answers of the teachers to the question whether they had paid attention to these criteria while searching the day before were grouped in three categories: 1)Yes I paid attention to these criteria, 2) No, I did not pay attention to these criteria, 3) I did not pay attention to all these criteria, only to a few. Answers to the question whether they ever paid attention to these criteria were grouped in three categories: 1) Yes I always pay attention to these criteria, 2) No, I never pay attention to these criteria, 3) I sometimes pay attention to these criteria. Answers to the question whether the criteria are the same for the courses language and culture, geography and physics were grouped in two categories: 1) Yes, criteria are the same, 2) No, criteria differ between courses.

The *task performance* (answer to the question on the posed task) of the participants was judged based on three criteria: 1) Answer contains information of more than one source. Zero points for only one site, one point for more than one site, 2) Quality of the sources is judged based on the evaluation criteria in Table 2. Quality can be zero to four points. Zero points if the used sources did not comply with any of the evaluation criteria (e.g., no author mentioned, the site had no connection to task). One point if one used source complied at least with some evaluation criteria. Two points if more sources complied with evaluation criteria. Three points if more sources complied with not only the basic evaluation criteria (e.g., language, connection to task) but also with more sophisticated criteria like references, goal and reputation. 4 points were given if every used source complied with basic and sophisticated criteria. 3) Suitability of information used was determined by identifying content related elements. Suitability 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 is nine.

To answer the research questions concerning if teachers differ in the used criteria when evaluating *1) search results, 2) information* and *3) source* when solving a task in their own domain or in a different domain non-parametric analyses for within subjects designs (2 related samples – Wilcoxon) were performed. A Wilcoxon signed-rank test is a non-parametric alternative for a t-test or ANOVA for the case of two related samples or repeated measurements on a single sample. This non-parametric test was used because of the small sample size (*N*=11).

Results

Before going into the results concerning the research questions, we report the analyses of the WWW Questionnaire to determine whether teachers were comparable in terms of WWW experience. Five teachers spent one to five h a day on the Internet. The other six spent less than one h a day on the Internet. The rated importance and usability of the WWW on the seven point Likert scale for education was 5.4 (*SD* = 0.61), with a minimum of 4.7 and a maximum of 6.6. This means that all teachers find the WWW important and usable for education.

The mean rated perceived WWW skills on the seven point Likert scale was 4.5 (*SD* = 1.09). Five teachers scored 4.5 or lower on this scale. The minimum score was 2.5 and the maximum score was 6.0. This means teachers differ somewhat in their perceived WWW skills.

Because no substantial differences between teachers' use of and conceptions about the WWW were found, these results of the questionnaire are not taken into account in further analyses.

Teachers' evaluations of search results, information and sources

To answer the research questions a) about which criteria teachers use, and b) whether these criteria are different when solving a task in their own or a different domain, the thinking aloud protocols were analyzed. Table 3 presents the criteria and the frequencies on how often teachers used these criteria for evaluating results, sources and information.

Evaluating search results. In total (22 performed tasks) 272 search results were viewed and they were evaluated explicitly 127 times (46.7 %). The teachers appeared to use only a few criteria during their searches. They evaluated search results in the hit list mainly on *title and summary*. They read the title and the summary and decided whether the information on the site could help them find an answer to the question. The other criteria, like *address* or *rank in hit list* were used only a few times.

Evaluating information. Information was evaluated 117 times out of 204 page views (57.3 %). The usability of the information found was often evaluated with the criterion *connection to task*. This means that teachers sought for information that could help answer the question. Sometimes the *agreement with prior knowledge* was used as a criterion. Reliability of information was checked four times with the criterion *kind of information*. This means teachers checked whether the information was a result from for instance research or an opinion in a newspaper article.

Table 3 Frequencies of criteria used for evaluating search results, information, and source

Sub skill	Criteria	Frequencies
Evaluate search results	1. Title/Summary	117
	2. Kind (site/PDF/)	2
	3. Address	2
	4. Rank in hit list	1
	5. Known to user	3
	6. Language	2
Evaluate information	**A Usability**	
	1. Language	2
	2. Connection to task	95
	3. Audience	0
	4. Topicality	0
	5. Amount	6
	B Verifiability	
	1. Author	0
	2. References	0
	3. Information agrees with more sites	1
	4. Information agrees with prior knowledge	7
	5. Organization	1
	C. Reliability	
	1.Kind of information	4
	2. Objectivity	1
	3. Primary/Secondary	0
	4. Goal	0
Evaluate source	**A Technical**	
	1. Appearance	1
	2. Speed	1
	B Usability	
	1. Language	0
	2. Connection to task	3
	3. Audience	2
	4. Topicality	0
	C Verifiability	
	1. Reputation	0
	D. Reliability	
	1. Kind (site/PDF)	0
	2. Primary/secondary	0

Evaluating sources. Only seven times out of 204 page views the source itself was evaluated (3.4 %). The source was evaluated on *speed, appearance*, *connection to task* and *audience*. The latter means that teachers sought for sources that were written for a certain audience.

In summary, teachers do not evaluate search results, information or source aloud every time they viewed them. What can be noticed is that the source is rarely evaluated. Furthermore, when evaluating search results, information and source not all possible criteria were used. Out of the possible 29 criteria (as specified in Table 3) only 18

criteria were used. Criteria most frequently used were 'title and summary' and 'connection to task'. Criteria used less often were 'rank in hit list', 'information agrees with prior knowledge' and 'appearance'. Criteria not used at all were 'topicality', 'author' and 'reputation' amongst others.

Criteria for evaluating: domain specificity

The second research question concerns whether the used evaluation criteria are different for teachers when solving a task in their own or a different domain. Table 4 presents the means and standard deviations of the used criteria for evaluating results, information and source for all tasks and for tasks in teachers' own domain or other domain.

Table 4 Means and standard deviations of criteria use

Sub skill	Criteria	All (N=22) Mean (SD)	Own domain (N=11) Mean (SD)	Other domain (N=11) Mean (SD)
Evaluate search results	Title/Summary	5.3 (4.5)	5.7 (4.8)	4.9 (4.5)
	Kind	0.1 (0.3)	0.1 (0.3)	0.1 (0.3)
	Address	0.1 (0.3)	0.1 (0.3)	0.1 (0.3)
	Rank in hit list	0.0 (0.2)	0.1 (0.3)	0.0 (0.0)
	Known to user	0.1 (0.5)	0.3 (0.6)	0.0 (0.0)
	Language	0.1 (0.4)	0.2 (0.6)	0.0 (0.0)
Evaluate information	Language	0.1 (0.3)	0.0 (0.0)	0.2 (0.4)
	Connection to task	4.3 (3.5)	4.5 (3.6)	4.1 (3.5)
	Amount	0.3 (0.7)	0.3 (0.9)	0.3 (0.5)
	Information agrees with more sites	0.0 (0.2)	0.0 (0.0)	0.1 (0.3)
	Information agrees with prior knowledge	0.3 (0.8)	0.3 (0.9)	0.4 (0.8)
	Organization	0.0 (0.2)	0.1 (0.3)	0.0 (0.0)
	Kind of information	0.2 (0.5)	0.2 (0.6)	0.2 (0.4)
	Objectivity	0.0 (0.2)	0.0 (0.0)	0.1 (0.3)
Evaluate source	Appearance	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
	Speed	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
	Connection to task	0.1 (0.5)	0.1 (0.3)	0.2 (0.6)
	Audience	0.1 (0.4)	0.2 (0.6)	0.0 (0.0)

To determine whether the use of criteria to *evaluate search results* differed within subjects and between tasks performed in teachers' own domain and outside their domain a Wilcoxon signed-rank test was performed. There was no significant difference between the number of evaluations of the results in tasks performed in teachers' own domain (Mdn. = 6) and tasks outside their domain (Mdn. = 4), T=20, *ns*, r = -.06. To test if teachers used certain criteria more while performing a task in their own domain than while performing a task outside their domain Wilcoxon signed-rank tests were performed on each of the criteria used for evaluating results (e.g., *title/summary, kind, address, rank in hit list, known to user* and *language*). None of these tests were significant. So, domain did not influence the use of criteria to evaluate the search results.

To identify whether the use of criteria to *evaluate information* differed within subjects and between tasks performed in teachers' own domain and outside their domain a Wilcoxon signed-rank test was performed. There was no significant difference between the number of evaluations of information in tasks performed in teachers' own domain (Mdn. = 4) and tasks outside their domain (Mdn. = 3), T=24.5, *ns*, r = -.07. To test if teachers used certain criteria more while performing a task in their own domain than while performing a task outside their domain Wilcoxon signed-rank tests were performed on each of the criteria used for evaluating information (e.g., *language, connection to task, amount, information agrees with more sites, information agrees with prior knowledge, organization, kind of information* and *objectivity*). None of these tests were significant. So, domain did not influence the use of criteria to evaluate information.

No further analyses were done with the use of criteria to evaluate source since the number of used criteria was too small.

In summary, the use of criteria for evaluating results and information did not differ significantly between tasks in teachers' own domain and tasks outside their domain.

Qualitative analysis

Besides a quantitative analysis, a qualitative analysis was also performed. All domain specific utterances were underlined. Examples of domain specific utterances are:

> *"Floating woods mummified into pit coal, of course that is absolute nonsense" (Teacher 5, geography: geography task).*

> *"Yes, the vowel shift, I know about that. In the English language there has been a vowel shift. They changed vowels, and words are now written differently" (Teacher 3, english: language and culture task).*

Utterances were also categorized by goal. Three goals were specified: activating prior knowledge, evaluation and deciding on strategy.

An example of an utterance with the goal of activating prior knowledge is:

> *If I am not mistaken, French and English are both Roman languages, and not German languages. If I'm not mistaken (Teacher 3, english: language and culture task).*

An example of an utterance with the goal of evaluating the information is:

> *This seems to be a nonsense story to me, this has got nothing to do with geography (Teacher 5, geography: geography task).*

An example of an utterance with the goal of deciding what the next search action (strategy) will be:

> *I think I know the answer, but I haven't found it on the Internet yet. I have to keep on searching. (Teacher 2, geography: geography task).*

Table 5 presents the results of the qualitative analyses.

Table 5 Domain specific knowledge utterances

	Task in domain of expertise	Task outside domain of expertise
Number of teachers	6	3
Number of utterances	29	7
Number of utterances with goal: activating prior knowledge	11	4
Number of utterances with goal: evaluating	9	2
Number of utterances with goal: make decision upon search strategy	9	1

Domain specific utterances were not found in all 22 protocols. Only 6 teachers explicitly used domain knowledge while solving a task in their domain of expertise, and only 3 teachers used domain knowledge while solving a task outside their domain of expertise.

The goal of the utterances does not differ between tasks in the domain of expertise and the tasks outside the domain of expertise ($\chi^2(2, N = 36) = 1.08$, $p = .58$).

The results of the quantitative analyses show that domain knowledge does not influence the use of evaluation criteria. The qualitative analyses show that not all teachers explicitly use domain related knowledge while solving a task. The goals of the utterances while solving a task in their domain of expertise do not differ from the goals while solving a task outside their domain of expertise.

Teachers' task performance

The third research question addresses the influence of the use of criteria and domain knowledge on the results of the task. Table 6 shows the means and standard deviations of the total product score and the scores on quality of sources and suitability of the information on the tasks where domain knowledge was expected, and on the task where less or no domain knowledge was expected.

Table 6 shows that the mean scores on all the aspects are low. The maximum possible total product score is 9 and the mean is 4.6. The maximum achieved total score was 9. The maximum (possible and achieved) score on quality of the answer is 4 and the mean is 1.9. Thus, teachers' scores are not optimal. The product scores do not differ significantly between tasks performed inside the domain of expertise (Mdn. = 5) and tasks performed outside the domain of expertise (Mdn. = 4), T = 17.50, *ns*, r = -.30. Quality of sources does not differ between tasks, T = 14, *ns,* r = -.30, nor does suitability of information, T = 5, *ns,* r = -.33. Domain knowledge does not seem to influence task performance.

Because teachers did not differ in their use of criteria in their own or another domain this distinction is not taken into account when analysing the relation between criteria use and performance. Overall the correlation between how often the teachers evaluated hits, sites or information on the one hand and the performance score was r = .12, *p* =.60.

Table 6 Means and standard deviations of the product score (Maximum = 9)

Tasks	N	Total Product Mean (SD)	Using more than one site (maximum = 1) Mean (SD)	Quality of Sources (maximum =4) Mean(SD)	Suitability of Information (maximum = 4) Mean(SD)
All	22	4.6 (2.2)	0.8 (0.4)	1.9 (1.0)	1.9(1.2)
In domain of expertise	11	5.3 (2.2)	0.8 (0.4)	2.2 (1.0)	2.3 (1.3)
Outside domain of expertise	11	3.9 (2.1)	0.8 (0.4)	1.5 (0.9)	1.5 (1.1)

Teachers' awareness of evaluation criteria

As shown in Table 3, teachers used 18 criteria out of the possible 29 criteria. The fourth research question focuses on teachers' awareness: are they aware of the criteria they use and criteria they can use?

Table 7 presents the criteria and the number of groups that mentioned these criteria during the group interviews. Two criteria were mentioned by all four groups; *title/summary, (results)* and *language (information)*. Three groups mentioned *connection to task (information), audience (information), author (information), references (information), objectivity (information), appearance (source)* and *kind (source)*. Two groups mentioned *address (results), rank in hit list (results), known to user (results), topicality (information), information agrees with more sites, information agrees with prior knowledge, organization (information), kind of information* and *primary/secondary (information)*. *Kind (results), amount* and *goal* were mentioned by only one group and *language (search results), speed, language (source), connection to task (source), audience (source), topicality (source), reputation (source) and primary/secondary (source)* were never mentioned.

Comparing Table 7 with Table 3 (the criteria used while solving the task) shows that most of the criteria actually used were also mentioned in the group interviews. The criterion *language* for evaluating search results and the criteria to evaluate sources, *speed, connection to task* and *audience* are used, but not mentioned. *Audience (information), topicality (information), author, references, primary/secondary (information)* and *goal* were mentioned during the interview but not stated during task performance. It seems that teachers are aware of more criteria than they explicitly use.

When asked about the importance of these criteria, teachers mentioned that it depends on the domain, domain knowledge and the kind of task whether criteria like *information agrees with more sites*, *author* and *topicality* should be used to evaluate the information. They felt that geography was a course where topicality was more important than in physics. Most teachers said that when they knew the information on a page was correct, they did not bother to check who the author was and if the information was up to date. They also felt they did not have to check for more sites, if they knew the answer was correct. However, they also agreed that their students should always compare more sites, and check for author and topicality. 'If my students would have used this site, I would not have accepted it.'(Teacher 6, physics teacher), remarked one teacher. The criterion *author* is important, but teachers wondered if

students were able to evaluate information with this criterion, because they lack the background knowledge. Teachers felt students might be better able to evaluate information based on the *organization* behind a site. The criterion audience could also be difficult for students to use.

Table 7 Criteria mentioned without prompting during group interviews

Sub skill	Criteria	Number of groups (maximum = 4)
Evaluate search results	1.Title/Summary	4
	2.Kind (site/PDF/)	1
	3.Address	2
	4.Rank in hit list	2
	5.Known to user	2
	6.Language	0
Evaluate information	**A Usability**	
	1. Language	4
	2. Connection to task	3
	3. Audience	3
	4. Topicality	2
	5. Amount	1
	B Verifiability	
	1. Author	3
	2. References	3
	3. Information agrees with more sites	2
	4. Information agrees with prior knowledge	2
	5. Organization	2
	C. Reliability	
	1.Kind of information	2
	2. Objectivity	3
	3. Primary/Secondary	2
	4. Goal	1
Evaluate source	**A Technical**	
	1. Appearance	3
	2. Speed	0
	B Usability	
	1. Language	0
	2. Connection to task	0
	3. Audience	0
	4. Topicality	0
	C Verifiability	
	1. Reputation	0
	D. Reliability	
	1. Kind (site/PDF)	3
	2. Primary/secondary	0

Teachers also mentioned that evaluating information and source in general is easier for themselves because they have more domain and basic knowledge and they are more aware of the fact that not all information is necessarily true. Teachers

admitted that they did not always use the criteria they would normally use during the task. One teacher remarked that he did not think about reliability for one second, he just had to answer the task (Teacher 4, English teacher). They said that searching differs between certain purposes. If they searched for a hobby, they would take more time, and read and evaluate information more in depth. However, the 30 min time limit forced them to read quickly and less accurately and 'show student behaviour' (Teacher 3, English teacher). While solving the task, teachers felt that they had to choose between giving an answer as soon as possible, or doing a decent job. They agreed that students would want to solve a task as quickly as possible, while the teachers wanted them to take more time. One teacher remarked: 'We are more process oriented while a student is result oriented' (Teacher 3, English teacher).

In summary: teachers mentioned 21 out of 29 criteria. Only language (search results), speed, language (source), connection to task (source), audience (source), topicality (source), reputation (source) and primary/secondary (source) were never mentioned. Teachers seemed to know more criteria than they explicitly used, and felt that all criteria were important, but that it depends on the domain whether they should be used. Topicality for example is more important in geography than it is in physics. Thus, teachers thought that the use of criteria differed between domains. They also felt there was a difference in use of criteria during the tasks and while searching in general. During the tasks, teachers were less focused on using the criteria. Students should use the evaluation criteria when searching for educational needs, but teachers think students would have difficulties using criteria like author and audience, since students lack domain knowledge.

Discussion

The goal of this study was to uncover criteria teachers use to evaluate search results, information and source while searching the WWW, and to determine the possible influence of domain knowledge on the evaluation skills. It was determined (1) which kind of criteria teachers use when evaluating results, information and source while searching the WWW for information, (2) whether a difference in the use of criteria depends on domain knowledge, (3) whether the use of criteria and domain knowledge influence the results of the tasks and (4) whether teachers are aware of the criteria they use and the criteria that can be used?

Teachers in this study evaluate search results by the title and summary they find in the hit list. Information and source are evaluated by checking whether the information or source can give an answer to the information problem at hand. Teachers do not pay attention to the author of the information or references in the information. They also do not always evaluate results, information and source every time they viewed search results or websites.

There was no difference in teachers' use of criteria when solving tasks within or outside their domain of expertise. Domain knowledge did not seem to influence the use of criteria. A qualitative analysis of the results revealed that not all teachers explicitly use domain knowledge. More domain knowledge was uttered while solving a problem

inside their domain of expertise. The goal of utterances does not differ between tasks in or outside the domain of expertise.

Domain knowledge did not influence the use of criteria, and it did not influence task performance. Teachers did not use better sources or found more suitable information on a task within their own domain than on a task outside their own domain.

Teachers were aware of the criteria they use and that can be used. During the group interviews they mentioned more criteria than they used. For instance, they mentioned *topicality (information), author* and *references*. They also admitted they did not use all the criteria they knew because they wanted to come up with an answer for the task. Teachers also stated that the use of criteria depends on the domain and domain knowledge.

The results reveal that teachers did not evaluate search results, information and source every time they viewed results or a website. Furthermore they mostly used more superficial criteria like 'title and summary' and 'connection to task'. It is possible that teachers used more criteria, but did not explicitly mention these criteria while thinking aloud; the thinking aloud method could have influenced the results. Not everyone is capable of thinking aloud. Sometimes participants verbalise their actions and not their thoughts (Young, 2005). Furthermore, thinking usually happens faster than verbalising thoughts. It is possible that while verbalising one thought, the participants had more thoughts, but did not verbalise them. And if they would verbalise every thought, it is possible that they would stop task execution. Recording eye movements of participants while solving a task and afterwards replaying these movements for them and letting them think aloud while watching their eye movements for instance on half speed could be a way of giving the participants more time to verbalise their thoughts. They only have to concentrate on their eye movements and thought, in stead of thinking about the task. However, eye movement is not suitable for tasks that take more than 10 min. Moreover, a study by Brand-Gruwel, van Meeuwen and van Gog (2008) where eye tracking was used to see which criteria psychology students used when evaluating search results, information and source did not reveal the use of different criteria. Brand-Gruwel et al. reported higher frequencies of the same criteria found in the study described in this Chapter.

Moreover, the complexity of the tasks may have influenced the results of this study. The tasks were formulated on student level. It is possible that teachers were not forced to explicitly use their domain knowledge for evaluating results, information and source. However, providing teachers with tasks on student level, made teachers aware that solving tasks while using the WWW is difficult and that their students could encounter difficulties. Furthermore, a way for students' evaluation skills to develop is by letting teachers exhibit model behaviour, e.g., showing students how they should evaluate search results, information and source. The teachers in this study did not show model behaviour, since they did not use very sophisticated criteria for their evaluations and their performance was poor. Teachers themselves also admit that searching for information is not as simple as it seems: 'As an English teacher, I thought I would need ten minutes to find the explanation for the equivalence between the English and French language, but no way! And I would have stopped, if I was not part of an experiment' (Teacher 4, English teacher). Another teacher remarked: 'Searching is an art' (Teacher

2, geography teacher). This study made teachers aware that solving tasks while using the WWW is difficult and that their students could encounter the same difficulties.

