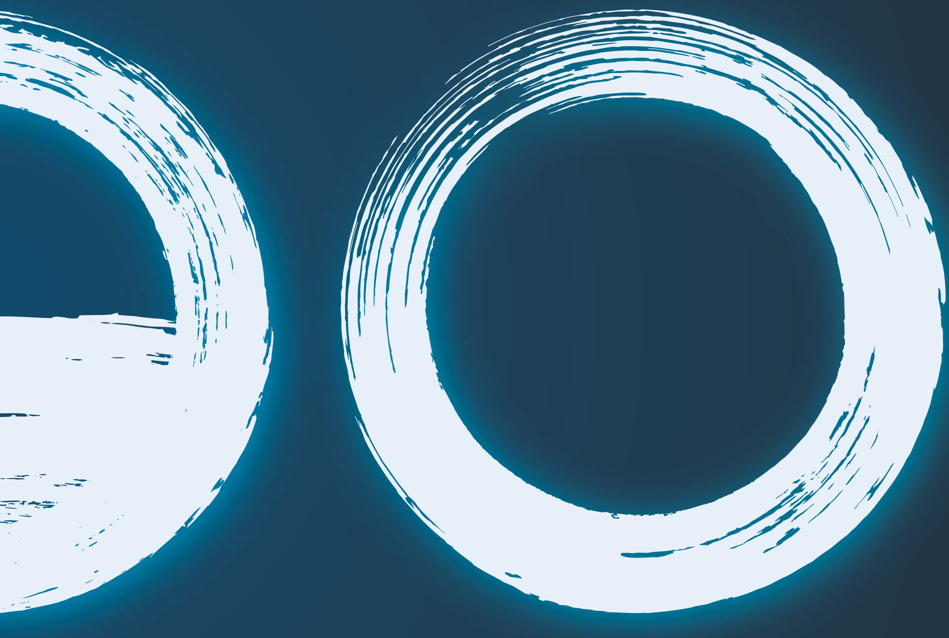
Instruction for information problem solving

Jimmy Frèrejean



The research reported here was carried out at the

Open Universiteit
www.ou.nl

in the context of the research school

ico

Interuniversity Centre for Educational Research

ISBN: 978-94-9273-905-6

Cover illustrations by omelapics & Milano83 - Freepik.com
Printed by Datawyse, Universitaire Pers Maastricht, The Netherlands

INSTRUCTION FOR INFORMATION PROBLEM SOLVING

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. mr. A. Oskamp
ten overstaan van een door het
College voor promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 8 december 2017 te Heerlen
om 13.30 uur precies

door

Jimmy Joseph Franciscus Antonius Frèrejean
geboren op 7 juli 1986 te Maastricht

Promotores
Prof. dr. F.L.J.M. Brand-Gruwel, Open Universiteit
Prof. dr. P.A. Kirschner, Open Universiteit

Co-promotor
Dr. J.L.H. van Strien, Universiteit Utrecht

Overige leden beoordelingscommissie
Prof. dr. J. Elen, Katholieke Universiteit Leuven
Prof. dr. L. Kester, Universiteit Utrecht
Prof. dr. J.M. Voogt, Universiteit van Amsterdam
Prof. dr. H.P.A. Boshuizen, Open Universiteit
Dr. A. Walraven, Radboud Universiteit Nijmegen

TABLE OF CONTENTS

Chapter 1

General introduction

Consider a teacher in a classroom of student teachers, discussing whether reading aloud to young children is an effective method to increase their vocabulary. As the group discusses this issue, they activate their prior knowledge about the problem and consequently find out which information is needed to reach an answer. After some discussion, the teacher helps the group formulate questions and instructs them to find answers by gathering information from online sources during a self-study period. The next time they meet, the students present their findings. This is a prime example of an authentic resource-based learning approach where students are required to find their own learning materials (Van Merriënboer & Kirschner, 2018). The rise of the Internet has provided quick and easy access to a wealth of online information sources that can serve as learning resources, making students less dependent on the library. While this rise of easily accessible resources might seem to benefit teachers and students alike, the abundance of information has significant drawbacks. The increased amount of information sources creates a choice overload and leads to difficulties when trying to separate the useful from the useless.

Whereas libraries have gatekeepers to guard against low-quality information, the Internet does not. Anyone can fill the web with anything ranging from completely correct and reliable information to false or fake information with absolutely no reliability. Apart from that, students also have to deal with distractions, such as online entertainment, online advertising, and social media. The world wide web is a world where numerous actors constantly compete for the attention of the visitor. Commercial companies plaster the web with specifically tailored advertisements, designed to elicit clicks. News agencies prioritize speed of publication over factual correctness, and lure visitors to their advertisement-driven websites with clickbait headlines or fake news posts. Internet companies track users' online behavior to profile them and present information that relates to their viewpoints and interests whilst hiding contradicting viewpoints, effectively isolating them in a filter bubble of agreeable information (Pariser, 2011). And those content producers that work hard to put objective, factually correct information online, such as scientific publishers, lock it away behind a paywall to make money. In the end, the world wide web forms a complex digital arena into which students must venture on their mission to gather information from high-quality, relevant and reliable sources. Too often it is assumed they are experienced gladiators who know how to fend off all the distractions and seductions. In reality, they are unprepared and unaware of who to trust or what to believe (Kirschner & De Bruyckere, 2017; Kirschner & van Merriënboer, 2013; Margaryan, Littlejohn, & Vojt, 2011).

The previous paragraph somewhat theatrically illustrates the problem addressed in this dissertation. Finding information online for educational purposes is becoming more conventional, but constitutes a complex task that requires knowledge, skills, and attitudes to perform correctly. This is often called information literacy or information problem solving (IPS; Brand-Gruwel, Wopereis, & Vermetten, 2005). While its importance as an essential 21st century skill is widely acknowledged, formal instruction to foster these skills is often limited or lacking (Badke, 2010). While some hold pervasive beliefs that modern students are *digital natives* and automatically acquire such IPS skills, research shows otherwise (Kirschner & De Bruyckere, 2017; Kirschner & van Merriënboer, 2013). This lack of deliberate instruction is striking, considering its necessity was recognized more than 20 years ago when students primarily retrieved their information from sources in physical libraries (Eisenberg & Berkowitz, 1990; Kuhlthau, 1988). While this research speaks of *library skills* or *study skills*, it found students encountered the same problems as our students today, and researchers expressed the same need for formalized instruction (Moore, 1995). The essence of the task has not changed significantly. Students still solve a problem by gathering information to formulate an answer, but the context in which this skill is performed has changed drastically. A shift from classroom-based and compartmentalized instruction towards student-focused and holistic instruction (Reigeluth, 1999) increased the adoption of approaches such as resource-based, problem-based or inquiry-based learning. These approaches strongly address students' self-directed learning and self-regulation skills. While working on problems, projects, or tasks, students increasingly need to search for information resources and study those materials to construct the knowledge they need to complete the task, but are confronted with an exponential growth in the diversity and amount of information. Ever since the Internet became publicly available in the nineties, it has been in rapid development, continually changing an already complex environment. As there are no signs this development will stall, up-to-date instruction on navigating the web to locate reliable information should become and remain an essential part of our educational programs, not only for new, young pupils, but also for lifelong learners.

A complex cognitive skill

Decomposing the IPS skill into its constituents, Brand-Gruwel et al. (2005) analyzed which knowledge, skills and attitudes are essential for IPS, followed by a similar analysis for online IPS (Brand-Gruwel, Wopereis, & Walraven, 2009). Based on several information seeking models developed in the past decades, such as the Big Six (Eisenberg & Berkowitz, 1990) and Kuhlthau's information search process (Kuhlthau, 1988), the IPS-I model was developed to describe IPS using the Internet (Brand-Gruwel et al., 2009). This model shows that those seeking information generally iterate between *problem definition*, *searching information*, *selecting information*, *processing information*, and *presenting information*, while *regulating* their process. Each of these skills encompasses several constituent skills. An overview of these skills is displayed in Figure 1.1.

Successful IPS starts with a problem definition, where searchers familiarize themselves with the problem and its domain. They investigate which information they already know and consequently determine which information is still needed in order to produce a satisfactory answer. Ideally, searchers then formulate one or more specific questions to guide their search process. Such a goal-driven approach using focused questions helps searchers stay on topic and recognize when they have gathered sufficient information. After defining the problem, the searcher decides on the best approach to collect the needed information. In most cases, they will be using a search engine such as Google™ to find online resources. The problem description, the formulated questions, and a searcher's background knowledge can be used to generate specific search terms to use in the search engine. It is important that searchers understand how search engines work, to determine which search strategies or combination of search terms are likely to lead to the best results.

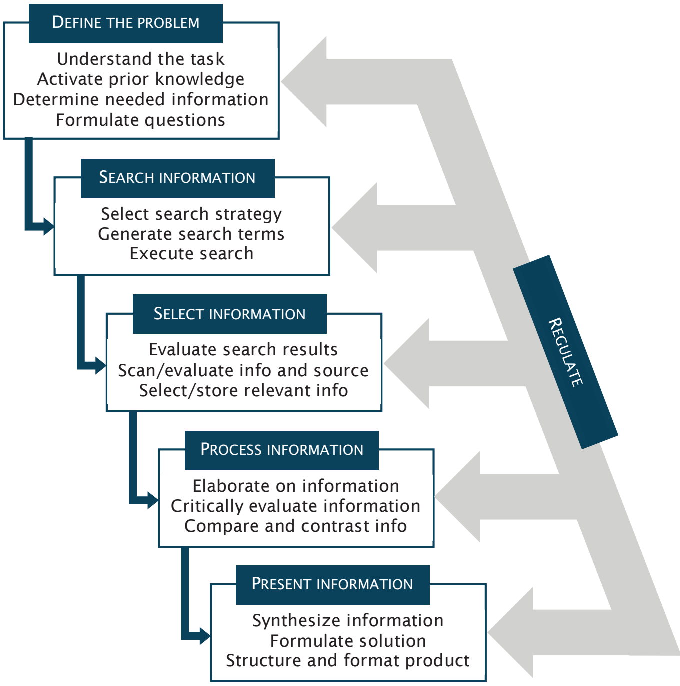


Figure 1.1. Overview of the skill 'information problem solving' (based on Brand-Gruwel et al., 2005)

Executing a search leads to a search engine results page (SERP), generally containing 10 links to information sources accompanied by the resource page's title, URL, and a small snippet of text. Careful evaluation of these information elements is needed to judge which seem useful enough to click on. Making good choices on a SERP avoids wasting time on irrelevant or untrustworthy sources of information and makes for an efficient IPS process. After accessing a source, its relevance and trustworthiness needs to be evaluated. *Relevance* relates to the amount of *on-topic* or *sought-for* information, while *trustworthiness* is an indication of the *reliability* of the information, determined by its publication date, author's expertise, reputation of the author's affiliation, quality of argumentation, etc. Searchers iterate between search queries, SERPs, and information sources to extract the information deemed useful for solving the problem. In the end, it is necessary to process the information to construct new knowledge and formulate an answer to the question(s). In the educational context, these solutions are often presented as reports, presentations, or essays.

Formulating good solutions always requires a certain degree of proficiency in all aspects of IPS. For example, good students with underdeveloped search skills will likely underperform, as their queries will mostly return less relevant results, unless they get a lucky hit. Similarly, even good searchers fail to reach good solutions if they did not define the problem carefully before starting their search. Therefore, it is important that students experience the interrelatedness of these skills and learn how errors early in the process lead to suboptimal performance. This allows students to regulate their IPS process and revisit previous phases when they encounter problems or struggle with a certain phase. For the same reason, assessment of IPS should focus also on the process, and not only on the product. Teachers who only assess outcomes (e.g., reports or essays), may be unable to diagnose which aspects of IPS are underdeveloped. Often, good searchers tend to find a perfectly good collection of sources to formulate a good answer to the question, but may not be able to present their reasoning in a well-structured product. Because they lack the presentation skills to showcase their solution, the teacher will then be unaware that most skills are already mastered, apart from presentation skills.

However, presenting is not the only aspect with which students struggle. Research exploring the problems students encounter with IPS shows there are major deficiencies in all aspects in people of all age groups (Walraven, Brand-Gruwel, & Boshuizen, 2008). This once more indicates that IPS is a complex cognitive skill, and instruction should focus on teaching these skills and subskills in an integrated and coordinated fashion, working on developing the necessary declarative knowledge, practicing the necessary skills, and forming the right attitudes in order to learn how to deal with new information problems.