Another possible limitation of this study is the experimental setting and the use of a maximum amount of time. Teachers admitted they did not always use the criteria they would normally use while searching for information. One teacher remarked that he did not think about reliability for one second, he just had to answer the task. They said there is a difference in searching for different purposes. If they searched for a hobby, they would take more time, and read and evaluate information more in depth. However, the 30 min time limit forced them to scan pages quickly and 'show student behaviour'. While solving the task, teachers had to choose between giving an answer as soon as possible, or doing a decent job. Perhaps teachers would have used more and more sophisticated criteria without the time limit.

The study described in this article did not find a relationship between the use of evaluation criteria and domain expertise. An explanation could be that since the Web is relatively new, its users have less experience with it, than with books or newspapers, and may find it difficult to evaluate search results, information and source (Metzger et al., 2003). Studies that have successfully shown a relationship between domain knowledge and evaluation all used printed text in stead of the Web (Britt & Aglinskas, 2002; Lundeberg, 1987; Rouet, Britt, et al., 1996; Wineburg, 1991). Perhaps inexperience with the Web causes users to not evaluate the information where they would evaluate it if they were reading printed text. Inexperience could therefore inhibit the effect of domain knowledge. It is possible that when these teachers become more experienced on the Web, the difference in domain knowledge will influence the use of evaluation criteria.

Use of evaluation criteria did not influence task performance. Perhaps, if teachers had used more sophisticated criteria for evaluating results, information and source, task performance would improve. Research by Brand-Gruwel, van Meeuwen et al. (2008) found a relation between the use of more sophisticated criteria (like author, references, and objectivity) when selecting sites and information, and the quality of the selected information by psychology students. Developing evaluation skills could therefore result in better task performance even before domain knowledge is enhanced.

The study described in this Chapter made clear that teachers are indeed equipped with the necessary knowledge on evaluation criteria, but are unaware of the fact that they do not always use these criteria and do not always function as a role model for students. Teachers should foster the development of evaluation skills. This could result in higher grades and consequently more motivated students.

Future research should focus on guidelines for instruction in evaluating search results, information and source. It is important that instruction is embedded in the curriculum and that students learn to use their evaluation skills in multiple contexts and courses (Brand-Gruwel & Wopereis, 2006; Walraven, Brand-Gruwel & Boshuizen, 2008b). A first step in this process could be to let more teachers become aware of their own use of criteria and their function as a role model for students in using these criteria. A form of design-based research, where instruction is designed, evaluated and re-designed together with teachers could be a good way of accomplishing this.

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

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.

This Chapter is based on Walraven, A., Brand-Gruwel, S, & Boshuizen, H. P. A. (submitted). *Fostering transfer of websearchers' evaluation skills: a field test of two transfer theories.*

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 et al., 1997; MaKinster et al., 2002; Wallace, Kupperman et al., 2000; Walraven et al., in press). 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 et al., 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 & Aglinskas, 2002; Metzger et al., 2003; Walraven et al., in press). It is not only important that students understand the need to evaluate and learn how to use evaluation skills, but also that these skills become adaptive to new situations and changed tasks (e.g., transfer). Failing this last quality will sooner or later lead to skills obsolescence.

The goal of the study in this Chapter 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. In the introduction of this article, first theories to foster transfer of a complex cognitive skill like evaluating information will be addressed. Second, the complex skill to evaluate search results, information and source will be discussed. Third, research concerning instructional methods for teaching the use of the WWW, evaluating information, and the way transfer is stimulated will be focussed upon. Fourth, the research questions addressed in this study will be presented.

Transfer of training

Transfer of complex cognitive skills can be fostered in several ways (e.g., Gick & Holyoak, 1983; Thorndike & Woodworth, 1901; Wertheimer, 1961). According to Perkins and Salomon (1989; Salomon & Perkins, 1989) 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. This so-called high road to transfer depends on mindful abstraction from the context of learning. It is 'the explicit conscious formulation of abstraction in one situation that allows making a connection to another' (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? 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. With backward reaching one abstracts in the transfer context, looking for features of the previous problem where new skills and knowledge were learned.

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) could be used to design instruction that fosters the transfer of evaluation skills. 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. 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 source 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.

To foster transfer Simons et al. (2000) emphasize the importance of a good, extensive and well organised knowledge base and the domain specific interpretation of the skills. This 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 of students 'are not restricted to one context but reach out to other contexts and situations.' (p. 2).

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 better anchored.

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 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., in press; Walraven, Brand-Gruwel, & Boshuizen, submitted). 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 et al., 2001; Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Koot & Hoveijn, 2005; Lyons et al., 1997; Wallace, Kupperman et al., 2000). Evaluating information and source is not always done based on clear, and well understood criteria but on intuition (Koot & Hoveijn, 2005).

Instruction in Web searching skills

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 et al., 2008b). Yet some research investigates instructional measures aiming to improve the evaluation of information from multiple documents.

Britt and Aglinskas (2002) developed the Sourcer's Apprentice, a computer application, to teach students sourcing, contextualization and corroboration in the context of researching a historical controversy. Sourcing refers to identifying critical features of the source like author, author's position, date, document type, etc;

contextualization refers to identifying relevant features of a source that can be useful in creating a context for historical information; corroboration refers to checking facts or interpretations from one source against other sources. The design of the Sourcer's Apprentice is based on five design principles. One of these principles aimed at support of transfer. This transfer was supported in two complementary ways. First a mapping was made between the environment in which the skills are learned and in which the skills must be applied and second the students are exposed to highly variable problems and text types to encourage students to abstract general heuristics and concepts. This abstraction of principles is in line with the ideas of the high road transfer theory of Perkins and Salomon (1989). A second principle used in the design of the Sourcer's Apprentice is supporting students' development of expert representations. The Sourcer's Apprentice supports the development of expert representations by providing various types of entities in the interface (e.g., source features, documents) and through explicit mentioning of relationships among entities (e.g., who else is cited in a particular document). The development of expert representations is in line with the rich representation theory for fostering transfer of cognitive complex skills. So, it seems that principles from the two theories have been used to design the Sourcer's Apprentice.

To determine the effect of this tool, different experiments were conducted. Using the Sourcer's Apprentice students (11th graders) read a structured set of documents about a single history problem, identified source and document features along with important content, and then used this information to answer questions and write a short essay. In different studies the group using the Sourcer's Apprentice was compared to a control group getting regular class instruction or a group getting the same material presented in a textbook. Students using the Sourcer's Apprentice showed better sourcing skills after the intervention (only 2 times of 40 min) than students in the other conditions. This was measured using a pre-test post-test control group design with tasks in the same domain of history but with a different type of controversy (military, economic, and social history) and different types of documents (e.g., treaties, letters, congressional reports, and historian essays). Because these tasks were not identical to the tasks used in the intervention it is reasoned that transfer occurred. One can however question if this type of transfer can be seen as far (e.g., transfer to a different domain) transfer, because the domain remained the same (history). Additionally, essays of the Sourcer's Apprentice group were graded higher than the essays of the group using the textbook materials. Essays of the Sourcer's Apprentice group included more document based information and more explicit citations and were better causally integrated.

Graesser, Wiley, Goldman, O'Reilly, Jeon, and McDaniel (2007) developed the web tutor SEEK (Source, Evidence, Explanation and Knowledge). SEEK was designed to promote college students' critical stance and learning while exploring Web pages. Critical stance requires presupposing that the quality of information is potentially suspect and requires close scrutiny with respect to its truth, relevance, and other dimensions of quality. SEEK consisted of a mock up Google page with three main facilities: 1) a hint button reminding the user of the goal of the task and giving suggestions what to do next; 2) a pop-up window asking to rate the expected reliability of the information and to type a justification; 3) a pop-up window with five questions, addressing core aspects of critical stance.

To test whether SEEK actually enhanced critical stance and thus influenced the processes of studying websites, ratings and rankings of websites' reliability, content learning and the use of what was learned in an essay, the results on a pre- and post-test of an experimental group (college students) were compared with results of a control group (same Google page, without the SEEK facilities). Results showed that after the intervention (50 min) students in both conditions studied reliable sites in more detail than unreliable sites, were able to differentiate reliable sites from unreliable sites and rated the reliable sites higher, and had more content knowledge about the subject of the websites; volcanoes. The essays of both groups had the same amount of correct and incorrect ideas on the subject matter, however, the essays of the SEEK group included more expressions of critical stance. So, there was a difference between pre and post-test, but only one difference between conditions; SEEK did not have the expected results. In a second study, prior instruction on critical stance and determining the reliability of websites in the form of an example was added to SEEK. Results showed that instruction had virtually no impact on the process of studying websites, ratings and rankings of websites' reliability, content learning and the use of what was learned in an essay. So, it is not sufficient to present information on critical stance in the form of an example to get students to apply critical thinking strategies in a flexible way in different settings.

Gerjets and Hellenthal-Schorr (2008) designed a training program called CIS-WEB (Competent Information Search in the World Wide WEB) aimed at improving cognitive and metacognitive aspects of pupils' (ages 12-13) ability to search for information competently on the WWW. CIS-WEB consists of six modules (total duration 12 lessons of 45 min). The six modules aimed at basic knowledge about the Internet, the WWW and search systems in the web, information-problems, structure of websites and use of web tools, evaluation with regard to credibility and actuality, segmentation of information-problems and processing of the resulting sub tasks. Students listened to presentations, worked collaboratively in a hypermedia environment and worked individually with paper and pencil materials like worksheets. 61 students followed the CIS-WEB program. It was tested whether CIS-WEB improved students search performance and their search-relevant and search-irrelevant declarative knowledge with regard to the web. Data on knowledge and search performance was gathered four times during the training. Declarative knowledge was measured with a multiple-choice test, and search performance was measured by the way students solved sets of information problems. Results showed that CIS-WEB enhanced pupils´ declarative knowledge about the web and their search performance. This program was based on a mix of didactical concepts as worked-out examples, symbolic and iconic visualizations, interactive multiple-choice questions with feedback, worksheets and exercises. With an intervention period of 12 lessons, it was effective and students became more competent websearchers. But it was only measured whether students used the skills in the same kind of context and not whether the skills were learned in a flexible way so that transfer would occur.

Kuiper et al. (2008) designed a curriculum for 5th graders to acquire Web skills. The implemented curriculum consisted of eight weekly lessons of about 2 h each. The first five lessons aimed at developing websearching, reading and evaluation skills. In the last three lessons, students received assignments and searched for information and

composed their own texts. Four schools participated in a multiple case study design. Learning results of the program in terms of content knowledge and web skills were investigated. Results revealed that students' knowledge about Web skills improved. But they still appeared to be inconsistent Web users, who did not always act upon their knowledge of websearching skills. Students showed unexpected, inconsistent or inflexible Web behaviour and little planning and reflection. Especially this reflection and planning aspects can be seen as metacognitive skills important to foster transfer. Although this study revealed that it is possible to teach students to become better websearchers, transfer measures were not included in this study.

Stadtler and Bromme (2008) used a different approach compared to the previously mentioned studies. They did not design and implement instruction, but a metacognitive tool. Adults (age 19 to 38) with little medical knowledge searched the WWW on a medical topic (cholesterol), and the tool 'met.a.ware' enabled users to store information they had found systematically. It provided them with different labelled tabs (ontological classification, like function and causes) under which they could store information. Participants also received prompts to evaluate the sources on credibility or to assess how well they comprehended information and how much they still needed to search for more information. To determine whether prompting (for evaluation of information and comprehension monitoring) would foster the acquisition of knowledge about sources and content knowledge, and would lead to more arguments to justify credibility judgements, participants in four conditions (evaluation prompts, monitoring prompts, both type of prompts, or no prompts) received preselected websites on the topic of cholesterol and were compared to a control group. Results revealed that prompts for monitoring and evaluation increased knowledge on content and sources, and ontological classification helped to structure notes and focused participants on important ontological categories. It can be conjectured that the approach taken in this study (metacognitive prompting) can foster transfer according to the assumptions of the high road transfer theory of Perkins and Salomon (1989). However transfer effect was not taken into account in this study.

Comparing the five discussed studies the following emerges: the type of instruction differs between these five studies. In two studies, participants were provided with a tool, but no instruction was given (Britt & Aglinskas, 2002; Stadtler & Bromme, 2008). In two studies participants received multiple lessons in the classroom (Gerjets & Hellenthal-Schorr, 2008; Kuiper et al., 2008). In one study instruction (worked example), was followed by a tool (Graesser et al., 2007). None of the instructional designs was strictly theory driven. Although, in most studies the use of metacognitive and reflection skills was emphasized and thus the principles supported by the high road transfer theory were used, effects on transfer were only measured in the study by Britt and Aglinskas (2002). The design of the Sourcer's Apprentice was based on a mixture of principles from both transfer theories: the high road theory and the rich representation theory. The students transferred evaluation skills to different and unsupported contexts, but the different contexts were all within the same domain.

These studies strongly suggest that instruction 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. The studies also showed that instruction

in evaluation skills can be designed according to principles derived from two transfer theories. However, we cannot draw any conclusion about transfer to new domains, nor did the studies reveal whether designs based on the two transfer theories differ in the effect within the domain of instruction and whether they are equally suitable for fostering transfer to a new domain.

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 quantitatively and qualitatively. 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?

Method

Participants

Four classes (84 students, age 14-15) of three different secondary schools participated in this study. Classes were randomly assigned to one of the educational programs. In the high road condition one class of 19 students and one class of 20 students participated. In the rich representation condition one class of 24 students and one class of 21 students participated. Each class had their own teacher, so in total four teachers and four classes participated.

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.

Materials

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 et al. (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.

High road program

In the high road program the focus was on the evaluation of results, information and source, 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?'.

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 to 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. Figure 1 shows two pages of a process worksheet.

Table 1 The lessons of the two educational programs

	Rich representation program	**High road program**
Lesson 1	Subject: Causes of World War 1 (WWI). Task: Find causes of WW I on the Internet.	Subject: Introduction to World War 2 (WWII). Task: Combine facts and dates on WWII.
Lesson 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.
Lesson 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 1924-1929. Students received a process worksheet with a high amount of support.
Lesson 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.
Lesson 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.
Lesson 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.
Lesson 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.

Lesson 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.
Lesson 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.
Lesson 10	Subject: Discuss the start of WWII. Task: None, normal lesson.	Subject: Conference of Munich 1938. Task: See lesson 9.
Lesson 11	Subject: The course of WWII till 1943. Task: None, normal lesson.	Subject: Economics between 1920-1939. Task: Match quotes and years.
Lesson 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 in to 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).
Lesson 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.
Lesson 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.
Lesson 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.

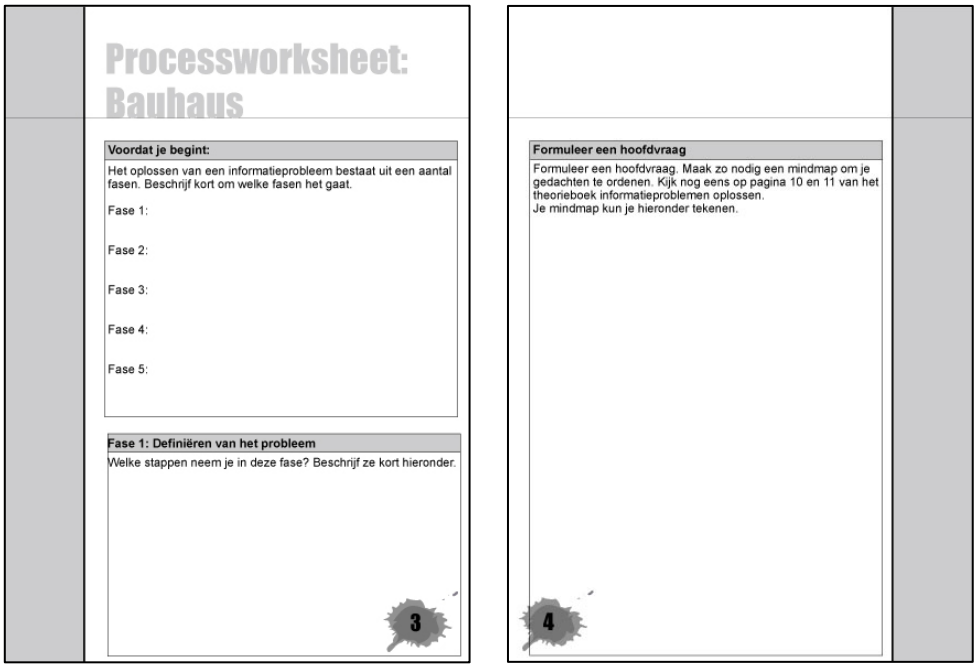


Figure 1 Process worksheet.

Rich representation program

In this program rich representation will be visualised 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 organised knowledge base (Ausubel, 1963; 1968; Ausubel, Novak & Hanesian, 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 interlinked, 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 knowledge structure (mind map) during the lessons. Figure 2 shows the knowledge web of class A after 15 lessons.

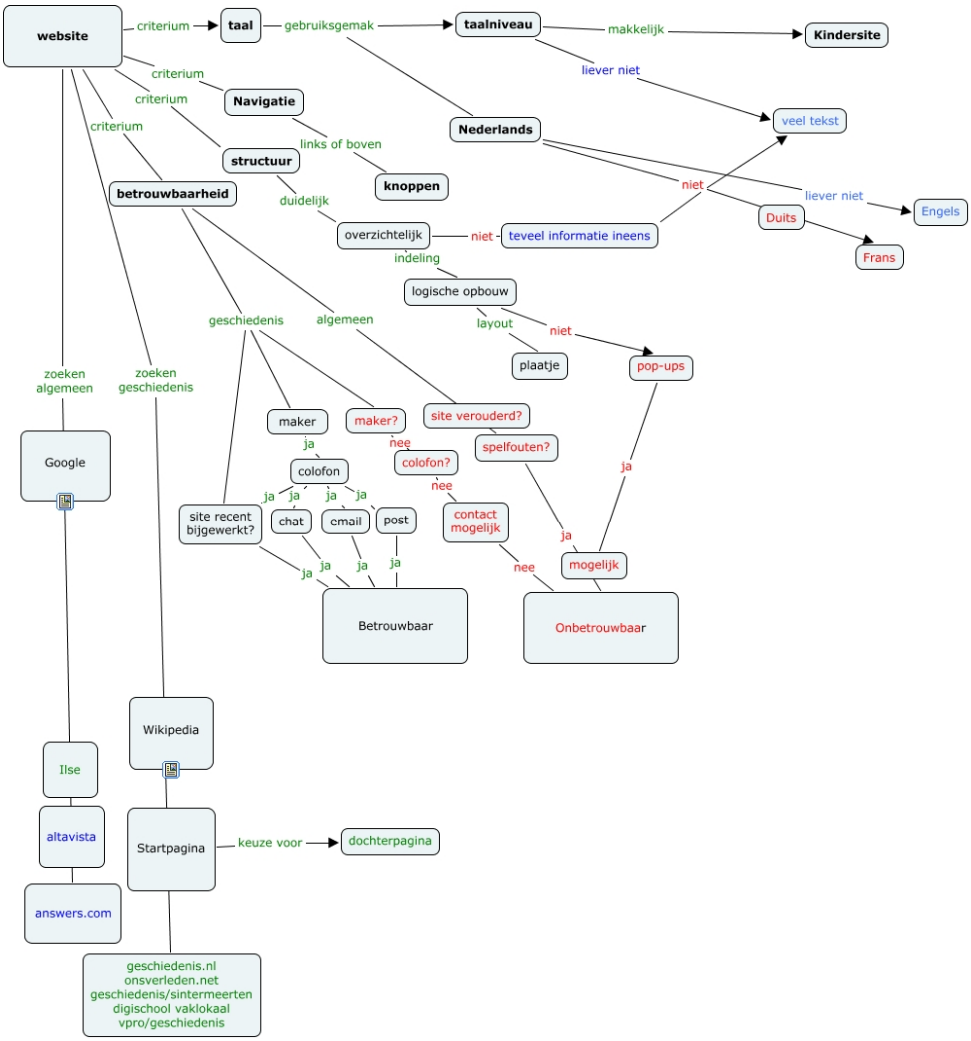


Figure 2 Knowledge web of class A after 15 lessons.

Measurements

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 list 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.

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 behind 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.

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 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."

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.

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 the 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 minutes 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 verse. 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)

Data analysis

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. (in press) 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).

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.

Results

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.

Table 3 Means and standard deviations of hit list en website evaluation task score

	High road program (N=39) Mean (SD)		**Rich representation program (N=45) 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)

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 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).

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.

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 evaluated 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.

Table 4 Frequencies of used criteria per domain and during pre- and post-test

		History		Biology	
Sub skill	**Criteria**	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
	3. Primary/Secondary	0	0	0	0
	4. Goal	0	0	0	0
Evaluate source	**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 evaluation 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

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 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.'