Instruction for IPS

For fostering complex cognitive skills, using real-world learning tasks can promote the transfer from the learning setting to daily practice. It provides an opportunity to present learners with whole tasks that encompass the whole range of constituent skills needed for task performance, while being properly supported and guided. Several models of learning and instruction support using real-world problems, such as cognitive apprenticeship (Collins, Brown, & Holum, 1991), first principles (Merrill, 2002), and four-component instructional design (4C/ID; Van Merriënboer, Clark, & de Croock, 2002; Van Merriënboer & Kirschner, 2018). In a review of task-centered learning models, Francom (2014) shows that many of these task-centered learning models share five common prescriptions that resemble Merrill's first principles of instruction (Merrill, 2002). First, learning should be centered around tasks that are based on real-world performance. Second, prior knowledge should be activated. Third, performance of a learning task should be demonstrated or modeled. Fourth, learners should be able to apply their new knowledge and skills in tasks while being supported. Fifth, learners should be able to integrate new knowledge and skills in daily practice. Instruction developed according to the 4C/ID model (Van Merriënboer & Kirschner, 2018) incorporates these principles and was found effective for the development of complex skills in domains of technical expertise (Sarfo & Elen, 2007), communication (Susilo, van Merriënboer, van Dalen, Claramita, & Scherpbier, 2013), electrical skills

(Melo & Miranda, 2015), and medical education (Vandewaetere et al., 2015). It was therefore chosen as the instructional design model for the instructional interventions that are evaluated in the studies in this dissertation. Figure 1.2 shows an overview of the four components.

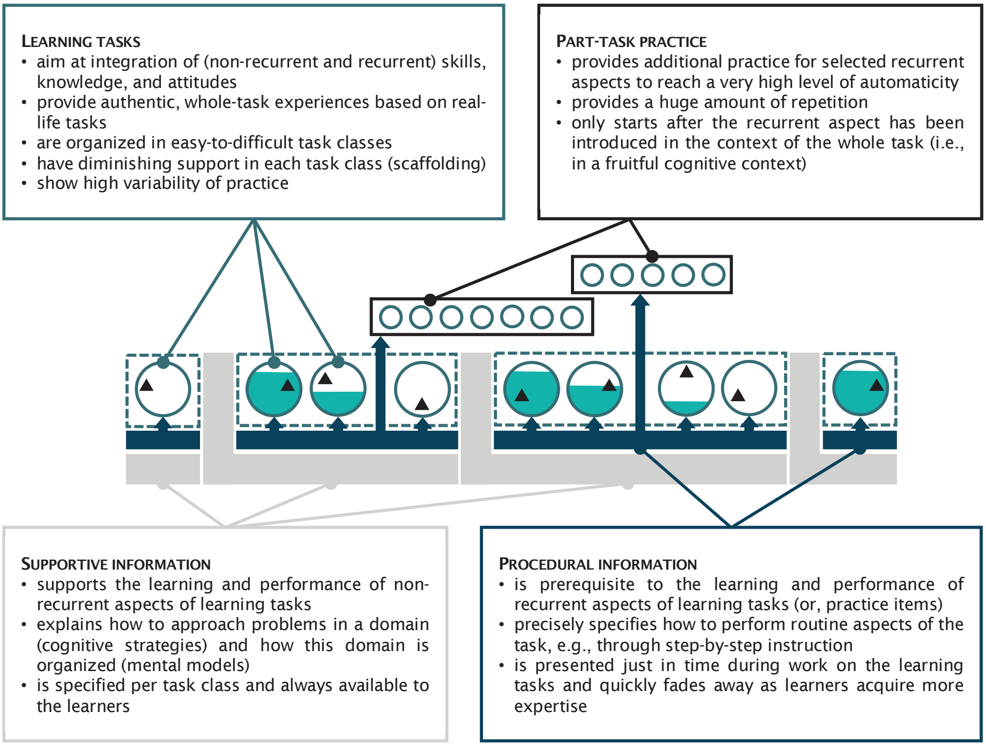


Figure 1.2. Four-component instructional design (4C/ID) model (based on Van Merriënboer & Kirschner, 2018)

For the design of IPS instruction, learning tasks constitute online information problems, appealing to all or most skills of the IPS process and their constituent parts. In such tasks, students are required to read the problem description, generate a problem statement or question, generate search terms, execute search queries, evaluate SERPs, sources, and information to produce a solution to the problem. Authentic and relevant tasks stimulate *inductive learning*: generalizing from concrete experiences. Supportive information can be included as instructional videos, web resources and worked-out examples. Studying this information leads to *elaboration*: connecting new information to existing cognitive schemas. As IPS is a primarily cognitive problem-solving process, not many recurrent and routine skills are required. The few relevant recurrent skills necessary for IPS (i.e., instrumental skills such as typing or operating a web browser) are often sufficiently developed, making procedural information and part-task practice mostly redundant (Van Deursen & van Dijk, 2009).

When designing instruction for fostering complex skills, it is essential that delivery methods are carefully considered. The experiments presented in this dissertation mainly make use of online or blended delivery, which requires the development of digital instructional materials such as instructional videos, digital modeling examples, or online presentations. It is therefore important those materials are designed to facilitate learning from multimedia formats. The design and quality of these materials can impact learning, as badly designed materials are known to increase unwanted cognitive load and hinder learning (Van Merriënboer & Ayres, 2005). To avoid such negative effects, Mayer's principles of multimedia learning were applied during the development of materials (Mayer, 2014). For example, for instructional videos, cognitive overload was avoided by applying the redundancy principle, which states that graphics, narration, and on-screen text should not appear simultaneously.

In practice, it appears most educational institutions struggle with the application of instructional design principles and encounter problems with the implementation of IPS instruction (Badke, 2010). Its importance is acknowledged, but most schools encounter great difficulty in finding a suitable place and time in their curricula, often leading to subpar instruction in short library training sessions (Derakhshan & Singh, 2011; Probert, 2009). Such part-tasks forego the benefits of whole-task instruction and lead to fragmentation (Lim, Reiser, & Olina, 2009). Research shows that whole-task approaches for teaching a complex skill such as IPS show potential (Wopereis, Frerejean, & Brand-Gruwel, 2015), and embedding IPS instruction within a meaningful context, presenting it simultaneously with domain-specific instruction can lead to deeper learning and improved transfer (Perin, 2011; Wopereis, Brand-Gruwel, & Vermetten, 2008). Research on instructional interventions for IPS often focus on a subset of the constituent skills (e.g., Britt & Aglinskas, 2002), or do not let learners apply their IPS skills in an authentic context, for example by restricting the number of potential information sources or making use of prefabricated SERPs (e.g., Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Gerjets, Kammerer, & Werner, 2011). In addition, many of the studies on IPS interventions focus on short-term learning effects and lack measurements of transfer or delayed learning effects.

This dissertation

The aim of the research carried out for this dissertation is to investigate instructional design principles in order to formulate practical guidelines for teachers and instructional designers who wish to design instruction for effective and efficient IPS. It attempts to overcome shortcomings of previous research and focuses on the application of IPS skills in ecologically valid and realistic settings, making use of authentic learning tasks that require integration of the skills, knowledge, and attitudes necessary for effective and efficient IPS. In addition, the instructional interventions presented incorporate measurements of transfer or delayed learning effects. More specifically, the following research questions are addressed:

- What are the effects of built-in task support (e.g., completion tasks, emphasis manipulation) on the acquisition of IPS skills?
- What are the effects of a modeling example on the acquisition of IPS skills?
- What are the effects of embedded IPS instruction on the acquisition of IPS skills?
- What are the general characteristics of students' IPS process, and how do student, query, and source characteristics predict the selection of relevant and trustworthy sources?

The research presented in *Chapter 2* investigates the principle of applying new knowledge and skills while receiving *built-in task supp*ort during task performance. More specifically, two approaches to task support are compared in a standalone online IPS training using whole tasks. The completion strategy, a sequence of learning tasks containing a decreasing number of worked-out steps, is compared to emphasis manipulation, an approach where students receive additional support on a single aspect of the task in each learning task. *Chapter 3* presents a study investigating the principle of demonstration. Using the same online training as described in Chapter 2, students receiving a video *modeling example* were compared to students performing a practice task. For these two studies, an online learning environment was developed consisting primarily of web search tasks and video materials for support. The studies reported in Chapters 2 and 3 were implemented as standalone training sessions as part of a university curriculum. A situational judgement test was developed to measure students' IPS skills in a short timeframe, based off an adaptation of a rapid online method for measuring domain-specific knowledge (Kalyuga, 2008). Invested mental effort was also measured using self-reports.

In contrast to the standalone sessions in Chapters 2 and 3, the study reported in *Chapter 4* deals with whole-task instruction *embedded* in an existing educational program. As such, it deals with the principle of application in an ecologically valid setting. In this study, an existing curriculum in a teacher training program was partly redesigned to include embedded whole-task IPS training. As the original program offered mainly face-to-face education, it was decided to include a parallel online environment for practicing IPS tasks embedded in the curriculum. The resulting blended learning setting was evaluated by comparing students receiving the regular curriculum with students receiving the redesigned curriculum including IPS training. Students' performance on authentic tasks was assessed by logging and retrospectively scoring all learner actions, such as selected sources and generated queries. The study presented in *Chapter 5* applied this method of assessment to provide detailed insight on students' search and evaluation skills. It elaborates on the method applied in Chapter 4 and uses the collected log files to perform a deep inspection of students' search processes. Discussion of the results and their implications for teachers, instructional designers and researchers is presented in the general discussion in *Chapter 6*.

CHAPTER 2

COMPLETION STRATEGY OR EMPHASIS MANIPULATION?

TASK SUPPORT FOR TEACHING INFORMATION PROBLEM SOLVING

ABSTRACT

While most students seem to solve information problems effortlessly, research shows that the cognitive skills for effective information problem solving are often underdeveloped. Students manage to find in- formation and formulate solutions, but the quality of their process and product is questionable. It is therefore important to develop instruction for fostering these skills. In this research, a two-hour online intervention was presented to first-year university students with the goal to improve their information problem solving skills while investigating effects of different types of built-in task support. A training design containing completion tasks was compared to a design using emphasis manipulation. A third variant of the training combined both approaches. In two experiments, these conditions were compared to a control condition receiving conventional tasks without built-in task support. Results of both experiments show that students' information problem solving skills are underdeveloped, which underlines the necessity for formal training. While the intervention improved students' skills, no differences were found between conditions. The authors hypothesize that the effective presentation of supportive information in the form of a modeling example at the start of the training caused a strong learning effect, which masked effects of task support. Limitations and directions for future research are presented.

THIS CHAPTER IS BASED ON:

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2016). Completion strategy or emphasis manipulation? Task support for teaching information problem solving. *Computers in Human Behavior, 62*, 90-104. doi:10.1016/j.chb.2016.03.048

INTRODUCTION

Searching the web for information seems effortless for students; they simply navigate to a popular search engine, type in a couple keywords, and select some of the sources that appear to be relevant (MaKinster, Beghetto, & Plucker, 2002). Most students easily find their way without any explicit instruction. They paraphrase, cite, or – in the worst case – copy and paste some of the text into their own document and the job is done (De Vries, van der Meij, & Lazonder, 2008). The abundance of information on the Internet is a bliss. While this may be viewed as a successful process in the eyes of the student, from an educational perspective it can be a waste of time. If the student is not equipped with the necessary skills, such as advanced search strategies and the ability to critically scrutinize information sources to determine relevance and reliability, chances are that the search process and the product fall short of what the teacher intended. It may be true that younger generations of students appear to quickly master the skills needed to navigate online information sources, but it is premature to claim that they automatically develop the skills to find correct and reliable online sources and learn from them (Kennedy, Judd, Churchward, Gray, & Krause, 2008; Kirschner & van Merriënboer, 2013; Rosman, Mayer, & Krampen, 2016a).

While most educational institutions acknowledge information problem solving (IPS) as an essential academic skill, they often struggle with implementation (Badke, 2010). To promote transfer of IPS to daily practice, it is advisable to practice these skills in different contexts and across different domains throughout the whole curriculum. This is problematic, and most schools experience great difficulty in finding a suitable place and time in the curriculum. Many, in turn, resort to providing nothing more than a short library training. To support teachers and faculty in embedding IPS skills in educational curricula, it is desirable to investigate which instructional approaches work well for IPS skills. This study takes a first step in that direction, describing the development and empirical testing of instruction for IPS skills, based on a solid instructional design model for teaching complex skills. Implications are discussed for both the domain of instructional design and information problem solving.

INFORMATION PROBLEM SOLVING

In educational settings, teachers often use information problems, where the necessary information to solve the problem is lacking, as an educational approach. The student is required to gather the missing information from external sources and combine the findings to construct a solution. Simple information problems, such as looking up the average monthly temperature in a country, pose little challenge for most students. Complex information problems, such as writing an essay on the effects of global warming on biodiversity, are a far more difficult challenge, because students will need to find, evaluate, and process sources of information that can vary greatly in terms of their trustworthiness, bias, reliability, or can contain contradictory information. Teachers often expect that having students search for information will automatically lead to their learning (Kirschner, Sweller, & Clark, 2006). But correctly and efficiently solving an information problem is a complex higher-order cognitive competence requiring a broad range of different cognitive skills that these students might not possess. The range of skills has been summarized as a 5-step model (see Figure 2.1) in which students iterate between the stages 'define the problem', 'search information', 'select information', 'process information', and 'present information', each step consisting of several constituent skills (Brand-Gruwel, Wopereis, & Vermetten, 2005; Brand-Gruwel, Wopereis, & Walraven, 2009).