Table 6 Examples of utterances containing knowledge of evaluation criteria

Categories	***Examples***
Utterances with goal justifying an action based on an evaluation criterion	"I 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"

Task performance

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

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

	High road program (N=6) Mean (SD)				**Rich representation program (N=5) Mean (SD)**			
	Pre test history	Post test history	Pre test biology	Post test biology	Pre test history	Post test history	Pre test biology	Post test 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)

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 (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.

Field notes

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 if they could contact the author of the site and more often checked if there were references on the site.

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 to 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 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.

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 did 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.

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 their approach to the WWW. Students complained about the process worksheet. Teacher D 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.

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 Chapter 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 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 representation 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 believes 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). These epistemological beliefs are activated when students search the WWW and have to evaluate information (Mason & Boldrin, 2008). In order to become a good websearcher, 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 & 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 flexibly 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. Although 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 makes 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 organised 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; Buzan, 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.

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.

Fostering students' evaluation behaviour while searching the Internet: Using a Design-Based Research methodology

In a design-based research cycle, a program for teaching 9th graders to evaluate the information found on the WWW was designed and tested. Goal of the program was both to teach evaluations skills, and to achieve transfer of these skills to a different domain. In the design of the program the principles of two transfer theories were combined using a design-based research approach. Results revealed that the program improved students' evaluation behaviour. Compared to the students of the control condition, the evaluation skills of students in the experimental condition improve to a higher level, but did not lead to transfer. Observations and reflective stories of the teachers in the design team showed that the program was not implemented as planned, which could be an explanation for the results. The study yields guidelines for instruction in evaluation skills when searching for information on the WWW.

This Chapter is based on Walraven, A., Brand-Gruwel, S, & Boshuizen, H. P. A. (submitted). *Fostering students' evaluation behaviour while searching the Internet: Using a Design-Based Research methodology*.

Introduction

Students in secondary education get more and more assignments in which they have to search for information by themselves, because the teacher does not provide them with information nor does he or she prescribe what they have to read and study. Students mostly use the Internet to gather information but the way they use this information can be critiqued. Teachers recognize that students show problematic 'cut-and-paste' behaviour when searching the Internet for information to write an essay. Students search, find some information (reliable or not), cut and paste it in a document and hand it over as an essay to the teacher. Teachers agree that instruction in evaluating information found on the Internet is needed, that students need support to use different criteria to evaluate web information, and that instruction must be implemented in domain-specific programs. Unfortunately, educational programs do not give so much attention to stimulating students to evaluate information and to the way a teacher could guide his students on the World Wide Web (WWW).

In this Chapter the results of a design experiment conducted in cooperation with four 9th grade teachers aiming at improving students' use of evaluation skills when searching the WWW will be discussed. In the introduction, first, the grounds for taking evaluation of information on the World Wide Web as the main focus of instruction will be discussed. Second, principles of design-based research will be elaborated on. Third, the design cycle will be described, in which the focus will be on the final design cycle. In this part also the didactical approach based on two transfer theories will be addressed. Finally, the research questions will be presented.

Instruction in evaluating information while searching on Internet

The WWW is often the only source of information secondary education students use for accomplishing school assignments (Beljaarts, 2006; Jones, 2002). Despite their frequent use of the WWW, students' search methods are far from ideal; most students do not evaluate their search results, the information they have found and the source of this information (Fidel et al., 1999; Hirsch, 1999; Kafai & Bates, 1997; Koot & Hoveijn, 2005; Lorenzen, 2002; Lyons et al., 1997; MaKinster et al., 2002; Wallace, et al., 2000; Walraven et al., in press). Evaluating what one has found on the WWW is crucial, since the WWW lacks centralized control and regulation and its contents can easily be altered (Metzger et al., 2003).

Evaluation of results (the hit list), information and source (the information on the website and the website itself) is part of the larger process of information problem solving (IPS). The process consists of the constituent skills *defining the information problem* (i.e., reading the task, activating prior knowledge), *searching information* (i.e., choose search strategy, specify search terms, evaluate search results), *scanning information* (i.e., read information global, evaluate information and source, elaborate on content), *processing information* (i.e., read in depth, evaluate information and source, store relevant information, elaborate on content) and *organizing and*

presenting the information (i.e., structure relevant information, realize the product) (Brand-Gruwel et al., 2005). There are three evaluation moments in this process: evaluating search results during the searching phase, evaluating information and source during the scanning and processing phase. Several criteria can be used to evaluate search results, information and source (Barker, 2005; Beck, 1997; Boekhorst, 2000; Kirk, 1996; Ormondroyd, 2004). Evaluating *search results* will answer the question: Which site am I going to open? Criteria that can be used for evaluating search results are the title and summary of the site, the kind, the address, the rank in the hit list, whether the result is known to the user and the language. After evaluating the results and opening a website or file, the *information* on that site or in the file has to be evaluated. This can be done with several criteria, grouped in three categories: usability, verifiability and reliability. Language, connection to task, author, references, kind of information and objectivity are examples of criteria to evaluate information. The *source* can be evaluated on technical, usability, verifiability and reliability grounds, like speed, appearance, audience and reputation. Using these criteria and thus evaluating results, information and source can help avoiding the use of incomplete, false and biased information. The criteria used in this study can be found in the Appendix.

From an educational point of view, a non-critical attitude towards information on the WWW can result in reports and learning that lack quality (Britt & Aglinskas, 2002), since students often cut and paste information without evaluating it (Grimes & Boening, 2001; Rothenberg, 1998). Unfortunately, although the importance of instruction in an effective and critical use of the WWW has been recognized for several years, instruction in information skills is rare, not always effective and transfer is often not measured (Walraven et al., 2008b).

If students are to become critical users of the WWW and use that ability throughout their lifes, it is important that they can use their evaluation skills in multiple contexts and various courses. Instruction in these skills should therefore be aimed at transfer. Transfer can be fostered in several ways (e.g. Gick & Holyoak, 1983; Thorndike & Woodworth, 1901; Wertheimer, 1961). According to the transfer theory of Perkins and Salomon (1989, Salomon & Perkins, 1989), 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. This can be done by abstraction from the context of learning, by answering questions like what is the general pattern? What is needed? Which step can I take now? This abstracting is closely related to metacognitive skills. High road transfer can be fostered by stimulating a person's metacognitive skills, like planning, monitoring, and evaluating. The transfer theory of Simons et al. (2000) emphasizes the importance of a good, extensive and well organised knowledge base, and the domain specific interpretation of the skills (rich representation). This knowledge base includes three representations of the information: conceptual representations, which are the concepts and their defining characteristics, episodic representations that refer to personal experiences with the concepts, and action representations that describe how the conceptual and episodic representations are used. When these representations are rich and well connected, learning outcomes become durable, flexible and generalizable. In the design of the present research these two theories are used to design two educational programs, each based on one transfer

theory, to foster students' evaluation behaviour when searching for information in the WWW. These programs are designed, tested and redesigned together with teachers.

Design-Based Research

Educational research is not always directly connected to problems that occur in educational practice, and results of educational research are often only communicated to other researchers and not to the teachers who could benefit from the discoveries (Garvey Berger & Baker, 2008). The Design Based Research Collective (DBRC) has stated earlier: 'Educational researchers, policymakers, and practitioners agree that educational research is often divorced from the problems and issues of everyday practice—a split that creates a need for new research approaches that speak directly to problems of practice (National Research Council [NRC], 2002) and that lead to the development of "usable knowledge" (Lagemann, 2002).' (DBRC, 2003, p.5). Design-based research can bridge this gap between theory and practice. According to the founders of design-based research in the educational field (Brown, 1992; Collins, 1992) the approach taken is to study learning in context through the systematic design and study of instructional strategies and tools. It can create and extend knowledge on the development, enactment and sustainment of innovative learning environments and can help researchers understand how theories on teaching and learning can be transformed into effective learning in educational settings (DBRC, 2003). Design experiments can be conducted in various settings, from one-on-one (teacher-experimenter and student), to classroom experiments, to school and school district restructuring experiments (Cobb, Confrey, diSessa, Lehrer & Schauble, 2003). Cobb et al. identified five features that all these different types of design experiments have in common. First, the purpose of the experiment is developing a class of theories on the learning process as well as the means designed to support that learning. Second, the methodology has a highly interventionist nature. Third, the experiments have a prospective and reflective side. Fourth, the design is iterative, and fifth, the developed theories during design experiments are concerned with domain-specific learning processes and are accountable for the activity of design. These five features show that design experiments have both a pragmatic and theoretical side: designing forms of learning and developing theories by systematically studying those forms of learning and the means of supporting them.

Reeves (2006) described the design-based research process in four steps: (1) the practical problems should be analysed by practitioners and researcher; (2) solutions must be developed using design principles; (3) these solutions must be tested and refined in practice, and (4) a reflection to develop design guidelines and enhance implementation must be conducted. Furthermore, to make design experiments successful, the research team has to have collective expertise to develop an initial design, conduct the experiment and carry out a systematic retrospective analysis (Cobb et al., 2003).

In this Chapter we discuss a study in which two design cycles were carried out, according to the principles of design-based research. In these two cycles different programs to foster students' evaluation skills when searching for information on the

WWW were designed, tested, redesigned and tested again in a team of teachers and researchers. In the next section these design cycles will be described. In this description we will use the four phases as described by Reeves (2006) as starting point, but we will give, as Collins, Joseph, and Bielaczyc (2004) address, special attention to the independent and dependent variables and the way they were measured to determine the effect of the designed instruction.

Design Research: the 'evaluation of Internet information'-program

Design team

The design team consisted of four secondary education history teachers and one researcher. Teacher A was a very experienced web user and maintained several websites, teacher B had less experience on the Web. Both teachers (from one school) participated in the study because they acknowledged the importance of teaching students how to evaluate information on the WWW. 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. Main reasons for these two teachers who were from two other schools for secondary education to participate were acknowledging the need for students to learn how to evaluate information and, for the first teacher, wanting to learn more about ICT and the Internet. The researcher (R) was an educational technologist and instructional designer with an interest in ICT, and expertise in evaluation of information on the WWW.

Analysing practical problems

As stated the problem is that students in secondary education do not evaluate search results (hit list), information and source (websites) in a proper way and use superficial criteria when evaluating. This often leads to products of less quality, 'cut and paste' behaviour and it does not lead to deep learning. In analysing the problem, questions the team put up were: what are the criteria students must use to evaluate when searching the WWW and what are the steps they must take to solve an information problem?

To answer these questions three group sessions were held by the team. The main aim of these sessions was to make the teachers in the design team aware of criteria used for evaluation and the difficulties that students have when searching the WWW. To experience how they search and which criteria they use and find important for students to use, they solved a few student level information-problems on the WWW and discussed which criteria they used to evaluate search results, information and source. These criteria were summarized and compared with the criteria formulated in previous studies by the authors. The result was a list of criteria students should use when evaluating search result, information and source. This was a crucial step in the design process, since it made the teachers aware of their own skills, and the need for instruction in these evaluation skills.

The next issue concerned the steps students should take when solving an information-problem. The steps the teachers found important were compared to a

model used in previous studies (Walraven et al., in press; Walraven et al., submitted), based on the skill decomposition by Brand-Gruwel et al. (2005). Teachers agreed with the steps prescribed by the model (defining the problem, searching, scanning, processing and organizing and presenting information).

Designing instruction using design principles

During the last group session the knowledge and skills of the design team regarding designing educational programs based on design and transfer theories were discussed. The teachers were content experts (history), but had little or no experience with instructional design based on a specific theory. Instructional design in general was something they had done before but everyday routine in school prevented them from paying much attention to redesigning their lessons.

Important design principles that were discussed with the teachers were the principles derived from the two transfer theories as described above. Since it was not known what the instructional and transfer effect of each of the two transfer theories was, two programs would be developed, each based on one theory. For the design of a program based on the high road program (the transfer theory of Perkins and Salomon, 1989) process worksheets would be developed to inform students about the steps to be taken and to support them to reflect on their search process. The guidance given in these sheets must fade from high to low. For the design of the rich representation program (based on the transfer theory of Simons et al., 2000) a mind mapping technique with discussion sessions to build a knowledge structure would be used. Furthermore, in both programs whole tasks would be used with a certain kind of variability. That is, tasks requiring from students not only to evaluate information, but to define the problem, search and select information and come up with a product (e.g., essay, role-play, comic). Finally, the decision concerning grade and content was made. Theme of the lessons for the 9th graders would be World War II, ranging from the Treaty of Versailles at the end of the First World War, to the end of the 2nd World War. The lessons would also pay special attention to the war as it progressed in the Netherlands and the way it affected daily life in the specific area around the schools (i.e., the war at local level).

So, the goal of the first experiment was to design two programs, each based on a different transfer theory (high road and rich representation) to teach 9th grade of pre university students (during history classes) how to evaluate search results, information and source when searching for information on the WWW, using different kind of tasks and stimulating students to use these skills in a variety of settings, that is foster the transfer of the skill.

When all the principles were clear, the teachers split up in teams for the first design cycle. Each team consisted of two teachers and the researcher (R). Both teams used design principles from a certain transfer theory. One team developed the lessons based on the high road theory; the other based their lessons on the rich representation theory (also see Walraven, Brand-Gruwel & Boshuizen, 2008a). Both teams independently designed 15 lessons. During this first design cycle teachers were responsible for the content. R, participating in both teams, was responsible for embedding the evaluation skills according to the specific transfer theory, for instance designing the process

worksheets for the assignments used in the high road program. The teams had 5 to 6 meetings during the first design cycle.

Testing the design

A quasi experimental study was conducted to test and evaluate the effect of both programs

on students' evaluation behaviour, that is use of criteria for evaluating results, information and source when solving information problems on the WWW and if the effects of both programs based on two transfer theories (high road versus rich representation) differ in terms of transfer achieved. The teachers gave the lessons they had developed themselves. Each teacher had one class. So two classes received the lessons based on the high road transfer theory, and two classes received the lessons based on the rich representation transfer theory. Before and after the intervention students were asked to evaluate a hit list, and websites and information in the domain of history and in the domain of biology (transfer tests). Furthermore, 11 students accomplished two tasks (one history, one biology) while thinking aloud during pre- and post-test. Each class was observed three times by R. During these observations teacher-student interactions, actions of students and actions of teachers were written down by the observer. Researcher and teachers had regular email contact on how the lessons were carried out by the teachers and how they were received by the students.

For the results of this study we refer to Walraven et al. (2008a). But in sum, both programs improved students' evaluation behaviour in terms of evaluation of sources and information in the domain of instruction (history). The rich representation program also had an effect on students' use of evaluation criteria for evaluating a hit list. Concerning the transfer effect the results on the biology tasks showed the same pattern.

Redesign of the program

After this first design cycle, a new cycle was started. The design team remained the same, as well as the target group, 9th graders. The goal of the second design cycle was to develop one new instructional program, based on the lessons learned from the first design cycle. It was decided that the good aspects of both programs should be integrated in the new program. The second design cycle started with a group session. In this session, the two design teams of the first cycle presented their instructional program to each other. Since the design had been done separately, the teams did not know what the other team had designed. Next, R. presented the results of the two instructional programs. After that, the group discussed the strong and weak points of both programs. In the high road program, the strong points identified were the variety in tasks and the focus on the entire search process with help of process worksheets. The process worksheets were also a weak point, since they were too extensive and students rebelled against filling them out. Strong points of the rich representation program were the focus on one evaluation criterion per lesson and the discussions with students about criteria. A weak point was that not enough attention had been paid to the use of the criteria in various contexts or the way criteria are connected. Teachers also noted that is was very important to convince the students of the importance and significance of evaluating information, and to make sure that the students did not see evaluation of

information as just another task during these lessons, but as a skill they should use in more courses in school. It was decided that the new program would use less extensive process worksheets, better structured discussions and the program would start with a confrontation with the importance of evaluating information. At the end of the group session, the teachers decided on the content of the lessons and divided five themes between them. They agreed to design three lessons per theme; two lessons with WWW and evaluation assignments, and one classical lesson. This decision was made based on the experience with the previous programs; students had complained that they spent too much time behind the computer and longed for a normal lesson. It was decided that R would develop adjusted process worksheets and a detailed teacher manual for the discussions on criteria in order to develop a rich knowledge structure on criteria. During the two following group sessions, teachers presented their designed lessons and R presented the new worksheets and the manual. The materials were discussed and adjusted if the group felt this was necessary. A description of the designed program will be presented in the method section.

Testing the redesigned program

The redesigned program was tested and evaluated using a pre-test post-test control group design. Reeves (2006) described this testing as the third phase in the design cycle. The evaluation of this program will be described in the next sections. The questions addressed are: (1) What are the effects of the program on students' evaluation behaviour, that is, how do students use criteria for evaluating results, information and source, when solving information problems on the WWW? (2) Does instruction based on a combination of two transfer theories lead to transfer? (3) What is the effect of the program on learning results? (4) What are teachers' perceptions with regard to designing the programs and being part of an experiment? Furthermore, the fourth and last phase of the design research cycles (Reeves, 2006) - develop design guidelines – will be attended in the discussion section of this Chapter.

Method

Participants

Five 9th grade classes (101 students, age 14-15) of three different secondary schools participated in this study. Four classes received an educational program and one class served as a control class. The four teachers who designed the program taught the four experimental classes.

Program

Goal and overview of the lessons

The goal of the program was to teach 9th grade students how to evaluate search results, information and source when searching for information on the WWW in a historical context, using different kinds of tasks and stimulating students to use these skills in a variety of settings, that is foster the transfer of the skill. The general subject of the program was World War II and it consisted of 15 lessons of 50 min. Table 1 gives an overview of the lessons.

Table 1 The 'Evaluation of Internet information'-program

Lesson	Theme	Content
1	Confrontation	Students answer four questions about the treaty of Versailles, using 7 websites provided by the teacher. These sites contain contradictory information, information on a different treaty of Versailles, newspaper articles etc. After 30 minutes a discussion is held on what students noticed about the sites and criteria for evaluating results, information and source.
2-4	Versailles	Students act out the negotiations of the Treaty of Versailles. The class is divided in groups and each group searches for information on the Treaty and the viewpoints of the main characters of these negotiations (France, US etc.). When enough information has been collected students act out the negotiations. Students receive a process worksheet that focuses on the first step of the process: Define the problem.
5-6	Weimar and art	Students read a text on art in Germany at the beginning of the War and choose a subject to write an article on art and war. The process worksheet focuses on the second step: search information, with special attention paid to the evaluation of results.
7-9	The rise of Hitler in 1933	Students draw a comic on the rise of Hitler. They search information about this rise, write a scenario for the comic and draw the comic. Instead of drawing they can also use pictures they find. The process focuses on the third step in the process, scan information, with special attention paid to the evaluation of information and source.
10-11	Chronology	Students play a card game (happy families/old maid/ kwartet (Dutch)) Each set contains four events during a certain year (e.g., 1939: Occupation of the Czech republic, Molotov Ribbentrop pact, invasion of Poland, Russians occupy east of Poland). Students play the game and organize the events according to chronology.
12	1938	Students act out the convention of Munich, in the same way they acted out the negotiations on the treaty of Versailles. Process worksheet focuses on evaluation of information and source.
13	Daily life in the war: collaboration and resistance	Students write an interview with a Dutch collaborator or a hero from the resistance. They base their questions and answers on information they find on the WWW.
14	Persecution of the Jews	Normal lessons with textbook or movie.
15	The war in our own region	Students make a picture of a war monument in their home town and write an article about it.

As stated earlier, based on experiences with the previous programs, it was decided that the new program would use less extensive process worksheets, better structured discussions and the program would start with a confrontation with the importance of evaluating information. So, the first lesson was a confrontation lesson.

Students were confronted with the importance of evaluating results, information and source. Students had to answer four questions about the treaty of Versailles. They could only use 7 sites provided by the teacher. These sites contained contradictory information, information on a different treaty of Versailles, newspaper articles etc. For instance, the first site was a Wikipedia site about the treaty of Versailles of 1783. This was not the treaty the assignment was about. The second site was a newspaper article about the correct treaty of 1919, but with different information than the third site, with an unknown source. Goal of this lesson was confronting students with incorrect, false, and biased information and having them think about the importance of evaluating information. After this confrontation lesson, three lessons on the treaty of Versailles, followed by two lessons on Weimar and art, three lessons on the rise of Hitler, three lessons on chronology/1938, one on daily life in the war, one on persecution of Jews, and finally a lesson about the war in the region of the school.

Reader

Students 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 et al. (2005), and described the necessary phases for information-problem solving (define the problem, search information, scan information, process information and organize and present information) and the 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. In the scan information phase during the step evaluate information and source, students were given hints like: check if you can see when the site was last updated.