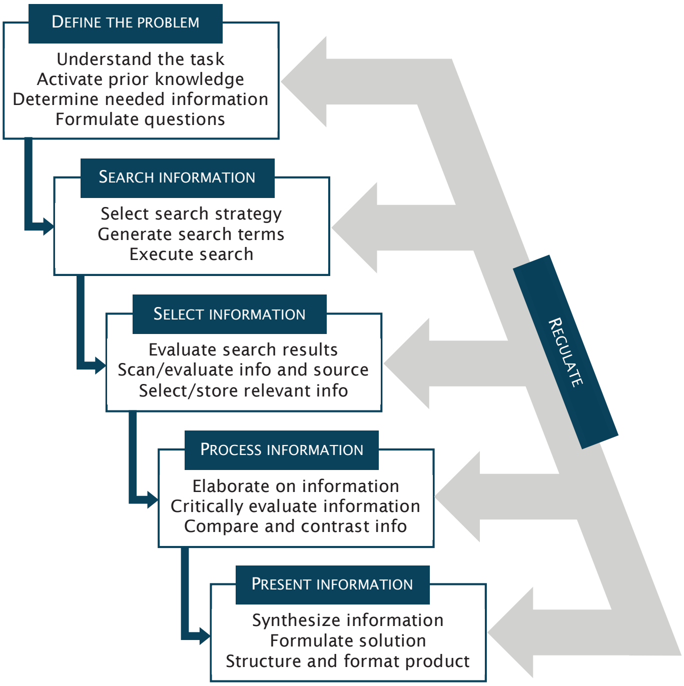


Figure 2.1. Overview of the skill 'information problem solving' (based on Brand-Gruwel et al., 2005)

To solve an information problem, the learner first needs to reach an understanding of the task and identify the needed information to define and delimit the task domain. In this step, formulating a clear and concise question is essential to stay focused and avoid unnecessary deviations while searching. Second, search terms need to be generated and tried out in a search engine. By identifying key concepts from the question and then systematically changing, adding, or removing terms while correctly using the available Boolean operators, the learner maximizes the chance to find relevant information sources. Third, it is important to maintain a critical attitude while evaluating the search engine results page (SERP), the subsequently visited information sources, and the information itself. Critical scrutiny avoids spending time on irrelevant websites or becoming occupied with information that is outdated, false, or which originates from unreliable or biased sources. Fourth, when relevant and reliable sources are found and stored, the learner needs to process their contents, deal with overlapping and conflicting information, and synthesize the different elements chosen from the separate sources. Finally, the solution can be presented in a product such as an essay or a presentation, depending on the task. It is important that the product clearly answers the question that was defined earlier in the task. Moreover, during all of these steps, the learner should regulate the search process, decide whether sufficient useful information has been found, and steer the process to avoid deviations or distractions.

Previous research indicates students may quickly develop the instrumental skills needed to operate digital devices and use software and Internet browsers, but IPS skills are generally underdeveloped or absent. In a comparison of experts and novices, Brand-Gruwel et al. (2005) found that novices took less time for orientation, chose less effective keywords, judged and evaluated sources less often, and hardly regulated their process. In a literature review, Walraven, Brand-Gruwel, and Boshuizen (2008) discuss several studies that show execution of IPS skills leave much room for improvement for all age groups. Similarly, studies by Van Deursen and van Dijk (2009) and Van Deursen and van Diepen (2013) show users of all ages experience problems with query formulation, evaluation of search results and processing of information.

Two things become clear from these findings. First, IPS is a complex higher-order cognitive skill. Successful problem solving depends on the existence of knowledge, the mastery and coordination of a set of skills and the adoption of a critical attitude. Second, research shows clear deficiencies in students of almost all ages. In general, students' IPS skills are often overestimated or expected to develop naturally over time. These IPS skills may not be of the level that is often expected of the student problem solver, or from the so-called 'digital natives' (see also: Kirschner & van Merriënboer, 2013). Providing students with a complex task for which they do not possess the required skills risks overloading their memory systems and lowering task performance and learning. Therefore, the development of evidence-based instruction for fostering IPS skills is warranted.

Instructional design for complex learning

Complex learning is defined as "the integration of knowledge, skills and attitudes; coordinating qualitatively different constituent skills; and often transferring what was learned in school or training to daily life and work" (Kirschner & van Merriënboer, 2009, p.244). The Four Component Instructional Design (4C/ID) model provides an extensive blueprint and approach for developing instruction to achieve complex learning, based on solid psychological and educational research (Van Merriënboer & Kirschner, 2013). First, the model advocates the use of authentic, whole tasks that require integration of knowledge, skills and attitudes, and coordination of constituent skills. Second, it provides guidelines to correctly provide the information needed to solve the problems: domain knowledge and a structured approach to solve the problem. Third, it advises providing just-in-time procedural information during the tasks to aid problem-solvers with routine tasks. The fourth component, part-task practice, is necessary when performance of these routine tasks needs to be automated.

This task-centered approach confronts learners with a series of whole tasks in the learner's zone of proximal development. Task complexity increases to keep up with learner progress. However, especially in the early phases of learning, tasks can be too complex for the learner, because they introduce too many interacting elements or the learner's knowledge schema are insufficiently developed. In these cases, the learner's memory system may become overloaded, which can negatively impact learning (Paas & van Merriënboer, 1994). In situations of complex learning and authentic tasks, there are many elements that potentially increase the amount of cognitive load experienced by the student. It is therefore essential that instructional designers take great care to reduce unnecessary load, while maintaining activities that induce germane load and lead to learning.

For IPS specifically, task complexity is not the only factor that influences the demands on working memory during problem solving, and in consequence, learning and instruction. Rouet (2009) summarizes additional factors in a conceptual framework comprising three dimensions: individual variables, information resources, and problem context. Instructional designers should be aware that personal factors, such as an individual's domain-specific knowledge (Monchaux, Amadieu, Chevalier, & Mariné, 2015), age (Chevalier, Dommes, & Marquié, 2015), attitudes and biases (Ford, Miller, & Moss, 2005; Van Strien, Brand-Gruwel, & Boshuizen, 2014), epistemic beliefs (Kammerer, Bråten, Gerjets, & Strømsø, 2012), and reading skills (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011) can affect the learning process and outcomes. Similarly, source factors (DeStefano & LeFevre, 2007) and task type (Wirth, Sommer, von Pape, & Karnowski, 2015) may influence variables in the learning process. While most of these factors lie outside the designer's influence, they all affect the demand imposed on working memory during the IPS process.

For situations where tasks may be too demanding for a learner to complete successfully, the problem-solving process must be supported (Van Merriënboer, 2013). The 4C/ID model stresses the importance of built-in task support. While learners can be supported in many ways (i.e. with case studies, modeling and/or worked examples, inducing reflection, etc.), the current experiments focus on two approaches that appear most applicable to IPS instruction, namely the completion strategy and emphasis manipulation.

Completion tasks

A completion task is a problem where the learner is provided with a given state and a partial solution. After studying the partial solution and the given information, the learner then has to complete the remaining solution steps in order to solve the problem (Van Merriënboer, 1990; Van Merriënboer & de Croock, 1995). This approach is effective for several reasons. First, completion tasks inherently stimulate active processing of the given solution steps because they contain essential information the learner needs to process before being able to continue. In addition, the provided solution steps are examples of a correct systematic approach to solving the problem. This enables learners to study the examples and by induction generate schemas of correct solution strategies themselves (Van Merriënboer, 2013). Studying correct examples (albeit partial solutions) can often be more effective than solving whole problems, especially early in the learning process (Renkl & Atkinson, 2003). When learners lack the necessary schemas and strategies, they will fall back to naïve and ineffective strategies such as means-end analyses or trial-and-error to solve the problem. Providing sufficient worked-out steps in this phase can avoid this (Van Gog, Paas, & van Merriënboer, 2004).

The second benefit of using completion tasks is that a designer can change the number of worked-out steps to adapt the task to the learner's level. For learners in an early learning phase, it would be beneficial to increase the number of worked-out steps (e.g., one or even no steps missing), providing ample examples of correct complete or partial solutions and allowing the learner to induce the necessary schemas and strategies (Atkinson, Renkl, & Merrill, 2003; Renkl, Atkinson, & Große, 2004). In later learning phase, learners benefit more from more conventional tasks that contain just a few worked-out steps. Offering too many worked-out steps to these learners would create the risk of inducing the expertise-reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga & Sweller, 2004). By gradually reducing the number of worked-out steps as a learner progresses, the amount of support that is offered corresponds more closely to the amount of support that is actually needed. In the context of IPS, this fading of solution steps can only be applied backward, meaning that worked-out solution steps late in the process will always fade before solution steps early in the process. To illustrate, consider the opposite: A worked-out example where the solution and information sources are given but the student needs to define the problem and generate search terms. Such a backward information problem is unrealistic, and practicing it has little purpose. In conclusion, a gradual transition from completely worked-out problems to conventional problems would be a good strategy for instruction: an approach dubbed the completion strategy.

Wopereis, Frerejean, and Brand-Gruwel (2015) implemented the completion strategy in a university-level IPS training program. In their training, an example completion task provides students with a problem orientation, a well-formulated problem statement and research question, and a partial list of search terms. In this case, the step 'problem definition' is completely worked out, and the step 'searching' is partially worked out. Students are required to process the problem orientation to become familiar with the task domain and to activate any prior knowledge. The given problem statement and research question provide a clear direction for the search and inform them which information is needed, and consequently, which information is not. Based on this orientation, students then extend the list of search terms and proceed with the search for information and the remaining solution steps ('select information', 'process information', and 'present information'). Compared to a conventional task where students perform the whole task, this approach requires less decision making - and therefore less room for error - and provides an additional example to learn from. The expectation here is that such tasks will impose fewer cognitive demands than conventional problems.

Emphasis manipulation

Students can also be supported by guiding them in the allocation of their attention to a certain skill (i.e., generating search terms) or a step in the process (i.e., select information) within a learning task. Students then perform the whole task from beginning to end, but just one aspect of the solution procedure is emphasized, often by instructions and feedback. In subsequent tasks, the emphasis and thus the allocation of the learner's attention shifts to a different aspect of the task. Note that the task is not broken up into part-tasks, but only the relative emphasis of the selected aspect varies. All skills are still performed in the context of the whole task. This approach, called emphasis manipulation or emphasis change (Gopher, 2007; Gopher, Weil, & Siegel, 1989), reduces strain on working memory because not all instruction needs to be kept available in working memory, and attention is focused on a single aspect, not divided over all aspects.

The emphasis change approach was effective in a training regime for a high-workload computer game called Space Fortress and in several dual-task settings (Gopher, 2007). In other research, students who received whole-task training with emphasis change were less easily disrupted by a concurrent task than students receiving part-task training (Fabiani et al., 1989). In addition, Yechiam, Erev, and Gopher (2001) demonstrated that an emphasis change approach is more effective than guided instruction in settings where searchers quickly converge to suboptimal strategies. The idea here is that problem solvers make only small changes to their current, suboptimal, strategy and insufficiently explore more diverse solution strategies, a process called melioration (Yechiam, Erev, Yehene, & Gopher, 2003). Emphasis change protocols facilitate the exploration of other, potentially more effective strategies.

The errors that can be observed when novices search the web may be a sign of melioration. Lacking sufficient skill, they employ naive strategies that will find some results (partly due to increasing quality of search engines), even though it may not be the information they are looking for. This will then lead them to obtaining suboptimal information, which in turn leads to a suboptimal solution to the task. Students experience the success of solving the problem, which reinforces their current behavior and leads to a similar approach to the next problem. Students see no reason to expend extra effort to significantly change their strategy. Emphasis change can encourage students to explore other strategies, such as more extensive planning, or using thesauri to generate keywords, which increases the chance of a more effective or efficient problem-solving process.

Placing emphasis on specific aspects of a task can be done by incorporating instruction and feedback during those specific aspects of the learning task. A simple and effective method to provide instruction and feedback in an online environment is by using prompts (see: Stadtler & Bromme, 2008). In the case of IPS, three types of prompts are effective: anticipative prompts delivered before execution of the targeted skill, instructional prompts delivered just in time before execution of the targeted skill, and reflection prompts delivered after performing the skill.

Consider a student working on a learning task where the skill evaluating sources is emphasized and therefore accompanied by prompts. Before she starts evaluating sources (i.e., the targeted skill), she is prompted: “Describe your approach to the next step. Where will you focus your attention?” By articulating her upcoming actions before performing the skill, anticipative reasoning, a skill found in effective problem solvers, is stimulated (Renkl, 1997). The student answers: “I’ll look at the result list and click on some of the titles that seem interesting. I’ll then read that text. If it seems relevant, I’ll probably use it.” The answer reveals that her solution schema is still incomplete, and that she has not yet learned to evaluate a search engine results page or an information source. Merely activating knowledge is therefore not sufficient. Her current schemas or strategies need to be corrected or completed.

She is prompted again, this time simply with instructions. The instructional prompt explains how to evaluate search results (i.e., pay attention to domain names, publication dates, snippets) before clicking a link and how to judge information sources (i.e., take into account author reputation, target audience, information goal, publication date). It essentially gives general feedback on her previous answer. The student will acknowledge that her previously articulated approach was incomplete and that she should not merely click ‘interesting’ links and use ‘relevant’ information. She learns that there are many more criteria to use to discriminate between interesting and relevant. She then processes this information and immediately carries out the solution step, with this new knowledge in memory. The subsequent application of the new knowledge stimulates assimilation into knowledge schemas.