Process worksheets

Students received a process worksheet with the assignment for the coming lessons and some questions they had to answer at the start of each theme (Versailles, Weimar and art, the rise of Hitler and chronology/1938). The questions on the sheets were linked to the phases of information-problem solving and corresponded with the phases and steps in the reader. In the first three lessons the first phase (define the information problem) was focused on, the next three lessons the scanning information phase, and so on. So, in stead of filling out questions for each phase like on the worksheets in the previous high road program, students only filled out questions from one phase of the information-problem solving process. To make the entire problem solving process visible to the students, and to point out which phase of the process was focussed on, every worksheet started with a figure of the process solving process with the central phase highlighted. Figure 1 presents an example of a process worksheet.

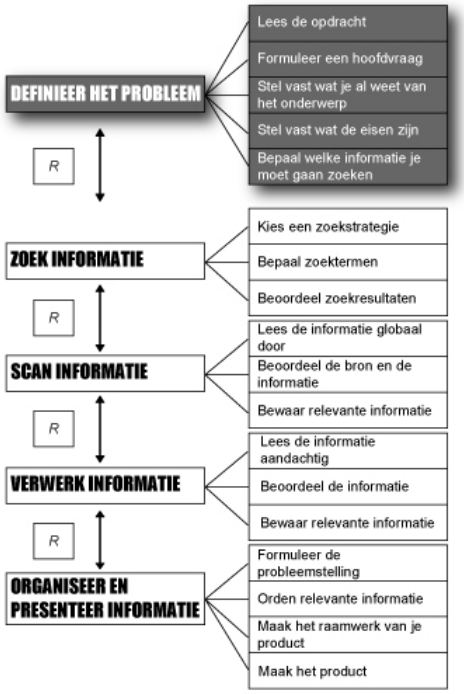
Opdracht Vrede van Versailles

Tijdens deze cyclus van drie lessen gaan jullie de besprekingen van Versailles, die de afronding waren van de Eerste Wereldoorlog, naspelen. Deze besprekingen vonden plaats in de eerste helft van 1919.

De klas wordt vooraf door de docent in twee groepen verdeeld, binnen elke groep spelen vier leerlingen de rol van een land. Dit zijn de belangrijkste landen die aanwezig waren bij de vredesonderhandelingen en de ondertekening van het verdrag van Versailles. De rest van de subgroep ondersteunt bij het voorbereiden van het spel.

Allereerst worden per groep de rollen verdeeld. Daarvoor moeten jullie eerst gaan uitzoeken welke hoofdrolspelers er waren in de onderhandelingen. Gebruik hiervoor je lesboek of het internet. Tijdens deze opdracht maak je gebruik van het proces voor het oplossen van informatieproblemen uit je theorieboek. Hieronder vind je het schema van dat proces.

Tijdens deze opdracht staat de fase 'definieer het probleem' centraal. Op deze lesbrief vul je de stappen in die je neemt in die fase. Straks wordt niet alleen het eindresultaat beoordeeld, maar vooral het proces, de manier waarop je tot je resultaat komt. Vergeet niet om voor je zelf alle stappen uit het proces te nemen, ook degene die je niet op dit formulier hoeft in te vullen!!



Lees de opdracht
Lees de opdracht goed door, en bepaal wie welk land straks in het rollenspel vertegenwoordigt.

Formuleer een hoofdvraag
Welke vraag moet je beantwoorden? Kijk nog eens op pagina 10 en 11 van het theorieboek informatieproblemen oplossen. Maak zo nodig een mindmap om je gedachten te ordenen. Je mindmap kun je hieronder tekenen.

Schrijf je hoofdvraag en eventuele deelvragen hier onder op:

……

……

……

……

Stel vast wat je al weet van het onderwerp
Weet je al iets over de Vrede van Versailles, en de rollen van de verschillende landen? Schrijf dat hier onder op. Je mag ook een mindmap maken.

Figure 1 Process worksheet

Discussions

At the end of every theme teachers and students had a discussion on evaluation criteria. The goal of these discussions was to develop a mind map or knowledge structure. Teachers received a manual for these discussions, based on the theory of Klausmeijer (1990). According to Klausmeijer several steps have to be taken when teaching concepts: (1) Providing an orientating instruction, by focussing students attention, pointing out the importance of the concepts to be learned, helping students developing a scheme of the concepts, providing students with a strategy to learn the concepts and creating the intention to learn the concepts. (2) Providing or eliciting a definition or defining characteristics. (3) Recalling what was learned from long term memory. (4) Using examples and non-examples. (5) Helping students with discovering defining characteristics. (6) Providing strategies for distinguishing examples and non-examples. (7) Giving feedback.

The manual prescribed how teachers should structure the discussions on evaluation criteria. The first discussion would start with telling the students that during the next 15 lessons they would learn more about criteria to evaluate information on the WWW. Next students were asked why they thought it is important to learn about criteria, and the teachers explained why he or she finds it important that students learn about criteria. The teacher then asked students which criteria they knew and used for evaluating information. The teachers drew these criteria on the blackboard and created a knowledge structure together with the students. When a student mentioned a criterion, the teacher asked for a definition of the criterion or characteristics of the criterion (i.e., a student mentions objectivity, and the teacher asks how can I tell if information is objective?). Students copied the knowledge structure in their notebooks. The discussion ended with the teacher stating that in the next 15 lessons they will do several assignments on the Internet, and that he will return to this knowledge structure every three lessons, and that he expects that they can enrich the structure together and that students are able to explain the criteria and know how they can use them more and more. This first discussion followed steps 1 and 2 of the theory of Klausmeijer (1990).

The discussions after this first discussion always followed the same routine: (1) Take a look at the knowledge structure, and let students summarize what they already know about criteria (e.g., step 3 of Klausmeijer); (2) Ask if students have new additions to the knowledge structure. This could be new criteria or additions to definitions of criteria); (3) Try to provide examples and non-examples of the criteria (So examples of an objective website, but also an example of a very subjective website) and discuss the differences and similarities between the examples (Klausmeijer step 4, 5 and 6), and (4) Discuss whether the criteria are equally important for every research question, or course.

Measurements

Evaluation hit list

To measure how students evaluated the hit list four information problems with a manufactured hit lists of 14 results on paper were developed. 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'. Per hit list, students had to select three sites they would open and three sites they would not open. They could highlight and circle parts of the results 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.

Evaluation of websites and information

To measure how students evaluated websites and information four information problems and booklets of eight printed websites each were developed. 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 behind the attacks of 9/11, and the second regarded the 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. Students were asked which sites and what information they would or would not use, given the problem provided and were informed that it was not impossible to find 5 to 10 features they based their decision on. They could highlight those features. If students had circled a certain area on the site or written down a comment like: 'Site is old' they received a point for recognizing the criterion, and if their evaluation was correct (e.g. the site was indeed old) they received another point. So, students could receive points for recognizing a criterion and for using a criterion in the correct way. Maximum score on all tasks was 200.

Learning results

To determine students' history knowledge a final exam about the topic 2nd World War was developed by one teacher. This test consisted of 10 content related open questions. Two of the questions also paid attention to information evaluation. A correction model was provided to the teachers in order to score the exams of their class in a correct way. Scores could range from 0 to 10.

Field notes

In each experimental class three lessons were observed. Field notes of these observations served as secondary material for possible explanations of the results. During these observations special attention was given to the interaction between the students and between the students and their teacher concerning evaluation behaviour and the use of evaluation criteria.

Reflective stories

To determine how the teachers had experienced this design-based research process the teachers of the experimental classes were asked to write a reflective story about the design process, educational design in general, designing with others, how the design works in practice, what they would do different next time, the influence of having to use a theory, and success and failure factors.

Design and Procedure

A pre-test post-test control group- design was used to determine the effect of the program on students' evaluation behaviour (e.g. evaluation of hit list, websites and information). Table 2 presents the design of the experiment.

Table 2 Design of the study

O1	X1	O2	*N*=80
O1	X2	O2	*N*=21

O1 = two tasks evaluation hit list (history and biology), two task evaluation information and source (history and biology).
X1 = intervention program (three observations per class)
X2 = regular lessons on the history content
O2 = two tasks evaluation hit list (history and biology), two task evaluation information and source (history and biology), final exam, reflective reports

Before the first lesson, all students (experimental and control condition) made a hit list and website evaluation tasks (one history and one biology). These tasks were counterbalanced and rotated for the pre- and post-test. Half of the students received the first history tasks (hit list and website) and the first biology tasks (hit list and websites) during the pre-test, and the remaining half received the second history tasks and the second biology tasks. Furthermore, half of the students started with the history task, and the other half started with the biology task. For completing the tasks students got a maximum time of 50 min. After the pre-test the experimental classes received the designed program and the control class received regular lessons on the 2nd World War. In each experimental class three lessons were observed by the first author. A week after the last lesson all students completed the parallel forms of evaluation tasks (different information problem). During the post-test students received a different task than the pre-test task. Students who had made task 1, now made task 2 and vice verse. Again, the order of tasks (starting with history or biology) differed between students.

Data analysis

Reflective stories. The reflective stories of the four teachers were read and utterances or paragraphs were grouped according to topic. Topics were the design process, educational design in general, designing with others, how the design works in practice, what they would do differently next time, the influence of having to use a theory, and

success and failure factors. The utterances were then summarized per subject and conclusions were drawn.

Results

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.

Table 3 Means and standard deviations of hit list en website evaluation task score

	Experimental condition (N=80) Mean (SD)		**Control condition (N=21) Mean (SD)**	
	Pre test	Post test	Pre test	Post test
Hit list history	3.2 (2.2)	3.2 (2.3)	3.8 (1.7)	1.3 (2.0)
Websites history	14.9 (8.8)	18.8 (9.5)	14.8 (7.7)	12.8 (7.6)
Hit list biology	3.1 (2.1)	3.4 (2.3)	4.2 (1.4)	1.6 (2.2)
Websites biology	16.5 (9.0)	18.4 (8.4)	14.2 (9.4)	15.2 (9.2)

Effects of the instruction

To answer the first research question on the effects of instruction on students' use of criteria for evaluating results (hit list) a repeated measures ANOVA analysis with condition as between factor was performed for the history task (the domain of the instruction). The analysis on the evaluation of the *history hit list* showed a significant main effect of 'time', $F(1,99) = 14.60$, $MSE = 49.56$, $p = .00$, $\eta^2 = 0.13$, and no main effect for 'condition', $F(1,99) = 2.21$, $MSE = 13.36$, *ns*. There was a significant interaction effect between 'time' and 'condition', $F(1,99) = 14.30$, $MSE = 48.55$, $p = .00$, $\eta^2 = 0.13$. This interaction effect is not caused by an increase in scores of the experimental classes, but by a decrease in scores in the control class. The scores in the experimental classes remain constant. No class effects were found between experimental classes.

The second part of the research question focuses on effects of instruction on students' use of criteria for evaluating information and source (website). Again, a repeated measures ANOVA was performed with the results on the *history websites evaluation task* and condition as between factor. No main effect of 'time' was found, $F(1,99) = .73$, $MSE = 32.36$, $p = .40$, $\eta^2 = 0.01$, and also no main effect for 'condition' was established, $F(1,99) = 2.73$, $MSE = 307.30$, *ns*. However, a significant interaction between 'time' and 'condition' was found, $F(1,99) = 6.46$, $MSE = 287.29$, $p = .01$, $\eta^2 = 0.06$. Scores in the experimental classes increase over time, while scores in the control class decrease. No class effects were found between experimental classes.

Transfer effect of instruction

To answer the second research question, whether instruction based on two transfer theories achieves transfer, a repeated measures ANOVA analysis with condition as

between factor was performed. This analysis on the evaluation of the *biology hit list* showed a main effect on 'time', $F(1,99) = 11.62$, $MSE = 45.12$, $p = .00$, $\eta^2 = 0.11$, and no main effect on 'condition', $F(1,99) = .75$, $MSE = 4.11$, *ns*. There was a significant interaction between 'time' and 'condition', $F(1,99) = 19.34$, $MSE = 75.06$, $p = .00$, $\eta^2 = 0.16$. This interaction effect is not caused by an increase in scores of the experimental condition, but by a decrease in scores in the control condition. Furthermore, a class effect was found between the four experimental classes, $F(3,76) = 2.68$, $MSE = 10.22$, $p = .05$, $\eta^2 = 0.10$. Post-hoc analysis revealed that the scores in one class decreased from pre to post-test, while the scores in the other three classes increased from pre to post-test.

To analyse the effects of instruction on transfer of evaluation of information and source an ANOVA was performed with the results on the *biology websites* evaluation task and condition as between factor. No main effect on 'time', $F(1,99) = 2.00$, $MSE = 69.95$, $p = .16$, $\eta^2 = 0.02$, and also no main effect on 'condition'', $F(1,99) = 2.00$, $MSE = 244.23$, *ns,* was found. There was no interaction effect between 'time' and 'condition', $F(1,99) = .19$, $MSE = 6.74$, $p = .66$, $\eta^2 = 0.00$. This indicates that there is no significant difference in scores between the students in the experimental and control condition, and thus no transfer effect was found. No class effects were found in the experimental classes.

Learning results

The average score on the final exam was higher (M = 6.3, SD = 1.1) in the experimental classes than in the control class (M = 5.8, SD = 1.1). This difference was significant, $t(99) = 1.97$, $p = .05$, $r = .19$.

Field notes

Experimental class A: the students. During the first lesson, students were confronted with websites they had to use to solve a task. Some of these websites contradicted each other or provided information on a different subject. Students noticed that there was something wrong with the sites, but did not seem to adjust their actions accordingly. They wanted to finish the assignment. In the discussion that followed the assignment, students mentioned criteria they used to select information. In general, things like using a site they had used before, the site has to look good, there have to be sources or references mentioned on the site, were mentioned by them.

During the next lessons, it became clear that some students in this class caused problems and had a negative influence on the other students. During the final lesson, where students presented good and bad websites to each other, students showed that despite the fact that they did not really showed it during the lessons, they actually had developed knowledge on how to evaluate websites (they could mention more and more sophisticated criteria) and were able to discuss websites with fellow students and defend their choice for a good or bad website.

Experimental class A: the teacher. During the group discussions, the teacher did not follow the manual. Since his students had some difficulties with the assignments, the

teacher had to spend more time on summarizing the historical content. Therefore, little time was left for discussions concerning the use of criteria when evaluating results, sources and information. The teacher did not check whether students filled out the process worksheet correctly.

Experimental class B: the students. During the first lesson, students worked concentrated. Many students walked into the Wikipedia-trap. A Wikipedia site was presented as the first site, but this site did not give information about the task at hand (The treaty of Versailles, 1919), but provided information about a different treaty of Versailles, in 1782. Most students used this Wikipedia site to complete the assignment. During the group discussion students mentioned that they found contradictory information, sites without references and sites with poor lay-out. One student remarks: 'If I have no knowledge on a subject, how do I know if the information on the site is reliable?' This is exactly the critical attitude we wanted to achieve with this first lesson. During a group discussion in one of the following lessons only a small group of students was prepared to discuss, while others remained silent. Not all students were happy with the new lessons; they feared that, since they had to find their own information, everyone would learn something else. These students would rather have more classical lessons. One student remarked: 'I have never wondered about information not necessarily being correct, I believed everything and now they make me distrust everything!' During the final lesson, students were able to discuss websites with fellow students and defend their choice for a good or bad website.

Experimental class B: the teacher. The teacher had to adjust the lesson program due to illness and classes being cancelled for meetings. The teacher tried to give the lessons according to the descriptions and followed the manual during the discussions on the use of criteria when evaluating information. Due to cancellation of lessons, the group discussions were not held as frequently as planned. The teacher did not check whether students filled out the process worksheet correctly.

Experimental class C: the students. Students in this class were not very enthusiastic about the new lessons. They felt they already knew everything about the WWW. When asked what they had noticed about the websites during the first lesson, they mentioned that there were no references on the sites and layout was bad. Some students felt they shouldn't use Wikipedia because it is an open source, others mentioned you can use Wikipedia because it is better than most sites and the content is checked. During this discussion some students were busy playing games or talking to each other. Only a few students were active. This didn't change in the following lessons; most students were not motivated and complained about the assignments. During the final group discussion in the last lesson, students were not able to mention more criteria than the few they mentioned during the first lesson.

Experimental class C: the teacher. The teacher in this class tried to follow the manual for group discussions but was not always sure how to structure the discussions. The teacher admitted not feeling competent enough for leading the discussions, since developing a knowledge structure in this way was new to this teacher. The teacher encouraged the children to fill out the process worksheets.

Experimental class D: the students. Due to a miscommunication the first lesson was not given according to plan. Students had already answered the questions, without

receiving the sites they had to use. When they were confronted with the sites for this assignment, reading and evaluating the sites was done very quickly, since they had already answered the questions. After a while students were asked to explain what they noticed about the sites. They mentioned that a newspaper is not always a reliable source, and that information differed between sites. In the next lessons, students seemed to understand what the lessons hoped to accomplish and why evaluating information is important, but they admitted they did not want to change their usual ways. They felt it took them more time to finish an assignment if they filled out the work sheets and evaluated every site. During the group discussion in the final lesson, some students were able to mention more criteria; some students did not engage in the discussion.

Experimental class D: the teacher. There was not always time to have a discussion after three lessons. The first time, the teacher mentioned more criteria than the students. During the final discussion this had improved and the teacher made sure the students did the work. The manual was not always followed. The teacher did not check whether students filled out the process worksheet correctly.

In summary, the instructional program was only partly executed as planned. Most teachers did not check whether students filled out the process worksheet correctly and there was too little time for the group discussions on evaluation criteria every three lessons. For students the used approach in the lessons was new and they had to get used to it. Although students did not seem to change their way of evaluation results, information and source, during the lessons, the final discussions showed students indeed had more knowledge on evaluation criteria.

Reflective stories

The perceptions of the teachers, as stated in their reflective stories are summarized per subject.

The design process. Teachers differentiate between the first and second design cycle. During the first design cycle, the small design teams and regular contact with each other were appreciated. Discussing the lesson plans, evaluating the lessons, adjusting the lessons and reflecting on each others work was made easy by short communication lines. Every teacher was involved with the design of all of the 15 lessons. Designing was inspiring.

Designing during the second design cycle was more an individual than a group process. The design of the lessons was divided in 5 themes and three teachers were responsible for one theme, one teacher was responsible for two themes. So, only a minority of the lessons was designed by a teacher him- or herself, while the majority was designed by one of the colleagues. There was not enough time to really discuss the lesson plans and reflect on each others work and there was also less contact with the researcher than in the first cycle. Teachers expressed that they did not have time to learn enough about the transfer theory they had not used in the first cycle. For instance, if a teacher had not used the group discussions to develop a knowledge structure during the first cycle, and the new program required the teacher to have these

discussions, the teachers experienced that they were not prepared enough. The program of the second cycle was not a group product, but consisted of a sequence of lessons designed by several teachers. One teacher remarked that the second cycle paid more attention to didactics than content. Other remarks made by teachers were that it was easy to integrate the WWW into the lessons and that the steps to be taken to solve an information problem took some time to get used to. Furthermore, teachers enjoyed designing the programs and being more involved with their domain. They did not have enough time to design education.

Designing with others. Working together with colleagues was very much appreciated and resulted in an abundance of ideas. Teachers learned from each other and found designing together very inspiring. Their colleagues provided the teachers with ideas for lessons they could also use in other grades besides the target group.

How the design works in practice. Again, a distinction has been made between the first and the second design cycle. Teachers expressed that the lessons in the first design cycle were implemented better than the lessons in the second cycle. This was due to the mentioned fact that *all* lessons of the first cycle were designed by the teachers themselves, while most of the lessons in the second cycle were designed by one of their colleagues.

The lessons in the first design cycle were probably filled with too many tasks. This resulted in students not being able to finish all tasks and students complaining about all the computer work. Teachers mentioned that they asked for more 'normal' lessons. Most students also needed a few summarizing lessons with the textbook.

In the second design cycle these summarizing lessons were part of the program, and there was more variation in tasks. Students mentioned to the teachers that they appreciated these lessons. However, not all students were actively involved with the lessons, and most students were fixated on form and not on content. There was a card game for instance, that was supposed to teach the students the order of events in the Second World War. But students were more focused on playing the game than on the final assignments of putting the events in the correct order.

Both designs were able to teach students how to evaluate information, but teachers doubt whether this knowledge will still be used after a few months and whether it was transferred.

Points of improvement. Teachers mentioned that they preferred the approach of the first design cycle, and would make sure next time that they know all the ins and outs of the lessons designed by their colleagues. Having to integrate a new transfer theory required more attention and time. Important for the students would be more variation in tasks and paying more attention to the reasons for the program, e.g., why do students have to learn how to evaluate results, information and source?

Designing instruction using a transfer theory. For designing instruction to foster a complex cognitive skill it is important to get grip of all the ins and outs of this skill. Teachers remarked that a short course about the steps to solve an information problem would help to gain more insight in the skills involved. The transfer theories were clearly stated and teachers experienced that it was not difficult to put these theories into the lessons plans.

Success and failure factors. Teachers stated that a better combination of textbook and WWW assignments would achieve better results. The textbook was hardly used in both the designs, and should be used more often in a new design. Furthermore, the teachers mentioned that more teachers in the school should have to use this method. All teachers should experience the importance of evaluating information on the WWW and point out this importance to their students. It is important to look at the long term results. Another important aspect was that adjusting students' search process by letting them fill out process worksheets was seen as very difficult. Students remarked that they have developed their own process and should have learnt this new, correct process much earlier than in the 9th grade.

Discussion

Design-based research was used to develop and test an educational program to teach 9th grade students not only historical content, but also to evaluate search results (a hit list), information and source (information on a website and the website itself) when searching for information on the WWW. Furthermore, the program should also stimulate students to use these skills in a variety of settings (e.g., lead to transfer). Effects of the program on knowledge and use of criteria in the domain of instruction, on the transfer of the skills to another domain, and on the learning results concerning history content were determined. Furthermore, the participating teachers were asked about their perceptions on educational design and design experiments.

We can conclude that instruction improved students' evaluation of information and source (websites). Students' scores of the experimental condition increased from pre- to post-test, while the scores of the control group decreased from pre- to post-test. Instruction seems to have had no effect on students' evaluation of search results (hit list); students in the experimental condition maintained their scores from pre- to post-test and the scores of the students in the control condition decreased from pre- to post-test. This could not be due to the difficulty of the tasks used during the post-test, because tasks were counterbalanced and rotated. An explanation could be that since the lessons were not always executed according to the lesson plans and the discussions were mostly focused on the evaluation of websites and information, students' use of criteria of evaluating hit lists was not triggered. This can explain the maintenance in scores of the students in the experimental condition. The scores of the students in the control condition decreased; this could be due to the fact that only one class participated in the control condition. Less time, or other interferences could have biased the results.

Furthermore, it can be concluded that the instruction did not achieve a transfer effect with regard to the use of criteria for the evaluation of websites. There was no difference in the scores on the post-test between students in the experimental classes and the control class. It was expected that the experimental classes would score higher on the post-test, if the program had led to transfer. Instruction did not have a transfer effect on students' use of criteria for evaluating search results as well. Students in the experimental condition maintained their scores from pre- to post-test and the scores of

the students in the control condition decreased from pre- to post-test. The same explanation as stated above holds in this case. While it is positive that instruction improved students' evaluations of websites, it was unexpected that transfer was not achieved. The program was based on two transfer theories and the combination of both theories was hypothesized to achieve transfer. The results of the earlier study (Walraven et al., 2008a), in which the two programs, each based on one transfer theory, gave rise to positive transfer effects, led the design team to believe that a combination of the strong points of both programs could even give rise to more transfer. The most likely explanation for not achieving a transfer effect is that the new program was not implemented to the full extent in all experimental classes. This is confirmed by the observations and the reflecting reports of the teachers concerning the design cycles and the implementation of the program.

Another question concerned the effect of the instruction on students' learning results. It was hypothesized that the scores on the final test concerning the history content would not differ between the two conditions. Teachers of the experimental classes had some doubt about embedding the evaluation skills, because it would mean less time for history content, and as a consequence maybe lower grades for history. But results reveal that the score on the final test of the experimental classes is significantly higher than the score of the control class. So, embedding evaluations skills does not cause lower grades. Last but not least teachers' perceptions concerning the design cycle were determined. Results reveal that the first design cycle, where teachers had more sessions together and developed all the lessons they would implement in their class themselves, was preferred above the second cycle. The second cycle was found to be too short and provided not enough information on the transfer theories. Therefore, teachers felt that the programs of the first design cycle were executed better than the program of the second cycle. Teachers also stated that transfer cannot be fostered by this program alone; teachers from all courses in school should integrate evaluation tasks in their lessons.

Regarding the implementation of the program, the design team believed that short process worksheets and structured discussions to develop a rich and well connected knowledge structure would be the best way to achieve transfer. In the first design cycle, the long process worksheets caused problems for the students (Walraven et al., 2008a). The team expected that if they provided students with shorter worksheets in the current study, students would not rebel against the worksheets, and use them in the intended way and transfer would be fostered. Unfortunately, the teachers did not encourage the students enough to fill out the sheets. Some students did not fill out sheets at all, others filled out sheets after completing the assignment in stead of during, and only a few students filled out the sheets correctly.

Another important educational measure in the program was the use of discussions on evaluation criteria to develop a rich and well connected knowledge structure about evaluation criteria. Important in these discussions was paying attention to the use of the evaluation criteria in various contexts or the way criteria are connected. A manual for the teachers had to make sure that all discussions were held in the same way and all aspects would receive the right amount of attention. Unfortunately, the discussions were often shortened due to time constraints and teachers did not always follow the manual.

The fact that two crucial factors of the educational program (filling out the sheets and the group discussions) were not completely implemented according to plan could – as stated – provide an explanation for the fact that no transfer was achieved. That the program was not executed according to plan is confirmed by the teachers who mentioned that the execution of this second program was more difficult than the execution of the first program.

A limitation of the study is the difference in number of students in experimental condition and control condition. There were four experimental classes and only one control class. Future research concerning the effect of the – further refined- program should be set up according to a more experimental design. Another focus of research could be testing whether the improvement in evaluations in the domain of instruction is still present a few months after the last lesson. It is not enough that students improve from pre- to post-test, students should acquire and use these evaluation skills throughout their lives. The role of the teacher to accomplish this should be further investigated. The teacher is an important factor, because the teacher is the one who must make students aware of the importance of evaluating information when searching the WWW for information. Future research should explore the role of the teacher in more detail and test whether being stricter to students with regard to the process worksheets influences the results. Next, the role of all teachers in the school should also be investigated. Integrating this evaluation skill throughout the curriculum is essential to foster transfer and prepare students for lifelong learning.

As mentioned in the introduction, the results of the fourth and last phase of the design research cycles as Reeves (2006) described, the design guidelines that can be distilled from this research cycle will be provided now. First, the design should be a result of a group process. Every member of the team that has to implement the program should have the same amount of knowledge on all the lessons. It is not enough for teachers to receive short instructions by a colleague on how to give a lesson. When the lessons are new for the students and the teachers this can result in lessons not being executed as planned. Second, when integrating two designs into one new design, it is important that the theory behind the two designs is thoroughly discussed between the team members and that knowledge on both previous designs is equal between the team members. It is important to make sure that all teachers have knowledge on all theories in the new program and can use methods based on these theories. Third, instruction in evaluation skills should have a good balance between text book lessons and tasks that require students to search for information on the WWW. Students need structure and become bored with too many WWW tasks. Fourth, teaching for transfer should be based on more than one program. Although the programs of the first design cycle achieved transfer to another domain, it was not tested whether the skills would still be used after time. It is critical that teachers in different domains pay attention to these skills and integrate instruction on these skills in their lessons. An important first step would be teaching teachers how to evaluate and how they can support their students to become critical web searchers.

CHAPTER 7

GENERAL DISCUSSION

General Discussion

Main findings and conclusion

The context of this thesis is information-problem solving, applied to information sought and found on the WWW. This is a complex cognitive skill that includes several consecutive steps and entails the evaluation of the search results that are presented on a hit list, the websites and information on those sites that have been selected from the hit list. The aim of this thesis was twofold: to gain theoretical insight in evaluation criteria websearchers use, and to design an educational program based on two transfer theories meant to teach students how to evaluate search results, information and source on the WWW. Gaining theoretical insight and instructional design are closely interrelated in this thesis, as improved theoretical insights informed design and the outcome of design application raised theoretical questions.

A literature review study (Chapter 2) and two empirical studies (Chapters 3 and 4) were conducted to gain insight in how students search the WWW and how they evaluate the usability, reliability and verifiability of websites and information while searching. The literature study showed that children, teenagers and adults have difficulties with specifying search terms, evaluating search results, evaluating information, evaluating the source of information, and regulating their search process. Regarding the evaluation of search results (hit list) young children appeared to use a search engine as if it were a magical machine; they expect the machine to provide them with the complete answer. Teenagers view every result without a clear evaluation of the results. Like teenagers, the strategy chosen by unsuccessful adults was to inspect the search results in the order they were presented. Successful adult searchers, who have considerable background knowledge, evaluate search results by looking at the title, the origin of the source, the description, and useful information or identifiers in the URL such as '.edu' or '.com'. Most groups partcipating in these reviewed studies evaluated websites and information solely based on usability of the expected information, and not on quality aspects such as validity, reliability and recency.

The outcomes of the studies described in Chapters 3 and 4 confirm these earlier results; students and teachers rarely explicitly evaluate search results, information found and source when searching the WWW for information; when they did so, they used criteria like title/summary, connection to the task and appearance of the site. Students and teachers knew more criteria than they used while accomplishing two tasks, and teachers mentioned more criteria than students. When interviewed after task performance they mentioned criteria like author, references and information agrees with more sites, but they did not use these criteria when actually searching the WWW. These results are in line with research by Koot and Hoveijn (2005), Lorenzen (2002), Monereo et al. (2000) and Rosell-Aguilar (2004). Confronting students with the observation that they mostly used superficial criteria and did not use all the criteria they knew, they mentioned that this was due to time pressure, motivation and convenience. If information that could possibly answer the question was presented on a website that looked ok, students did not bother to look any further. They said they would pay more attention to the criteria when they had to write an essay or when they were searching

for something they were really interested in. Surprisingly, teachers showed the same behaviour; they said that when they would have searched for real, they would have taken more time, and read and evaluated information more in depth. Within the context of the experiment they scanned pages quickly and 'showed student behaviour'; they felt that they had to choose between giving an answer as soon as possible, or doing a decent job.

The studies not only gave insight in students' and teachers' search behaviour, they also yielded a list of criteria, categorized in usability, reliability and verifiability, for the evaluation of results, information and source on the WWW. This set of criteria was derived by analysing think-aloud protocols of students' and teachers' task performance in different domains. Students and teachers in both studies accomplished two tasks in two domains. The results revealed that the approaches for evaluating search results (hit lists), information and source (websites) did not differ between domains. This suggests that these evaluation processes and the use of the derived criteria are generic in nature, or at least have common characteristics. However, knowing when certain criteria are more important to use than others in different situations and the flexible use of the criteria is what makes a websearcher a critical websearcher. So, knowing the categories and criteria is not enough, it is also important to have more sophisticated knowledge on the use of criteria. Furthermore, having domain knowledge influences how criteria are applied. For instance, a person who knows all the important scientists in the field of biochemistry will deal differently with the criterion 'author' than a novice, who will use this criterion at a more global level.

A distinguishing feature of the present studies was that spontaneous evaluation processes during information-problem solving were investigated, in contrast to most other studies done in the field of information-problem solving. In previous studies (e.g., Britt & Aglinskas, 2002; Graesser et al., 2007; Stadtler, 2006) the standard methods of data collection was prompting information evaluation by means of the experimental instruction. However, using prompts for evaluation may result in experimental artifacts and in an overestimation of naturally occurring evaluation processes. Scenarios and instructions which explicitly prompt evaluation processes need to be avoided when studying spontaneous evaluation processes, participants should not even be aware of the research interest in evaluation processes. Our studies, in which participants were not prompted to evaluate and did not know what the researcher's interests were, revealed that students and teachers use fewer criteria when searching for information, than they know. Prompting could have resulted in an increase of experimental artifacts.

The insights on use of evaluation criteria from Chapters 3 and 4 were input for the design, implementation and evaluation of two educational programs (Chapter 5). These two programs were each based on a different transfer theory. The first transfer theory was the high road theory of Salomon and Perkins (1989). This theory states that students have to be stimulated to pay explicit attention to the various steps in a process that have to be taken and to the way these steps can be used flexibly in different situations. Process worksheets were used in the design of the program to stimulate flexible use of the steps. The second theory was the theory of Simons et al. (2000), which emphasizes the importance of a good, extensive and well-organised knowledge base and the domain specific interpretation of the skills. The development of this knowledge

base was fostered by using mind mapping techniques and discussions. The effects of both programs on students' evaluation skills and on the transfer of the skills to a new domain were determined. Success and failure factors of the way the transfer theories were implemented were analysed and the success factors were combined in a new designed program. Effects of the new program on students' evaluation skills in the domain of instruction and in a transfer domain, as well as the effect on students' learning performance (history exam) were determined (Chapter 6).

Results of these studies revealed that the program based solely on the high road theory, the program based solely on the rich representation theory and the program based on success factors of the previous designed programs improved students' evaluations of information and source in the domain of instruction (history). However, none of the programs improved the evaluation of search results (hit list). An explanation could be that during the lessons the teachers spent more time on evaluation of websites and information and due to lack of time the assignments with regard to the evaluation of hit list got less attention.

The goal of the programs was, besides improving the use of evaluation skills in the domain of instruction, achieving transfer of the evaluation skills to a different domain. The programs that were based on one transfer theory (i.e., the high road theory or the rich representation theory) both led to transfer. Students improved their evaluation of information and source in the transfer domain (biology). Evaluation of search results again did not improve. So, providing students with process worksheets (the high road program) or having group discussions to create a rich and well-connected knowledge structure (the rich representation program) fostered transfer. This leads to the conclusion that the translation of the two theories to the didactical measures used are effective. Process worksheets can foster the systematic approach of information-problem solving in a flexible way, and discussions and mind mapping technique lead to a well defined knowledge base on evaluation criteria.

However, the improvement in students' scores concerning the evaluation of websites and information was far from optimal in both programs. It was hypothesized that by improving the process worksheets and structuring the group discussions about criteria, and combining the success factors of the two programs in a new program, the transfer effect would increase. This was also based on the assumptions that in order to learn complex cognitive skills and foster transfer of these skills students need not only a good knowledge base (mental models) but also need to have well established cognitive strategies (van Merriënboer, 1997). Unfortunately, the results of the final experimental study revealed that no transfer was achieved. This unexpected result could be due to the fact that the program was not implemented as planned in all classes. The teachers did not encourage the students enough to fill out the process worksheets. Some students did not fill out sheets at all, others filled out sheets after completing the assignment instead of during and only a few students filled out the sheets correctly. Furthermore, the teachers did not lead all the discussions according to the manual and discussions were often shortened due to time constraints. The teachers confirmed that the program was not executed according to plan, and mentioned that the implementation of this new program was more difficult than the execution of the program based on one transfer theory, because they had not designed all the lessons

themselves. This last study showed that the correct implementation of the program is crucial. So, there is a difference between a theoretical design and putting theory into practice. However, it cannot be concluded that the combination of theories to design instruction does not work, since the program was not implemented correctly. More studies are needed to examine whether the combination of success factors of the programs based on one theory actually can improve transfer (i.e., whether the transfer of the combination equals the sum of both transfer effects) or whether the combination inhibits the positive effects of both programs.

The design and implementation of the designed programs followed a design-based research approach. In the first cycle, the two programs based on one transfer theory were designed and in the second cycle one program was designed based on the results of the first cycle. Several conclusions can be drawn from these two cycles. First, knowledge on transfer theories behind the program played a role in the correct implementation of the programs. During the second design cycle too little attention was paid to these transfer theories. This meant that teachers who had used the high road theory in the first cycle did not have enough information about the rich representation theory and vice versa. For design-based research this means that although a design team remains the same throughout the various cycles, it is important to re-establish the common knowledge of a team at the start of every cycle. Furthermore, teachers should have a good overview of the lessons that have to be implemented, especially when they did not design the lessons themselves. In their reflective stories, teachers mentioned that this was a weak point of the second cycle. They found the first design cycle more inspiring and satisfactory, and that the implementation of the first program was easier than the second, since they knew all the ins and outs of the first program and were less familiar with the lessons designed by their colleagues. It was not enough to provide teachers with a manual. Finally, it can be concluded that more design cycles are necessary to adjust the design to practice and for teachers to become familiar with the program. Perhaps the step from two programs based on one theory to one program that combined the success factors of both programs was too big.

The two aims of this thesis (theoretical insight and designing instruction) have a complex relationship with each other. Theory does not guarantee a good design as the non-optimal outcomes in Chapter 5 and 6 showed. Conversely, testing two contrasting designs in practice opened our eyes for the non-stated and non-questioned assumptions underlying the two theories and the question was raised whether the scope of both theories was the same, more specifically whether both approaches would lead to the same amount, extent and sustainability of transfer. Testing a program in practice and adjusting it according to the results is a necessary step in the design process and should be implemented in every design process.

The two transfer theories are widely used for educational design, but are not design theories themselves. Although the high road transfer theory stated that providing the required steps of the process could be effective, it did not describe how these steps could be provided. Our initial design showed that providing students with too many steps could cause problems. So, when designing instruction based on transfer theories, it is important to translate the theory to practice and ask questions about how the theory could be translated and what the most effective translation could be.

Practical implications

When students are to become critical websearchers, it is important that first, their teachers are aware of what constitutes a critical websearcher, and how they can stimulate their students to become one. Students and teachers should be made aware of the importance of evaluating search results, information and source, and must learn how to use different kinds of evaluation criteria when searching the web for reliable and useable information.

A first important aspect of designing instruction in evaluation of search results, information and source is embedding the instruction in the curriculum. Providing students with whole tasks, emphasizing domain content as well as evaluation skills, improves students' domain knowledge as well as their evaluation skills. The tasks should be constructed in such a way that the entire information problem solving process (from the problem definition to organization and presentation of the information) is covered. At the same time, it should be realized that students could be overwhelmed by such tasks, where all aspects of the task may be new to them. Working with completion problems in which part of the steps are already filled out can help students to focus on certain steps in the process and not lose sight of the whole process (e.g., van Merriënboer, 1997). This can be done by giving students process worksheets in which certain steps are already taken. The sheets can also be designed in such a way that they give more or less support depending on the students' need. Furthermore, when students search information and accomplish these tasks, it is important that the teachers make sure that these tasks are well performed, that performance standards are met and that the content knowledge students acquire is sufficient to pass the final exam.

Implementing a program, with an emphasis on searching for information and on whole tasks, requires adaption from both teachers as well as students. Teachers must adapt to this new approach and use different didactics. This will take time and adjustment when evaluations reveal that the instruction does not exactly fit the students' needs. Designing and implementing instruction on evaluation skills should be done in several rounds, so teachers become familiar with the new approach and can learn to adapt the lessons to the situation at hand. Students also have to learn to cope with their new role; instead of receiving the information they need, they have to search for it themselves. This can cause anxiety in students; they fear they do not learn what they are supposed to learn when they have to find information themselves. Furthermore, students have to make an epistemological shift from the view that information can be trusted, towards a view that information has to be evaluated (Mason & Boscolo, 2004). To accomplish this, students need to be convinced of the necessity of evaluating information. Similar to the adaptation of teachers, time is an important factor in the adaptation of students to a new approach. When students work for some time with the approach, they will become more comfortable with information-problem solving and evaluating information and have solved several tasks, and their anxiety will fade.

The tasks students receive should have a certain degree of variation. The programs in this thesis used role playing, card games, drawing a comic and writing a newspaper article for instance. This is one way to realize variability. Also the use of different kinds of information problems in terms of fact finding, finding definitions of

concepts, relations between concepts or the use of wicked problems can make students use the evaluation skills in a flexible way. The mind map technique is very helpful in providing students the insight in the relation between criteria and discusses the difference in use of these criteria in different settings. Using a variety of tasks can also foster transfer.

Future research

The present studies could not tell us whether one transfer theory provided better measures to stimulate transfer than the other. The programs based on one transfer theory led to a similar level of transfer. However, no information is available about long-term effects. Yet the two theories might have different long-term implications since research in other domains shows that students remember information and skills longer when they have a well developed mental model (Woods, 2007). Future research should determine whether there are differences in the long term effects of the two transfer theories.

The program with the combination of transfer theories did not yield any transfer. Difficulties with the implementation and earlier findings (van Merriënboer, 1997) preclude that we draw the conclusion that it does not work. Future research should address whether instruction designed based on the combination of both theories actually improves transfer (i.e., whether the transfer of the combination equals the sum of both transfer effects) or whether the combination inhibits the positive effects of both programs.

Another interpretation problem results from the kind of measurement used. A paper and pencil test on criteria is not the only way to measure improvement of evaluation skills. It is a good way to test a large group of students' use of criteria. But besides being able to point out features connected to criteria on a site, and point out which information they would use, it is desirable to get more insight in the information-problem solving process of students. The studies in this thesis used think aloud protocols to uncover the problem solving process. Drawbacks of this method are the fact that not everyone is capable of thinking aloud and think aloud protocols only reveal part of the problem solving process. Sometimes participants verbalise their actions and not their thoughts (Young, 2005). Furthermore, thinking usually happens faster that verbalising thoughts. It is possible that while verbalising one thought, participants have more thoughts, but cannot verbalise all of them. And if they would verbalise every thought, it is possible that they would stop task execution. Finally, a first evaluation of information or a website is done very quickly and only leads to verbalising when the information or a website has striking features. Recording eye movements of participants while solving a task and afterwards replaying these movements for them and letting them think aloud while watching their eye movements for instance on half speed could be a way of giving the participants more time to verbalise their thoughts. This cued retrospective reporting (van Gog, Paas, van Merriënboer, & Witte, 2005) has been proven to be more effective than concurrent reporting, and thus could result in a better insight in the use of criteria and students' problem solving process. However, cued retrospective reporting

can only be used with short tasks of about 10 min, since replaying longer tasks makes that students can not remember their thoughts anymore.