To enforce this process, a reflection prompt can be delivered after the step is performed: “How did it go? Did you encounter any problems?” This prompt induces reflection and forces her to look back at how she applied the new knowledge, which should reinforce the use of a correct or more effective solution strategy (Saito & Miwa, 2007; Stark & Krause, 2009). Taken together, this combination of three prompts, the prompt triad, fulfills the purpose of emphasis manipulation by first lowering cognitive demand by focusing student attention to a particular aspect of the task while leaving the whole task intact and then promoting improvements in strategies by activating and correcting current knowledge schema.

The present study

Seemingly little research has focused on the development of holistic instruction for IPS. Most studies either focus on elements of instruction, such as feedback (e.g., Timmers, Walraven, & Veldkamp, 2015), restrict the search space to prefabricated portals (De Vries et al., 2008), or focus instruction on elements of the skill, such as source evaluation (Walraven, Brand-Gruwel, & Boshuizen, 2010). Some are focused on classroom interventions (Argelagós & Pifarré, 2012; Kuiper, Volman, & Terwel, 2008). In the current study, a holistic approach for teaching the complete skill in individual (online) instruction is adopted and a first step is taken towards developing instruction based on whole tasks with built-in task support. Two experiments were conducted to investigate the effects of two forms of task support (*completion strategy* vs. *emphasis manipulation*) on the acquisition of IPS skills in a short online training. This training was embedded as a standalone practical assignment in university students' first-year curriculum. As an intervention in a naturalistic setting, this training aimed to develop students' IPS skills while detecting differences in the extent of learned skills due to the different methods of support. It was expected that students who receive at least one form of task support (i.e., completion tasks and/or emphasis manipulation) will perform better than students who do not receive task support (Hypothesis 1) and students who receive a combination of both forms of support will perform better than students who receive only a single form of support (Hypothesis 2). To help explain differences in learning outcomes, students were asked to report the required mental effort at several points during the learning phase.

Experiment 1

Method

Participants

A total of 96 students between 18 and 24 years old (M_{age} = 18.7 years) participated in this experiment, 89 of whom were female (92.7%) and seven were male (7.3%). All participants were first-year Pedagogical Science students at a Belgian university.

Experimental design

The experiment was a regular pretest-posttest design with four conditions. All conditions received a two-hour online training consisting of an instructional video, a modeling example, and four learning tasks. Each condition received a different form of task support during three of the four learning tasks. The first condition received task support in the form of the completion strategy combined with emphasis manipulation (CS+EM). The second condition received completion tasks, but no emphasis prompts (CS). The third condition received emphasis prompts, but no completion tasks (EM). The fourth condition was a control condition and received conventional learning tasks without support. The different forms of task support are further detailed in the section 'Task support'.

Materials

Online training

In a two-hour classroom session, students received an online training that started with a 14-minute instructional video introducing the five steps of the IPS process (i.e., 'define', 'search', 'select', 'study', 'present') including their constituent skills. The instructional video was followed by a modeling example: a 10-minute screencast in which a fictitious expert showed a systematic approach to solving an information problem. This modeling example was split into four short fragments that ended with the questions "What do you think of the actions of the expert?" and "How does this differ from your current approach?" intended to stimulate students to formulate explanations and stimulate active processing of the example (Atkinson et al., 2003;

Renkl & Atkinson, 2002). These elements formed the supportive information component in the 4C/ID model.

The training further comprised four learning tasks in the form of a web search exercise. Students received a problem description and had approximately 15 minutes to search the web for information and formulate a solution to the problem. The topics were: effects of stretching before sports, effects of electromagnetic radiation from cellphones, effects of violence in videogames, and effects of using media devices before sleeping. The learning tasks guided students through the problem-solving steps with on-screen instructions. Students were asked to explicitly formulate research questions and search terms, and list the URL of four sources that contributed to their solution, along with an explanation of why they chose these sources. At the end of the learning task they formulated a solution in a few sentences. Each of the experimental conditions received a different form of support during learning tasks 1 to 3. A fourth and final task was presented that did not include any support or guidance, but simply gave a problem description and a textbox for an answer. This task was identical for all students and contained no explicit instruction.

Task support

For the three experimental conditions, learning tasks 1 to 3 contained built-in task support in the form of completion tasks, emphasis prompts, or both. These tasks were designed in a way to support the problem-solving process without overloading the student. While the IPS-I model (Brand-Gruwel et al., 2009) describes a five-step approach, to comply with time constraints in this experiment the steps 'select information' and 'process information' were merged to a single step and no task support was supplied on the final step: 'present information'. Presenting information can be done in countless ways, and providing support on this skill would be very time-consuming. Additionally, students likely benefit more from support on the first four steps than from support on presenting information. In conclusion, task support is offered on the steps 'define the problem', 'search information', and 'select & process information'.

The control condition received no task support at all, meaning that they work through each learning task by following the guidance on the screen that take them through the steps 'define the problem', 'search information', 'select & process information' and 'present information'.

The EM condition received emphasis manipulation, meaning that each learning task contained one solution step that was emphasized with a prompt triad: an anticipative prompt and an instructional prompt before execution of the step, and a reflective prompt afterwards. In learning task 1, the step 'select & process information' was emphasized, in learning task 2 the step 'search information', and in learning task 3 the step 'define the problem'.

The CS condition received completion tasks. In these tasks, some solution steps are already worked out and replaced with a very short video (approximately one to two minutes) of a fictitious expert reasoning through the solution step. No further action was required. The worked-out steps were faded backwards, meaning each subsequent learning task contained one less worked-out step and students were therefore required to perform one step more in each learning task. In learning task 1, the steps 'define the problem' and 'search information' were worked out and students had to select and process sources from the given search results to formulate a solution. In task 2, only 'define the problem' was worked out and all other steps had to be performed. In task 3, no steps were worked out.

Finally, in the CS+EM condition completion tasks and emphasis prompts were combined, meaning that the prompt triad was added to the first step that followed the worked-out steps. This entails that in task 1, the first two steps were worked out: 'define the problem' and 'search information'. The next step, 'process & select information' was emphasized with a prompt triad. The final step had to be performed without support. In task 2, only the first step was worked out and emphasis shifted to 'search information'. In the third task, no worked-out steps were given and 'define the problem' was emphasized. See Figure 2.2 for a graphical overview of the experimental design.