Since 15 lessons is relatively short, and the lessons were packed with many assignments in which students had to search for information, future research should also address whether a program of perhaps one year, with only a few lessons per month concerning searching the WWW would be more effective. Furthermore, it would be wise to discuss with teachers in different domains how they could integrate WWW assignment in their lessons throughout the entire curriculum. Finally, it should be investigated what motivates students to use evaluation skills. Students may be motivated extrinsically by receiving grades for their evaluations. Instead of this esteem by teachers, one could also think about esteem by peers. Students could teach and stimulate their peers. Perhaps when students get the chance to show their peers their evaluation skills, they become more aware of their skills and will want to show off their skills more and more. Future research should also focus on uncovering students' intrinsic motivation to evaluation information. The studies in this thesis provide insight in the evaluation criteria involved when searching the WWW for information and although instruction designed according to different transfer theories yielded a transfer effect, this effect was not established when instruction was designed based on a combination of theories. Although, some questions remain unanswered and new questions came up, this thesis provides guidelines for fostering students to become critical websearchers.

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.

Bannert, M. (2004). Designing metacognitive support for hypermedia learning. In: H. Niegemann, D. Leutner & R. Brünken (Eds.), *Instructional Design for Multimedia-Learning.* (pp. 19-30). Münster: Waxmann.

Barker, J. (2005). *Critical evaluation of resources*. Retrieved May, 16, 2006, from UC Berkeley Library Web site http://www.lib.berkeley.edu/TeachingLib/Guides/Evaluation.html

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

Beck, S. (1997). *"Evaluation Criteria." The Good, The Bad & The Ugly: or, Why It's a Good Idea to Evaluate Web Sources*. Retrieved May 10, 2006, from http://lib.nmsu.edu/instruction/evalcrit.html

Beljaarts, M. (2006). *Internet, een populair medium voor het zoeken van informatie bij schoolopdrachten [Internet, a populair medium for searching information for school assignments]*. 's-Hertogenbosch: Malmberg.

Berner, E. S., McGowan, J. J., Hardin, J. M., Spooner, S. A., Raszka Jr., W. V., & Berkow, R. L. (2002). A model for assessing information retrieval and application skills of medical students. *Academic Medicine*, *77*, 547-551.

Bilal, D. (2000). Children's use of the Yahooligans! Web search engine: I. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society of Information Science*, *51*, 646-665.

Boekhorst, A. K. (2000). *Informatievaardig worden in het onderwijs, een informatiewetenschappelijk perspectief: Een vergelijkende gevallenstudie in Nederland en Zuid-Afrika [Becoming information literate in education, an information science perspective: a comparative case study in The Netherlands and South Africa*]. Unpublished dissertation. Retrieved May 10, 2006 from http://cf.hum.uva.nl/akb/publicaties/prom-akb-tot.pdf

Boekhorst, A. K. (2003). Becoming information literate in the Netherlands. *Library Review*, *52*, 298-309.

Brand-Gruwel, S., van Meeuwen, L., & Van Gog, T. (2008, June). *Surfen op de informatieoceaan: beoordelen van informatie op Internet [Surfing the information ocean: evaluating information on the Internet]*. Paper presented at the 35th Onderwijs Research Dagen (ORD). Eindhoven, the Netherlands.

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.

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*, 487-508.

Bransford J. D., Brown A. L., Cocking R. R. (Eds.) (1999). *How People Learn: Brain Mind, Experience, and School.* Washington, DC, National Academy Press.

Bransford J. D., Brown A. L., Cocking R. R. (Eds.) (2000). *How People Learn: Brain Mind, Experience, and School.* National Academy of Sciences. National Academy Press, Washington, D.C.

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., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*, 485-522.

Britt, M. A., & Gabrys, G. L. (2000). Teaching advanced literacy skills for the World Wide Web. In: C. R. Wolfe (Ed.), *Learning and teaching on the World Wide Web* (pp. 73-90). San Diego: Academic Press.

Brown, A. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist*, *52*, 399–413.

Brown, A.L. (1992). Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *Journal of the Learning Sciences, 2*, 141-178.

Buzan, T. (1995). *The mind map book*. (2nd edn). London, BBC Books.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

Colaric, S. M. (2003). Instruction for web searching: An empirical study. *College and Research Libraries*, *64*, 111-122.

Collins, A. (1992). Towards a design science of education. In: E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). Berlin: Springer.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the draft of reading, writing and mathematics. In: L.B. Resnick (ed.), *Knowing, learning and instruction. Essays in honor of Robert Glaser*. Hillsdale, NJ.: Erlbaum.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: theoretical and methodological issues. *The Journal of the Learning Sciences, 13*(1), 15-42.

Design Based Research Collective (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32*(1), 5–8.

De Vries, B., van der Meij, H., & Lazonder, A. W. (2008). Supporting reflective web searching in elementary schools. *Computers in Human Behavior*, *24*, 649-665

Dirkx, A., Theuns, A., & Timmers, C. (2006). In drie uur bewust onbekwaam! Informatievaardigheden [In three hours aware of incompetence! Information skills]. *Informatieprofessional, November*, 16-19.

Dochy, F. J. R. C. (1993). De invloed van voorkennis op leerresultaten en het leerproces [The effect of prior knowledge on learning outcomes and process]. In: W. Tomic & P. Span (eds.), *Onderwijspsychologie: Beïnvloeding, verloop en resultaten van leerprocessen* (pp. 97-119). Utrecht, The Netherlands, Lemma.

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*. Alphen a/d Rijn: Samsom.

Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information problem-solving: The big six skills approach to library and information skills instruction*. Norwood, NJ: Ablex.

Eisenberg, M. B., & Berkowitz, R. E. (1992). Information problem-solving: the big skills approach. *School Library Media Activities Monthly*, *8*(5), 27-29, 37, 42.

Eskola, E.-L. (2005). Information literacy of medical students studying in the problem-based and traditional curriculum. *Information Research*, *10*(2), paper 221. Available at http://informationr.net/ir/10-2/paper221.html (October 2005).

Feddes, R., Vermetten, Y., Brand-Gruwel, S., & Wopereis, I. (2003). Strategische kennis over het oplossen van informatieproblemen: een exploratief onderzoek [Strategic knowledge on information problem-solving: an explorative study]. *Pedagogische Studiën*, *80*, 210-225.

Fidel, R., Davies, R. K., Douglass, M. H., Holder, J. K., Hopkins, C. J., Kushner, E. J., Miyagishima, B. K., & Toney, C. D. (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.

Garvey Berger, J., & Baker, R. (2008, January). *Developing new knowledge and practice through teacher-researcher partnerships?* Paper presented at International Congress for School Effectiveness and Improvement (ICSEI), Auckland, CA.

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

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational Psychologist, 38*, 33-41.

Gick, M., & Holyoak, K. (1983). Schema-induction and analogical transfer. *Cognitive Psychology, 12*, 1-38.

Goldstein, F. C., & Levin H. S. (1987). Disorders of reasoning and problem-solving ability. In: M. Meier, A. Benton, & L. Diller (Eds.), *Neuropsychological rehabilitation*. London: Taylor & Francis Group.

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.

Hill, J. R. (1999). A conceptual framework for understanding information seeking in open-ended information services. *Educational Technology, Research and Development*, *47*(1), 5-27.

Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the World Wide Web. *Educational Technology, Research and Development*, *45*(4), 37-64.

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.

Hoffman, J. L., Wu, H.-K., Krajcik, J. S., & Soloway, E. (2003). The nature of learners' science content understandings with the use of on-line resources. *Journal of Research in Science Teaching*, *40*, 323-346.

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.

Kirk, E. E. (1996). *Evaluating information found in the Internet*. Retrieved May, 10, 2006 from Johns Hopkins University, the Sheridan Libraries Web site http://www.library.jhu.edu/researchhelp/general/evaluating/index.html

Klausmeijer, H.J. (1990). Conceptualizing. In: Jones, B.F., & Idol, L. (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 93 - 138). Hillsdale NJ: Erlbaum.

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.

Land, S. M., & Greene, B. A. (2000). Project-based learning with the World Wide Web: A qualitative study of resource integration. *Educational Technology, Research and Development, 48*(1), 45-68.

Large, A., & Beheshti, J. (2000). The web as a classroom resource: Reaction from the users. *Journal of the American Society of Information Science, 51*, 1069-1080.

Larkin, J. E., & Pines, H. A. (2004). Developing information literacy and research skills in introductory psychology: A case study. *The Journal of Academic Librarianship*, *31*(1), 40-45.

Lazonder, A. W. (2001). Minimalist instruction for learning to search the World Wide Web. *Education and Information Technologies*, *6*, 161 – 176.

Lazonder, A. W. (2003). Principles for designing web searching instruction. *Education and Information Technologies, 8*, 179-193.

Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on Web search behavior and search outcomes. *British Journal of Educational Technology, 36*, 465-475.

Lorenzen, M. (2002). The land of confusion? – High school students and their use of the World Wide Web for research. *Research strategies, 18*(2), 151-163.

Lundeberg, M. A. (1987). Metacognitive aspects of reading comprehension: Studying understanding in legal case analysis. *Reading Research Quarterly, 22*, 407-432.

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.* Paper presented at the meeting of the National Association for Research in Science Teaching, Chicago, IL.

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.

Marchionini, G. (1995). *Information-seeking in electronic environments.* New York: Cambridge University.

Marchionini, G. (1999). Educating responsible citizens in the information society. *Educational Technology, 39*(2), 17-26.

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.

Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. *Contemporary Educational Psychology, 29*, 103-128.

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.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analyses: An expanded sourcebook (2nd ed.)*. Thousand Oaks, CA: Sage.

Monereo, C., Fuentes, M., & Sànchez, S. (2000). Internet search and navigation strategies used by experts and beginners. *Interactive Educational Multimedia, 1*, 24-34.

Moore, P. (1995). Information problem solving: a wider view of library skills. *Contemporary Educational Psychology, 20*, 1-31.

Moore, P. (1997). *Teaching information problem solving in primary schools: an information literacy survey*. Paper presented at the 63rd IFLE general conference, Copenhagen, Denmark.

Mosenthal, P. B. (1998). Defining prose task characteristics for use in computer-adaptive testing and instruction. *American Educational Research Journal, 35*, 269-307

Novak, J. D. (1990). Concept mapping: a useful tool for science education. *Journal of Research in Science Teaching, 27*, 937–949.

Ormondroyd, J. (2004). *Critically analysing information sources*. Retrieved May 10, 2006, from Cornell University Library, Reference Department, Instruction, Research, and Information Services (IRIS), Ithaca, NY, USA Web site http://www.library.cornell.edu/olinuris/ref/research/skill26.htm

Paul, R. C. (1992). *Critical thinking: What every person needs to survive in a rapidly changing world*. (2nd revised ed.). Santa Rosa, CA: Foundation for Critical Thinking.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher, 18*, 16-25.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*, 801–813.

Pritchard, A. & Cartwright, V. (2004). Transforming what they read: helping eleven-year-olds engage with Internet information. *Literacy, 38*(1), 26-31

Reeves, T. (2006). Design research from a technology perspective. In: J. van den Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.), *Educational Design Research*. Londen: Routtledge.

Rosell-Aguilar, F. (2004). Well done and well liked: online information literacy skills and learner impressions of the web as a resource for foreign language learning. *ReCALL, 16*(1), 210-224.

Rothenberg, D. (1998). How the Web Destroys Student Research Papers. *Education Digest, 63*, 59-61.

Rouet, J. -F., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88*, 478-493.

Rouet, J. -F., & Eme, E. (2002). The role of metatextual knowledge in text comprehension. In: P. Chambres & M. Izaute (Eds.), *Metacognition: Process, function and use* (pp. 121-133). Kluwer Academic Publishers.

Rouet, J.-F., & Levonen, J. J. (1996). Studying and learning with hypertext: Empirical studies and their implications. In: Rouet, J.F., Levonen, J.J., Dillon, A. & Rand, J. (eds.), *Hypertext and Cognition* (pp 9-23). Hillsdale, NJ.: Erlbaum.

Rouet, J. -F., Levonen, J. J., Dillon, A., & Spiro, R .J. (1996). An introduction to hypertext and cognition. In: J. F. Rouet, J. J Levonen, A. Dillon, & R. J. Spiro (Eds.), *Hypertext and cognition*. New Jersey, Lawrence Erlbaum Associates.

Saljö, R. (draft). *A sociocultural perspective on learning and education: Material and intellectual culture and the formation of minds.*

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist, 24*, 113-142.

Schacter, J., Chung, G. K. W. K., & Dorr, A. (1998). Children's internet searching on complex problems: performance and process analyses. *Journal of the American Society for Information Science, 49*, 840-849.

Schmeck, R. R., & Geisler-Brenstein, E. (1989). Individual differences that affect the way students approach learning. *Learning and Individual Differences, 1*, 85-124.

Shapiro, J. J., & Hughes, S. K. (1996). Information literacy as a liberal art: enlightenment proposals for a new curriculum. *Educom Review, 31*(2). Retrieved October, 31, 2005. Available from www.educause.edu/pub/er/review/reviewarticles/31231.html.

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*. (pp 1 – 20). Dordrecht: Kluwer.

Spitzer, K. L. (2000). What every educator should know about information literacy. In M. B. Eisenberg & R. E. Berkowitz (Eds.), *The Big6 collection: The best of the Big6 newsletter* (pp. 3-13). Worthington, OH: Linworth.

Stadtler, M. (2006). *Auf der Suche nach medizinischen Fachinformationen. Metakognitionen bei der Internetrecherche von Laien*. Münster: Waxmann.

Stadtler, M., & Bromme, R. (2004). Laypersons searching for medical information on the web: The role of metacognition. In: K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society,* (p. 1638), Mahwah, NJ: Erlbaum.

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the web search of laypersons. *Computers in Human Behavior, 24*, 716-737.

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

Ten Dam, G., & Volman, M. (2004). Critical thinking as a citizenship competence: Teaching strategies. *Learning and Instruction, 14,* 359-379.

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.

Todd, R. J. (1995). Integrated information skills instruction: Does it make a difference? *SLMQ, 23*(2).

Tutzauer, F. (2003). On the sensible application of familywise alpha adjustment. *Human Communication Research, 29*, 455-463.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction, 16*, 154-164.

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.

Van Merriënboer, J. J. G (1999). *Cognition and Multimedia Design*. Inaugural address. Open University of the Netherlands.

Wallace, M. C., Shorten, A., & Crookes, P. A. (2000). Teaching information literacy skills: An evaluation. *Nurse Education Today, 20*, 485-489.

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. (2008a, July). *A Field Test of Two Conflicting Transfer Theories; Metacognition versus Cognition*. Paper presented at 29th International Congress of Psychology (ICP). Berlin, Germany.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2008b). 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. (in press). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education*.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (submitted). *Teachers and the World Wide Web: How teachers evaluate search results, information and source.*

Wertheimer, M. (1961). *Productive Thinking*. London: Tavistock.

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*, 73-87.

Woods, N. N. (2007). Science is fundamental: the role of biomedical knowledge in clinical reasoning. *Medical Education, 41*, 1173–1177

Wopereis, I, Brand-Gruwel, S. & Vermetten, Y. (2005, August). *The effect of embedded instruction on solving information problems.* Paper presented at the 11th European Conference for Research on Learning and Instruction (EARLI), Nicosia, Cyprus.

Wopereis, I., Brand-Gruwel, S., & Vermetten, Y. (2008). The effect of embedded instruction on solving information problems. *Computers in Human Behavior, 24*, 738-752.

Young, K. A. (2005). Direct from the source: the value of 'think-aloud' data in understanding learning. *Journal of Education Enquiry,* 6(1), 19-33.

Summary

Teenagers nowadays spend more time online than they watch television. The World Wide Web (WWW) is their playground. They use it to communicate with friends, watch movies, download music and play games. Since they use the Web so frequently in everyday life, it is not surprisingly that students rely on the Web for educational tasks, such as writing essays and preparing for presentations, as well. Using the Web for educational purposes requires that they identify their information needs, locate information sources, extract and organize information from each source, and synthesize information from a variety of sources. This set of activities is frequently defined as information-problem solving (IPS) (Brand-Gruwel, Wopereis & Vermetten, 2005).

In the process of solving information problems the student's ability to evaluate the hit list and the information on websites including the website itself plays a critical role. However, the way students evaluate is far from ideal, because it is not always based on clear criteria but on intuition (Koot & Hoveijn, 2005). A non-critical attitude towards information on the WWW can seduce students to cut and paste information for accomplishing assignments without evaluating it, which results in reports and learning that lack quality (Britt & Aglinskas, 2002).

The importance of instruction in these skills and in the critical use of the WWW has been recognized by teachers. This instruction should not only address the evaluation skills to evaluate search results, information and source, but should also focus on the adaptation of these skills to new situations and changed tasks (i.e., transfer), because failing this last quality will sooner or later lead to skills obsolescence. Different theories could be used to design instruction that fosters transfer of evaluation skills. One of these theories is the high road theory of Salomon and Perkins (1989), which 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 flexibly in different situations. Another transfer theory is the rich representation theory of Simons, van der Linden and Duffy (2000), which emphasizes the importance of a good, extensive and well organised knowledge base and the domain specific interpretation of the skills.

The aims of this research are gaining insight in the evaluation skills used by students and teachers and to design, implement and evaluate instruction to foster the transfer of these skills to various settings. Research questions were: 1) which evaluation criteria do students and teachers of secondary education use for evaluating search results, information and source and do the criteria used differ when solving tasks in different domains, 2) can instruction designed according to the high road transfer theory and the rich representation theory foster the transfer of students' evaluation skills?

These questions were addressed in five studies. One review study, two experimental studies to determine the criteria students and teachers use to evaluate results, information and source on the WWW in different domains, and two experimental studies to examine the (transfer) effect of designed instruction were conducted.

Chapter 2 describes the review study. The goal of this study was to determine what kinds of problems children, teenagers and adults experience when solving information problems on the WWW, and what kind of instructional support can help to solve these problems. Fifteen articles concerning problems children, teenagers and adults experience when searching for information on the WWW were found after a systematic search in PsycINFO and ERIC, and using references of the articles found. The articles were analyzed using the decomposition of the information-problem solving skill developed by Brand-Gruwel, Wopereis and Vermetten (2005) as a framework. Furthermore, twelve studies regarding instructional support to foster students' information-problem solving skills were found. They are described based on age of the participants involved and on use and effectiveness of instructional measures. Results show that people in every age group experience some problems with information-problem solving. Some skills are mastered during the process of growing up, other skills remain problematic throughout life. Children, teenagers and adults have problems with specifying search terms, evaluating search results, evaluating source and information and regulating their search process. Children and teenagers also have trouble with the skills define information problem and process information. Instruction in information-problem solving should therefore take into account the age of the target group and adjust the instruction accordingly.

The studies reviewed concerning instruction and support to foster students' information- problem solving ability can be grouped based on several features: the way the instruction is offered (either embedded in the curriculum or as a separate course); the way the instruction is followed by participants (individually or collaboratively), tools used during the instruction, and the skills addressed in the instruction. The review showed that there are only a few empirically tested instructional or support methods for IPS. Most of the methods found were stand-alone courses for individual use. Tools used in these methods differ from a web-based portal or a computer application, to worked-out examples and visualizations, to work sheets, to paper material only. It is promising that all methods aimed at (some of) the problematic skills of their target group and that most of them were effective. However, the effectiveness of the methods has not been established without doubt. The first, and perhaps one of the biggest, question marks that can be placed by the instructional methods was the fact that only one of them tested for transfer. It remains unclear whether or not instruction in information-problem solving IPS should best be given embedded or stand-alone and whether it should be given collaboratively or individually. The review suggests that IPS-instruction should strive to encourage students to actively engage in the process and not only focus on the 'mechanical' aspects. This also implies that the whole process should be taken into account. In sum, research about instructional support provides ideas and guidelines for designing instruction or support in IPS, like working with whole tasks and a focus on the whole process. However, further research should aim at the mentioned issues and should especially include transfer of the IPS skill.

Chapter 3 presents the results of the first empirical study, aiming at determining the evaluation criteria 9th grade students in secondary education use to accomplish assignments. Twenty-three students solved two tasks from different domains while thinking aloud. They were asked to answer the question posed in the task by selecting

information from the WWW and pasting this into a word document. The protocols were transcribed verbatim and using a coding scheme based on a scheme used by Brand-Gruwel, Wopereis and Vermetten (2005) evaluation criteria were scored. Task performance was measured by three criteria: answer contains information of more than one source, quality of the sources and suitability of information used. A day after the students individually completed the tasks, they returned in groups of three. These focus groups were interviewed to obtain students' knowledge and conceptions on criteria for evaluating information and source.

Results showed that while solving information problems students spent most of their time on searching and scanning and only a small amount of time on processing and organizing information. During the IPS process, students did not evaluate results, information and sources very often. When students evaluated the results of a hit list, they mostly used the title and the given summary. When evaluating information the criterion used most often was whether the information is connected to the task. No differences were found between the different domains. The criteria mentioned by the students, when asked which are important for evaluating information, are not always the same criteria they used while solving the information problems. Students mentioned more criteria, but also admitted that they did not always use them while searching the Web. The reasons were time pressure, motivation and convenience. Furthermore, the students did not mention differences between the domains. All criteria are applicable in the different domains, although type of task can make criteria less or more important. Students' task performance was far from optimal, however no relation between evaluation behaviour and task performance was found.

The goal of *Chapter 4* was to uncover criteria teachers use and the possible influence of domain knowledge on their evaluation skills. Eleven teachers solved two information problems while thinking aloud, one in the domain they taught, and one in a different domain. This means that they were domain experts while accomplishing one assignment and non-experts while solving the other assignment. All think-aloud protocols were transcribed verbatim. Again, the criteria used to evaluate results, information and source were coded. Furthermore, to gain more insight in the use of prior knowledge when evaluating information, sites or hit lists, utterances showing use of domain knowledge were categorized based on the goal of the utterance: 1) activate prior knowledge, 2) evaluate results, information or source, or 3) to make a decision upon the search strategy. The performance was measured in the same way as in the previous study. A day after teachers individually completed the tasks, they returned in groups of three. These focus groups revealed teachers' knowledge and conceptions on criteria for evaluating information and its source in the two different domains. Results shed light on how teachers evaluate and how they think domain knowledge influences the search process. Teachers in this study evaluated search results by the title and summary they found in the hit list. Information and source were evaluated by checking whether the information or source can give an answer to the information problem at hand. Teachers did not pay attention to the author of the information or references in the information. They did not always evaluate results, information and source every time they viewed search results or websites.

There was no difference in teachers' use of criteria when solving tasks within or outside their domain of expertise. Domain knowledge did not seem to influence the use of criteria. A qualitative analysis of the results revealed that not all teachers explicitly used domain knowledge. More domain knowledge was uttered while solving a problem inside their domain of expertise. The goal of utterances did not differ between tasks in or outside the domain of expertise. Domain knowledge did not influence the use of criteria, and it did not influence task performance. Teachers did not use better sources or found more suitable information on a task within their own domain than on a task outside their own domain.

Teachers were aware of the criteria they used and that can be used. During the group interviews they mentioned more criteria than they used. They also admitted they did not use all the criteria they knew because they wanted to come up with an answer for the task. Teachers also stated that the use of criteria depends on the domain and domain knowledge.

In *Chapter 5* a study is reported in which two instructional programs were designed to foster 9th grade students' evaluation skills, as well as the transfer of these skills to another domain. The first program was designed using the transfer theory of Salomon and Perkins (high road) as a starting point; for designing the second program the transfer theory of Simons, van der Linden and Duffy (rich representation) was used. Both programs were developed together with four secondary education history teachers, using a design-based research approach. The same teachers also implemented the programs. Two 9th grade classes received the high road program; two other classes participated in the rich representation program. Three lessons were observed in every class. Effects of the programs on students' use of criteria for evaluating search results, information and source were measured with two paper and pencil tests, consisting of a hit list and a website evaluation task. Students received a manufactured hit list on paper and had to select three sites they would open and three sites they would not open. Students also received a booklet of eight printed out websites and were asked which sites and what information they would or would not use. These paper and pencil tests came close to reality while enabling us to test a large group of students at the same time. Furthermore, a small group of students thought aloud while accomplishing two tasks; one task in the domain of instruction (history) and one in the transfer domain (biology). Students in both programs benefited from the lessons in terms of improved evaluation behaviour regarding the evaluation of websites and information in the domain of instruction. Although students in both conditions improved, 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. The think aloud protocols showed that students evaluated hit lists by the title and summary and that they evaluated information and source by the connection to the task. Results concerning task performance of these students did not reveal a large improvement after the intervention.

The findings regarding the evaluation of websites and information 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. 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.

A success factor in the rich representation program 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. The success factor of the high road program 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.

In *Chapter* 6 a study is described in which the best of both programs, as studied in the previous experiment, is combined in the design of a new program. The same four history teachers co designed and implemented this new program. The goal of this program was to teach students how to evaluate results, information and source and fostering transfer of these evaluation skills to other domains. This new program was tested with four experimental classes and results were compared to a control class. Effects of the program on students' use of criteria while evaluating search results, information and sources were again measured using two paper and pencil tests (history and biology). Furthermore, because the instruction of the evaluation skills was embedded in history lessons, also students' knowledge on the subject matter (Second World War) involved was measured after the intervention. Furthermore the teachers' experiences with design-based research (DBR) were investigated by comparing the reflective stories written by the teachers. We can conclude that instruction improved students' evaluation of information and source (websites) in the domain of instruction. Students' scores of the experimental condition increased from pre- to post-test, while the scores of the control group decreased from pre- to post-test. Instruction seems to have had no effect on students' evaluation of search results (hit list); students in the experimental condition maintained their scores from pre- to post-test and the scores of the students in the control condition decreased from pre -to post-test. The program did not cause a transfer effect with regard to the use of criteria for the evaluation of websites. There was no difference in the scores on the post-test between students in the experimental classes and the control class. It was expected that the experimental classes would score higher on the post-test, if the program had led to transfer. Instruction did not have a transfer effect on students' use of criteria for evaluating search results as well. Students in the experimental condition maintained their scores from pre- to post-test and the scores of the students in the control condition decreased from pre- to post-test. The most likely explanation for not achieving a transfer effect is that the new program was not implemented to the full extent in all experimental classes. This is confirmed by the observations and the reflecting reports of the teachers concerning the design cycles and the implementation of the program.

The score on the final test of the experimental classes was significantly higher than the score of the control class. Teachers preferred the first design cycle, where they had more sessions together and developed all the lessons they would implement in their class themselves, above the second cycle. The second cycle was found to be too short and provided not enough information on the transfer theories. Therefore, teachers felt that the programs of the first design cycle were executed better than the program of the second cycle. Teachers also stated that transfer cannot be fostered by this program alone; teachers from all courses in school should integrate evaluation tasks in their lessons.

Main conclusions and findings are presented in *Chapter 7*, as well as some practical implications and suggestions for future research. It can be concluded that stimulating students' use of evaluation skills and increasing their knowledge on these skills is important. While it is important for students to recognize and use categories like usability and reliability, they should also realize that the use of criteria can differ between domains. Knowing the categories and criteria is not enough, it is also important to have more sophisticated knowledge on criteria.

A program designed to foster 9th grade students' evaluation skills based solely on the high road theory, a program based solely on the rich representation theory and a program based on success factors of the theories improved students' evaluations of information and sources (websites) in the domain of instruction. However, none of the programs improved the evaluation of search results (hit list). Furthermore, the programs that were based on one transfer theory (i.e. the high road theory or the rich representation theory) both led to transfer. Students improved their evaluation of information and source in the transfer domain. Evaluation of search results again did not improve. Unfortunately, a combination of the strong points of both transfer theories did not lead to transfer. This unexpected result could be due to the fact that the program was not always implemented as planned.

Practical implications that can be derived from the studies in the thesis are that embedding the instruction in the curriculum and providing students with whole tasks, emphasizing domain content as well as evaluation skills, improves students domain knowledge as well as their evaluation skills. Implementing a program, with an emphasis on searching for information, and whole tasks, requires adaption from teachers as well as students and this adaptation takes time. Finally, for schools facing the implementation of a new curriculum DBR could be an effective way to involve teachers more with the implementation and thus increase support of the implementation.

Future research should aim at determining the long term effects of the two transfer theories, and should address whether the combination of both theories actually improves transfer or whether the combination inhibits the positive effects of both programs. Furthermore it would be wise to discuss with teachers of more courses how they could integrate WWW assignment in their lessons throughout the entire curriculum. Finally, it should be tested what motivates students to use their evaluation skills.

Samenvatting

Tegenwoordig surfen tieners vaker op Internet dan dat ze televisie kijken. Het World Wide Web (WWW) is hun speeltuin. Ze gebruiken het om met vrienden te communiceren, films te kijken, muziek te downloaden en te gamen. Omdat leerlingen het Web zoveel gebruiken in hun dagelijks leven, is het niet verwonderlijk dat ze ook op het Web vertrouwen voor schoolse taken, zoals het schrijven van een werkstuk en het voorbereiden van een presentatie. Het gebruiken van het Web voor onderwijsdoeleinden vraagt van leerlingen dat ze hun informatiebehoefte onderkennen, bronnen lokaliseren, informatie selecteren en organiseren uit die bronnen, en informatie uit verschillende bronnen samenvoegen tot een product, zoals bijvoorbeeld een werkstuk. Deze verschillende activiteiten samen wordt het oplossen van informatieproblemen genoemd (Brand-Gruwel, Wopereis, & Vermetten, 2005).

De vaardigheid van de student om de informatie die zij vinden te evalueren, speelt een cruciale rol in het hele oplossingsproces. Dit geldt zowel voor de hit lists, de informatie op websites, en de websites zelf. De manier waarop leerlingen evalueren is echter verre van ideaal; dit gebeurt lang niet altijd op basis van duidelijke criteria maar op basis van intuïtie (Koot & Hoveijn, 2005). Een niet-kritische houding ten opzichte van informatie op het WWW kan leerlingen verleiden tot het knippen en plakken van informatie zonder de informatie te evalueren, wat kan leiden tot werkstukken en kennisconstructie van mindere kwaliteit (Britt & Aglinskas, 2002).

Het belang van instructie in evaluatievaardigheden en het kritisch gebruik van het WWW wordt onderkend door docenten. De instructie zou zich niet alleen moeten richten op vaardigheden om zoekresultaten, informatie en bronnen te evalueren, maar zou ook aandacht moeten besteden aan de aanpassing van deze vaardigheden aan nieuwe situaties en veranderde taken (d.w.z. transfer). Wanneer dit laatste niet gebeurt, zal vroeg of laat de vaardigheid in onbruik raken. Er kunnen verschillende theorieën gebruikt worden om instructie te ontwerpen die de transfer van evaluatievaardigheden bevordert. Een van deze theorieën is de 'high road' theorie van Salomon en Perkins (1989), die stelt dat leerlingen gestimuleerd moeten worden om expliciet aandacht te besteden aan de verschillende stappen die genomen moeten worden in het oplossingsproces en de manier waarop deze stappen flexibel kunnen worden gebruikt in verschillende situaties. Een andere transfertheorie is de 'rich representation' theorie van Simons, van der Linden, en Duffy (2000), die het belang van een goede, uitgebreide en goed georganiseerde kennisbasis en de domein specifieke interpretatie van de vaardigheden benadrukt.

Het doel van het onderzoek gepresenteerd in dit proefschrift was tweeledig. Enerzijds inzicht verkrijgen in de evaluatievaardigheden van leerlingen en docenten, en anderzijds het ontwerpen, implementeren en evalueren van instructie voor het bevorderen van deze vaardigheden binnen verschillende domeinen. De onderzoeksvragen waren: 1) welke evaluatiecriteria gebruiken leerlingen en docenten van het voortgezet onderwijs voor het evalueren van zoekresultaten, informatie en

bronnen en verschillen deze criteria tussen verschillende domeinen, 2) kan instuctie, ontworpen volgens de 'high road' transfertheorie en de 'rich representation' theorie, de transfer van evaluatievaardigheden door leerlingen bevorderen?

Deze vragen werden in vijf studies beantwoord: een review studie, twee experimentele studies om te bepalen welke criteria leerlingen en docenten gebruiken om zoekresultaten, informatie en bron te evalueren in verschillende domeinen, en twee experimentele studies om het (transfer)effect van ontworpen instructie te bepalen.

Hoofdstuk 2 beschrijft de review studie. Het doel van deze studie was te bepalen welke problemen kinderen, tieners en volwassen ondervinden wanneer ze informatieproblemen oplossen en informatie zoeken op het WWW, en hoe instructie deze problemen kan helpen oplossen. Een systematische zoektocht in PsycINFO en ERIC, en het gebruik van referenties van gevonden artikelen, leverde vijftien artikelen op over problemen die kinderen, tieners en volwassenen ondervinden tijdens het zoeken van informatie op het WWW. De decompositie van de vaardigheid 'oplossen van informatieproblemen' van Brand-Gruwel, Wopereis, en Vermetten (2005) is gebruikt als raamwerk voor de analyses. Daarnaast werden er twaalf studies gevonden over instructie om deze vaardigheid van leerlingen te bevorderen. Deze studies zijn eerst verdeeld op basis van de leeftijd van de doelgroep waarvoor de instructie bedoeld was. Per doelgroep werden de verschillende instructiemethodes geanalyseerd op basis van soort intructie, gebruike materialen en effect. De resultaten laten zien dat personen in elke leeftijdscategorie problemen ondervinden met het oplossen van informatieproblemen. Sommige problemen lossen zich op wanneer men ouder wordt en deze vaardigheden vanzelf leert, andere vaardigheden blijven ook voor volwassenen problematisch. Kinderen, tieners en volwassenen hebben problemen met het specificeren van een zoekterm, met het evalueren van zoekresultaten, informatie op websites en bronnen, en met het reguleren van hun zoekproces. Daarnaast hebben kinderen en tieners problemen met het definieren van het probleem en met het verwerken van informatie. Instructie in het oplossen van informatieproblemen zou daarom moeten worden afgestemd op de doelgroep.

De twaalf studies over instructie in het oplossen van informatieproblemen kunnen gegroepeerd worden op basis van verschillende aspecten: de manier waarop instructie wordt aangeboden (ingebed in het curriculum of als een aparte cursus), de wijze waarop deelnemers de instructie doorlopen (individueel of samen), de materialen die gebruikt worden tijdens de instructie, en de vaardigheden die centraal staan tijdens de instructie. De review laat zien dat er maar weinig empirisch geteste instructiemethodes in het oplossen van informatie problemen zijn. De meeste gevonden methodes waren op zichzelf staande cursussen voor individueel gebruik. Gebruikte materialen verschillen van een webgebaseerde portal of een computerapplicatie tot uitgewerkte voorbeelden en visualisaties, tot werkbladen en uitsluitend papieren instructie. Het is veelbelovend dat al deze methodes zich veelal richten op problematische vaardigheden, maar de effectiviteit van de methodes kon niet met zekerheid worden vastgesteld. Zo heeft slechts één van de studies gekeken naar transfer van de vaardigheid naar een andere omgeving (van het WWW naar een electronische database). In geen enkele studie werd onderzocht of het effect van de instructie ook in andere domeinen of na langere tijd nog zichtbaar was. Daarnaast blijft

het onduidelijk of de instructie het beste geïntegreerd of zelfstandig moet worden gegeven en of het individueel of samen met anderen moet worden gevolgd. Uit de review blijkt dat instructie in informatievaardigheden ernaar moet streven leerlingen actief te betrekken in het gehele proces en niet alleen aandacht moet besteden aan de mechanische aspecten, als het intypen van trefwoorden en op de juiste knoppen drukken. Dit impliceert ook dat het hele proces van informatieproblemen oplossen aandacht moet krijgen. Kortgezegd geeft onderzoek naar instuctie richtlijnen voor het ontwerpen van instructie in het oplossen van informatieproblemen en het gebruik van informatievaardigheden, zoals werken met hele taken en aandacht voor het hele proces. Er is echter meer onderzoek nodig naar de genoemde onduidelijkheden, en voornamelijk naar de transfer van informatievaardigheden

Hoofdstuk 3 presenteert de resultaten van de eerste empirische studie, gericht op het vaststellen van de evaluatiecriteria die leerlingen uit de derde klas van het VWO gebruiken bij het selecteren van informatie gevonden op Internet voor het maken van opdrachten. Drieëntwintig leerlingen kregen twee taken van een half uur voorgelegd binnen verschillende domeinen. Deze taken, die bestonden uit een informatieprobleem, werden hardopdenkend gemaakt. De hardopdenkprotocollen werden uitgetypt en de gebruikte evaluatiecriteria werden gescoord met behulp van een codeerschema gebaseerd op een schema van Brand-Gruwel, Wopereis, en Vermetten (2005). De score op de taak werd vastgesteld op basis van drie criteria: het antwoord bevat informatie uit meer dan één bron, de kwaliteit van de bron, en de geschiktheid van de informatie. Een dag nadat de leerlingen de taken individueel hadden gemaakt, kwamen ze in groepen van drie terug. Deze focusgroepen werden geinterviewd om de kennis en concepties van de leerlingen over criteria voor het evalueren van informatie en bronnen te achterhalen.

De resultaten laten zien dat leerlingen zoekresultaten, informatie en bronnen niet vaak evalueerden tijdens het zoekproces. Wanneer leerlingen wel de zoekresultaten evalueerden, deden ze dit voornamelijk op basis van de titel en de samenvatting. Bij het evalueren van informatie op een site was het meest gebruikte criterium of de informatie aansloot op de vraag die ze moesten beantwoorden. Er werden geen verschillen gevonden tussen verschillende domeinen. De criteria die leerlingen noemden wanneer ze gevraagd werden welke criteria belangrijk waren voor het evalueren van informatie, waren niet altijd de criteria die ze gebruikten tijdens het uitvoeren van de taken. Leerlingen noemden meer criteria dan ze gebruikten en gaven toe dat ze de criteria niet hadden gebruikt tijdens het zoeken op het WWW. De redenen hiervoor waren tijdsdruk, motivatie en gemak. Leerlingen noemden ook geen verschillen tussen domeinen. Aangegeven werd dat alle criteria bruikbaar zijn binnen verschillende domeinen, maar ook dat niet alle criteria altijd even belangrijk zijn. Het hangt af van het taaktype. Het resultaat op de taak was verre van optimaal, maar er werd geen relatie gevonden tussen evaluatie gedrag en resultaat.

Het doel van de studie beschreven in *hoofdstuk 4* was het vaststellen van evaluatiecriteria die docenten gebruikten en de mogelijke invloed van domeinkennis op de evaluatievaardigheden van docenten. Elf docenten losten, terwijl ze hardop dachten, twee informatieproblemen op, één binnen hun eigen domein en één binnen een ander domein. De hardopdenk protocollen werden uitgetypt en wederom werden

de gebruikte evaluatiecriteria gecodeerd. Om meer inzicht te krijgen in het gebruik van voorkennis bij het evalueren van websites, informatie of zoekresultaten werden uitingen waarin domeinkennis voorkwam gecategoriseerd op basis van het doel van de uiting: 1) voorkennis activeren, 2) zoekresultaten, informatie of bron evalueren, 3) een beslissing nemen over de zoekstrategie. De score op de taken werd op dezelfde wijze als in de voorgaande studie vastgesteld. Een dag nadat de docenten de taken individueel hadden gemaakt, kwamen ze in groepen van drie terug. Deze focusgroepen hadden tot doel de kennis en concepties van de docenten betreffende criteria voor het evalueren van informatie en bronnen te achterhalen. De resultaten laten zien hoe docenten evalueerden en hoe ze dachten over de invloed van domeinkennis op het zoekproces. Docenten in dit experiment evalueerden zoekresultaten op basis van titel en samenvatting in de hit list. Informatie en bron werden geëvalueerd door te kijken of de informatie of bron een antwoord gaf op de vraag. Docenten besteedden nauwelijks aandacht aan de auteur of de referenties op een site. Ze evalueerden zoekresultaten, informatie en bronnen niet elke keer dat ze deze bekeken.

Docenten lieten geen verschil zien in het gebuik van evaluatiecriteria bij de taak die ze maakten in hun eigen vakgebied en de taak in het andere vakgebied. Domeinkennis leek het gebruik van criteria dus niet te beïnvloeden. Een kwalitatieve analyse van de resultaten liet zien dat niet alle docenten expliciet gebruik maakten van domeinkennis. Er werd wel meer domeinkennis geuit wanneer er een taak werd gemaakt binnen het eigen vakgebied. Het doel van deze uitingen verschilde niet tussen taken binnen of buiten het eigen vakgebied. Domeinkennis had ook geen invloed op de score op de taak. Docenten gebruikten geen betere bronnen, en vonden geen betere informatie bij een taak binnen hun eigen vakgebied dan bij een taak binnen een ander vakgebied.

Uit de focusgroepen bleek dat de docenten zich bewust waren van de criteria die ze gebuikten tijdens het uitvoeren van de taak en de criteria die gebruikt hadden kunnen worden. Opvallend was dat docenten tijdens de focusgroep discussies meer criteria noemden die gebruikt hadden kunnen worden dan dat ze gebruikten tijdens het maken van de taken. Ze gaven zelf ook toe niet alle criteria die ze kenden te hebben gebruikt omdat ze graag snel een antwoord wilden vinden op de taak. Alhoewel er bij de uitvoering van de taken geen verschil te zien was in gebruik van criteria tussen domeinen, gaven de docenten in de focusgroep aan dat het gebruik van criteria afhangt van het domein en domeinkennis.

In *hoofdstuk 5* wordt verslag gedaan van een studie waarin twee lessenseries zijn ontworpen voor het bevorderen van het gebruik van evaluatiecriteria door derdeklassers van het VWO, en van de transfer van het gebruik van deze criteria naar een ander domein. De eerste lessenserie werd ontworpen met de transfertheorie van Salomon en Perkins (high road) als startpunt; voor het ontwerp van de tweede serie werd de theorie van Simons, van der Linden, en Duffy (rich representation) gebruikt. Beide lessenseries werden samen met vier geschiedenisdocenten ontwikkeld, met behulp van een design-based research aanpak. Deze docenten implementeerden de lessenseries ook in hun eigen klassen. Twee derde klassen van het VWO werkten met de 'high road' lessenserie, en twee andere derde klassen van het VWO werkten met de

'rich representation' lessenserie. In elke klas werden drie lessen geobserveerd. Het effect van de lessenseries op het gebruik van criteria voor het evalueren van zoekresultaten, informatie en bronnen door de leerlingen werd gemeten met twee taken, één in het domein van instructie (geschiedenis) en één in het transferdomein (biologie). Beide taken bestonden eruit een hit list en een website te evalueren. Leerlingen ontvingen een gefabriceerde hit list op papier en moesten drie sites selecteren die ze wel en drie sites die ze niet wilden openen. Daarnaast kregen leerlingen ook een boekje met acht uitgeprinte websites en ze werden gevraagd welke sites en welke informatie ze wel of niet zouden gebruiken. Deze papieren taken benaderden de realiteit terwijl we op deze manier een grote groep leerlingen tegelijkertijd konden testen. Daarnaast voerde een kleine groep leerlingen (N = 11) twee taken online uit terwijl ze hardop dachten, een taak in het instructiedomein (geschiedenis) en een taak in het transferdomein (biologie). Leerlingen scoorden bij de geschiedenistaak bij de website-evaluatieopdracht na afloop van de lessen hoger dan voorafgaand aan de lessen. Dit betekent dat leerlingen na de lessen beter in staat waren websites te beoordelen in het domein van instructie. Deze resultaten werden bij beide typen instructie, in gelijke mate, gevonden. De resultaten van de klassen die werkten met de 'rich representation' lessenserie verschilden echter wel van elkaar. De scores op de evaluatieopdracht na afloop van de lessenserie in de ene klas waren hoger dan voor de lessen, terwijl de scores in de andere klas licht daalden.

De hardopdenk protocollen lieten zien dat leerlingen zowel voor als na de lessenserie zoekresultaten evalueerden met behulp van titel en samenvatting en dat ze informatie en bronnen evalueerden op basis van aansluiting op de vraag. De resultaten op de score op de taak lieten geen grote verbetering zien na de instructie.

Als we kijken naar de resultaten op de taken die de leerlingen maakten in het domein biologie, het transferdomein, dan blijkt dat de bevindingen overeenkomen met de resultaten op de taak in het instructiedomein, geschiedenis. Leerlingen in beide instructiecondities verbeterden hun evaluatiegedrag met betrekking tot het evalueren van websites (bronnen) en informatie, ze scoorden na afloop van de lessen hoger op de website-evaluatieopdracht dan voorafgaand aan de lessen. Net als bij de geschiedenistaak, verbeterden de scores van de ene klas in de 'rich representation' conditie zich na afloop van de lessen, de scores van de andere klas bleven nagenoeg gelijk. We kunnen dus concluderen dat beide lessenseries, één gebaseerd op de 'high road' principes van Perkins en Salomon (1989) en één gebaseerd op de 'rich representation' theorie van Simons et al. (2000), leerlingen kunnen stimuleren om kritische websearchers te worden, dat wil zeggen, websites beter te evalueren. Dit geldt voor taken in het instructiedomein en voor taken in het transferdomein. Men zou zelfs mogen zeggen dat de rich representation serie een groter effect had, omdat leerlingen in deze conditie meer verbeterden in hun evaluaties van zoekresultaten ten opzichte van leerlingen in de high road conditie.

Een succesfactor van de 'rich representation' lessenserie waren de groepsdiscussies over evaluatiecriteria. Maar deze groepsdiscussies zouden ook een faalfactor kunnen zijn; als docenten niet in staat zijn deze discussies te structureren, zouden ze wellicht niet zo effectief zijn. Een zwak punt in het ontwerp van dit programma was het gebrek aan kennisconstructie op het gebied van het totale

probleem-oplossingsproces. Dit was juist een succesfactor van de 'high road' lessenserie: de systematische aanpak van het gehele proces. Een faalfactor daar was het feit dat leerlingen de gebruikte processworksheets niet waardeerden en dat een stricte aanpak in het gebuik nodig was. Deze combinatie kan grote problemen opleveren voor een docent.

In *hoofdstuk 6* wordt een studie beschreven waarin het beste van beide lessenseries gecombineerd is in het ontwerp van een nieuwe lessenserie. Dezelfde geschiedenisdocenten ontwierpen en implementeerden deze nieuwe lessenserie. Het doel van de lessenserie was leerlingen te leren hoe ze zoekresultaten, informatie en bronnen kunnen evalueren met behulp van criteria en de transfer te bevorderen van het gebruik van deze criteria naar andere domeinen. Deze nieuwe lessenserie is geëvalueerd met behulp van vier experimentele klassen en de resultaten werden vergeleken met een controleklas. De effecten van de lessenserie op het gebruik van criteria tijdens het evalueren van zoekresultaten, informatie en bronnen, werden opnieuw gemeten door twee taken (geschiedenis en biologie). Omdat de instructie in het gebruik van evaluatiecriteria was ingebed in geschiedenislessen, werd ook de inhoudelijke kennis van de leerlingen over het onderwerp van de lessen (de Tweede Wereldoorlog) na de 15 lessen gemeten. Daarnaast werden de ervaringen van de leraren met design-based research onderzocht door de reflectieve verhalen geschreven door de docenten te analyseren.

We kunnen concluderen dat de instructie het gebruik van criteria voor het evalueren van informatie en bron (websites) van leerlingen verbeterde in het instructiedomein; leerlingen gebruikten na de instructie meer criteria en gebruikten de criteria ook op een correcte manier. De scores op de website evaluatie taak van leerlingen in de experimentele conditie gingen omhoog van pre naar post test, terwijl de scores van de controlegroep verslechterden van pre naar post test. Instructie leek geen effect te hebben op de evaluaties van de zoekresultaten (hit list) van de leerlingen; leerlingen in de experimentele conditie behielden hun score op de hit list evaluatie taak en de scores van leerlingen in de controle conditie verslechterden van pre naar post test. De lessenserie zorgde niet voor een transfereffect op het gebied van het evalueren van websites. Leerlingen in de experimentele conditie gebruikten na de instructie niet signifcant meer criteria dan voor de instructie. Binnen de controlegroep werd eveneens geen verschil gevonden tussen scores op de pre- en posttest. Bovendien was er geen verschil tussen de beide condities; de experimentele conditie scoorde niet hoger dan de controle groep. Wanneer er transfer bereikt zou zijn, dan zouden leerlingen in de experimentele klassen net als bij de geschiedenistaak na afloop van de instructie hoger scoren op de website-evaluatietaak dan voor de instructie, en zou er een verschil zijn in scores tussen leerlingen uit de experimentele conditie en leerlingen uit de controle conditie. Dit was nu niet het geval. De instructie had ook geen transfereffect op het gebruik van criteria voor het evalueren van zoekresultaten door de leerlingen. Leerlingen in de experimentele conditie behielden hun score en de scores van leerlingen in de controle conditie verslechterden van pre naar post test. De meest waarschijnlijke verklaring voor het ontbreken van een transfereffect is dat het nieuwe programma niet volledig is geïmplementeerd in alle experimentele klassen. Dit wordt

bevestigd door observaties en de reflectieve verslagen van docenten over de designcycli en de implementatie van het programma.

De score op het proefwerk van de klassen in de experimentele conditie was significant hoger dan de score van de controle conditie. De docenten hadden een voorkeur voor de eerste ontwerp-cyclus, waarin ze meer gezamenlijke sessies hadden en alle lessen die ze moesten implementeren zelf ontwierpen. De tweede cyclus werd tekort bevonden en verschafte niet genoeg informatie over de transfertheorieën. Daarom hadden de docenten het gevoel dat lessenseries uit de eerste cyclus beter door hen werden uitgevoerd dan de lessenserie uit de tweede cyclus. De docenten gaven ook aan dat transfer niet kan worden bevorderd door deze lessenserie alleen; docenten van alle vakken zouden evaluatietaken moeten integreren in hun lessen.

De belangrijkste conclusies en resultaten worden gepresenteerd in *hoofdstuk 7,* evenals enkele praktische implicaties en suggesties voor verder onderzoek. Er kan geconcludeerd worden dat het stimuleren van de het gebruik van evaluatiecriteria van leerlingen en het vergroten van hun kennis over evaluatiecriteria belangrijk is. Hoewel het belangrijk is dat leerlingen categorieën als bruikbaarheid en betrouwbaarheid herkennen en gebruiken, moeten ze zich ook realiseren dat het gebruik van criteria tussen domeinen kan verschillen. Het kennen van categorieën en criteria is niet voldoende, het is ook belangrijk om meer diepgaande kennis over criteria te hebben.

Een lessenserie, ontworpen om de evaluatievaardigheden van derde klassers te bevorderen, gebaseerd op de 'high road' theorie, een lessenserie gebaseerd op de 'rich representation' theorie, en een lessenserie gebaseerd op de succesfactoren van beide theorieën, verbeterden allemaal de evaluaties van informatie en bronnen van de leerlingen binnen het instructiedomein. Na het doorlopen van een van deze lessenseries gebruikten leerlingen meer criteria voor het evalueren van websites. Geen van de lessenseries verbeterde de evaluaties van zoekresultaten (hit list). Wanneer we kijken naar het bereiken van transfer zien we dat de series gebaseerd op één transfertheorie (d.w.z. de high road of de rich representation theorie) beide transfer bereikten. Na afloop van deze lessenseries gebruikten leerlingen meer criteria voor het evalueren van websites in het transferdomein. De lessenseries bleken niet effectief voor het verbeteren van de evaluaties van zoekresultaten in het transferdomein. Er werd verwacht dat een combinatie van sterke punten van beide transfertheorieën ook tot transfer zou leiden. Helaas bleek dit niet het geval, leerlingen gebruikten na afloop van de gecombineerde lessenserie niet meer criteria om websites te evalueren, en hun evaluaties van zoekresultaten verbeterde ook niet. Dit onverwachte resultaat kan komen door het feit dat die laatste lessenserie niet altijd geïmplementeerd is zoals gepland.

Praktische implicaties die uit de studies in dit proefschrift kunnen worden afgeleid, hebben betrekking op het integreren van instructie in het curriculum en het gebruik van hele taken. Het implementeren van een lessenserie, dat de nadruk legt op zoeken naar informatie en hele taken, vereist aanpassing van zowel docenten als leerlingen. Docenten moeten gebruik maken van een andere didactiek (meer nadruk op leerlingen zelf laten zoeken) en een ander soort opdrachten, en leerlingen moeten leren omgaan met hun nieuwe rol. In plaats van het ontvangen van informatie, moeten ze zelf informatie zoeken. Deze aanpassing kost tijd. Tot slot, voor scholen die staan voor

het implementeren van een nieuw curriculum kan design-based research een goede manier zijn om docenten meer te betrekken bij de implementatie en dus de steun voor de implementatie te vergroten.

Toekomstig onderzoek zou zich moeten richten op het achterhalen van de langetermijneffecten van onderwijs ontworpen volgens de twee transfertheorieën, en of de combinatie van beide theorieën in het ontwerp de transfer bevordert. Daarbij zou het verstandig zijn om met vakdocenten van verschillende domeinen te overleggen over hoe WWW opdrachten zouden kunnen worden geïntegreerd in lessen binnen het hele curriculum. Tot slot zou er onderzocht moeten worden wat leerlingen motiveert om evaluatievaardigheden te gebruiken.

Appendix: Evaluation criteria

Sub skill	Criteria	
Evaluate search results	1. Title/Summary	What is the title given by the search engine and what is the content of the summary?
	2. Kind (site/PDF/)	What kind of source is it, a website, a word document, a PDF file?
	3. Address	What is the address? Is it a .com or .org address?
	4. Rank in hit list	How many results are there in total and what is the rank of the result I am evaluating?
	5. Known to user	Have I used this site before, or have I heard good or bad things about it?
	6. Language	Is the site in a language I prefer and/or understand?
Evaluate information	**A Usability**	
	1. Language	In what language is the information written? Are there many grammatical or type errors? Is it filled with domain specific language?
	2. Connection to task	Does the information answer (part of) the information problem?
	3. Audience	Is the information aimed at a specific group of readers?
	4. Topicality	Is the information up to date?
	5. Amount	Is there enough information on the page? Or only a part of the information I'm looking for?
	B Verifiability	
	1. Author	Who has written the information? Can I contact him/her?
	2. References	Are there references on the page to used sources? Or links to more websites on the same subject?
	3. Information agrees with more sites	Can I find the same information on more pages or is this information only available on this site?
	4. Information agrees with prior knowledge	Does the information confirm what I already know?
	5. Organization	Which organization is behind the information? A governmental organization or a health organization? Can I find their logo on this site?
	C. Reliability	
	1.Kind of information	What kind of information is it? A newspaper article or a forum? Is it an opinion or results from research?
	2. Objectivity	Is the information objective or coloured by a certain point of view? Are there a lot of advertisements on the page?
	3. Primary/Secondary	Is the information first hand or is it someone telling about someone who did something?
	4. Goal	What does the (author of) information want to achieve. Sell something? Convince me of something or just inform me?
Evaluate source	**A Technical**	
	1. Appearance	Does the site appeal to me? Does it have pictures or only text?
	2. Speed	Does it take a lot of time to load the page?

	B Usability	
	1. Language	In what language is the site written? Are there many grammatical or type errors?
	2. Connection to task	Does the site have a connection to (part of) my information problem?
	3. Audience	For whom is the site meant? Who are it's visitors?
	4. Topicality	Is the site updated regularly?
	C Verifiability	
	1. Reputation	Is this site famous or infamous? Does it have a good/bad reputation?
	D. Reliability	
	1. Kind (site/PDF)	What kind of source is it, a website, a word document, a PDF file?
	2. Primary/secondary	Is the site an original source or a site telling about what is written somewhere else?

ICO dissertation series

ico

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the following universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Groningen, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University (and formerly Radboud University Nijmegen and Tilburg University).

134. Van Gog, T. (28-04-2006). *Uncovering the problem-solving process to design effective worked examples.* Heerlen: Open University of the Netherlands.
135. Sins, P.H.M. (18-05-2006). *Students' reasoning during computer-based scientific modeling.* Amsterdam: University of Amsterdam.
136. Mathijsen, I.C.H. (24-05-2006). *Denken en handelen van docenten.* Utrecht: Utrecht University.
137. Akkerman, S.F. (23-06-2006). *Strangers in dialogue: Academic collaboration across organizational boundaries.* Utrecht: Utrecht University.
138. Willemse, T.M. (21-08-2006). *Waardenvol opleiden: Een onderzoek naar de voorbereiding van aanstaande leraren op hun pedagogische opdracht.* Amsterdam: VU University Amsterdam.
139. Kieft, M. (19-09-2006). *The effects of adapting writing instruction to students' writing strategies.* Amsterdam: University of Amsterdam.
140. Vreman-de Olde, G.C. (27-09-2006). *Look experiment design: Learning by designing instruction.* Enschede: University of Twente.
141. Van Amelsvoort, M. (13-10-2006). *A space for debate: How diagrams support collaborative argumentation-based learning.* Utrecht: Utrecht University.
142. Oolbekking-Marchand, H. (9-11-2006). *Teachers' perspectives on self-regulated learning: An exploratory study in secondary and university education.* Leiden: Leiden University.
143. Gulikers, J. (10-11-2006). *Authenticity is in the eye of the beholder: Beliefs and perceptions of authentic assessment and the influence on student learning*. Heerlen: Open University of the Netherlands.
144. Henze, I. (21-11-2006). *Science teachers' knowledge development in the context of educational innovation.* Leiden: Leiden University.
145. Van den Bossche, P. (29-11-2006). *Minds in teams: The influence of social and cognitive factors on team learning.* Maastricht: Maastricht University.
146. Mansvelder-Longayroux, D.D. (06-12-2006). *The learning portfolio as a tool for stimulating reflection by student teachers.* Leiden: Leiden University.
147. Visschers-Pleijers, A.J.S.F. (19-01-2007). *Tutorial group discussion in problem-based learning: Studies on the measurement and nature of learning-oriented student interactions.* Maastricht: Maastricht University.
148. Poortman, C.L. (16-02-2007). *Workplace learning processes in senior secondary vocational education.* Enschede: University of Twente.
149. Schildkamp. K.A. (15-03-2007). *The utilisation of a self-evaluation instrument for primary education.* Enschede: University of Twente.
150. Karbasioun, M. (20-04-2007). *Towards a competency profile for the role of instruction of agricultural extension professionals in Asfahan.* Wageningen: Wageningen University.

151. Van der Sande, R.A.W. (04-06-2007). *Competentiegerichtheid en scheikunde leren: Over metacognitieve opvattingen, leerresultaten en leeractiviteiten.* Eindhoven: Eindhoven University of Technology.
152. Pijls, M. (13-06-2007). *Collaborative mathematical investigations with the computer: Learning materials and teacher help.* Amsterdam: University of Amsterdam.
153. Könings, K. (15-06-2007). *Student perspectives on education: Implications for instructional design.* Heerlen: Open University of the Netherlands.
154. Prangsma, M.E. (20-06-2007). *Multimodal representations in collaborative history learning.* Utrecht: Utrecht University.
155. Niemantsverdriet, S. (26-06-2007). *Learning from international internships: A reconstruction in the medical domain.* Maastricht: Maastricht University.
156. Van der Pol, J. (03-07-2007). *Facilitating online learning conversations: Exploring tool affordances in higher education.* Utrecht: Utrecht University.
157. Korobko, O.B. (07-09-2007). *Comparison of examination grades using item response theory: A case study*. Enschede: University of Twente.
158. Madih-Zadeh, H. (14-09-2007). *Knowledge construction and participation in an asynchronous computer-supported collaborative learning environment in higher education.* Wageningen: Wageningen University.
159. Budé, L.M. (05-10-2007). *On the improvement of students' conceptual understanding in statistics education.* Maastricht: Maastricht University.
160. Meirink, J.A. (15-11-2007). *Individual teacher learning in a context of collaboration in teams.* Leiden: Leiden University.
161. Niessen, T.J.H. (30-11-2007). *Emerging epistemologies: Making sense of teaching practices.* Maastricht: Maastricht University.
162. Wouters, P. (07-12-2007). *How to optimize cognitive load for learning from animated models.* Heerlen: Open University of the Netherlands.
163. Hoekstra, A. (19-12-2007). *Experienced teachers' informal learning in the workplace.* Utrecht: Utrecht University.
164. Munneke-de Vries, E.L. (11-01-2008). *Arguing to learn: Supporting interactive argumentation through computer-supported collaborative learning*. Utrecht: Utrecht University.
165. Nijveldt, M.J. (16-01-2008). *Validity in teacher assessment. An exploration of the judgement processes of assessors*. Leiden: Leiden University.
166. Jonker, H.G. (14-02-2008). *Concrete elaboration during knowledge acquisition*. Amsterdam: VU University Amsterdam.
167. Schuitema, J.A. (14-02-2008). *Talking about values. A dialogue approach to citizenship education as an integral part of history classes*. Amsterdam: University of Amsterdam.
168. Janssen, J.J.H.M. (14-03-2008). *Using visualizations to support collaboration and coordination during computer-supported collaborative learning*. Utrecht: Utrecht University.
169. Honingh, M.E. (17-04-2008). *Beroepsonderwijs tussen publiek en privaat: Een studie naar opvattingen en gedrag van docenten en middenmanagers in bekostigde en niet-bekostigde onderwijsinstellingen in het middelbaar beroepsonderwijs.* Amsterdam: University of Amsterdam.
170. Baartman, L.K.J. (24-04-2008). *Assessing the assessment: Development and use of quality criteria for competence assessment programmes*. Utrecht: Utrecht University.
171. Corbalan Perez, G. (25-04-2008). *Shared control over task selection: Helping students to select their own learning tasks*. Heerlen: Open University of the Netherlands.
172. Hendrikse, H.P. (22-05-2008). *Wiskundig actief: Het ondersteunen van onderzoekend leren in het wiskunde onderwijs*. Enschede: University of Twente.