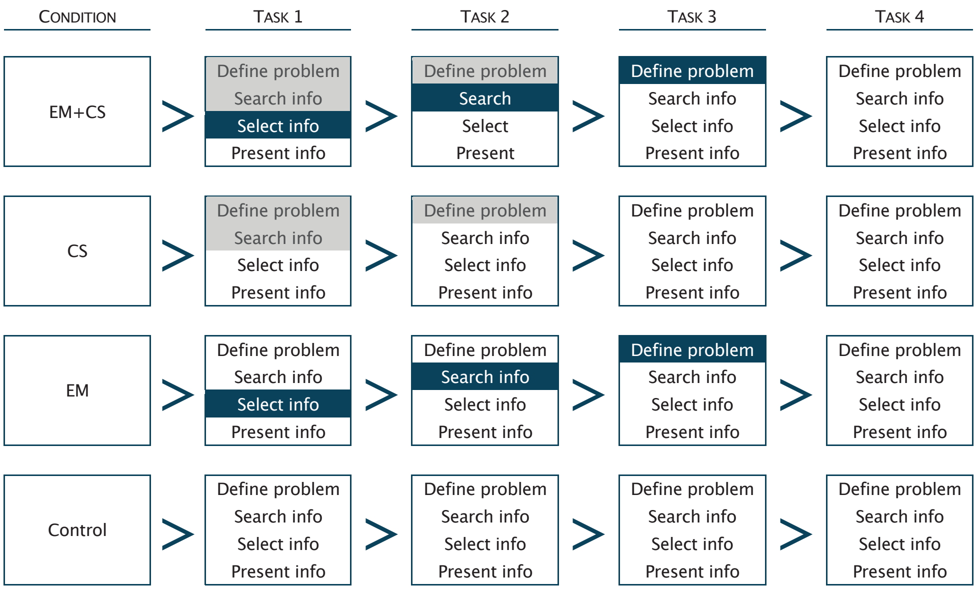


Figure 2.2. Overview of the experimental design: four conditions (rows) received four learning tasks (columns) that consist of four steps. Worked-out steps in these tasks are marked with gray, emphasized steps are colored. Steps that are not colored contained no built-in support.

Measurement of IPS skill

To measure IPS skills in such a short timeframe, an online skills test was developed that aimed to reveal the student's level of performance without requiring the performance of another whole task. The tests confronted students with seven fabricated situations that occur during an information problem and asked them to formulate their next action. This closely mimics a realistic task situation. To ensure validity, the items were based on important subskills in the first four steps of the IPS-I model by Brand-Gruwel et al. (2009). The step 'present information' was not measured because presentation of a problem solution is a multifaceted skill too difficult to measure quickly, and the training did not include support on this step. See Table 2.1 for an overview of the pretest and posttest items. For example, the item corresponding to 'select information' showed a fabricated SERP and asked students to indicate which sources they would select and why. Answers were scored blindly, based on a task-specific rubric that resulted in a maximum subscore of four points per step, for a maximum total of 16 points. The scoring sheet and procedure are included in Appendix 1. The items in the pretest concerned the topic (i.e., problem domain) of gender-specific education. The posttest items were identical to the pretest items, but on the topic of the malleability of intelligence. A second experimenter rescored 20 randomly chosen participants in order to obtain a measure of inter-rater agreement.

Table 2.1. *Overview of pretest and posttest*

Item	Step	Subskill	Given	Question
1	Define the problem	Problem orientation	A problem description	How would you start this task? What is your first step and why?
2	Define the problem	Formulating a question	A problem description	Which problem statements would you formulate? Why do you choose these?
3	Search information	Generating search terms	A problem description	Which search query would you type into Google? Formulate two alternative search queries.
4	Select information	Evaluating search results	A fabricated SERP	Which three websites would you select? Why did you select these websites?
5	Process information	Scanning a source	A screenshot of a text-rich website, zoomed out so the text is unreadable	What do you do when you visit a text-rich website and want to find out if it contains relevant information? How do you proceed?
6	Process information	Evaluating information	A short text fragment containing an argument given by an expert	Which criteria do you use to determine whether information is useful for your task? What are your conditions for use?
7	Process information	Contrasting information	Two short, contradicting arguments	How do you deal with contradicting information? How does this affect your solution? Explain.

Mental effort rating

Integrating and coordinating the skills, knowledge and attitudes that are required to effectively and efficiently solve an information problem is a complex activity that places high demands on the learner's memory system. To bring down this complexity, built-in task support is incorporated in the learning tasks. It can be expected that different types of support impose different amounts of cognitive load on the students. Lacking an objective, direct way to measure cognitive load, experienced mental effort was measured as a proxy. During the learning phase, each learning task ended with a short measurement of experienced mental effort: a 9-point mental effort rating scale

(Paas, 1992): *How much effort did it take to perform this task?* While all students were instructed to spend approximately 15 minutes on each learning task, working through the extra prompts and worked-out steps may have increased time on task for those students and perhaps put the students under time pressure. Performing the task under high time pressure might cause an increase in experienced mental effort. Therefore, time pressure was explicitly measured with the temporal demand item from the NASA-TLX (Hart & Staveland, 1998): *How hurried or rushed was the pace of the task?*

Data analysis

The scores on the pretest and posttest were analyzed with a repeated measures analysis of variance with *type of support* (CS+EM vs. CS vs. EM vs. Control) as a between-subjects variable and *time of test* (pretest vs. posttest) as a within-subjects variable. The same analysis was conducted on the subjective mental effort rating and the time pressure rating but with *learning task* as a within-subjects variable. In addition, an analysis of variance was conducted on the ratings per learning task to investigate differences in required mental effort between conditions.

Procedure

The training was embedded in the students' current curriculum as a practical assignment and offered in four different timeslots. Students were free to choose a timeslot that fit their schedule. During the two-hour training session, students took place at a computer in the university computer room and logged in to the online learning environment. After logging in, students first filled out a short preliminary questionnaire and were automatically randomly assigned to one of four conditions. They were instructed to work individually through the tasks they received on screen and informed that their screen content could differ from that of the other students. The experimenter asked students to spend approximately 15 minutes on each learning task, comparable to similar tasks used in other research (Lazonder, 2000; Lazonder, Biemans, & Wopereis, 2000). They then received the following: pretest, instructional video, modeling example, four learning tasks, and posttest. Each learning task concluded with the mental effort and time-pressure ratings. The instructional video and modeling example remained available via a link during the learning tasks. Before the posttest, students filled out a short evaluation and a final mental effort rating for the training as a whole. After the posttest, students signed for informed consent, received course credit and were subsequently dismissed. A debriefing with preliminary results followed eight weeks later.

Results

The four randomly generated conditions did not differ significantly on any of the items on the preliminary questionnaire, such as age or prior education. They reported equal amounts of time spent behind a computer per day, and no differences in the use of the computer for information retrieval (either for personal or educational goals), news, social media, chatting, and entertainment. The sample can therefore be considered homogeneous. Some data were scored as missing due to the fact that students answered questions with a dash or a space, and some data were lost due to incidental technical problems. On the posttest, missing values were substituted for their corresponding scores on the pretest as a best-guess – and indicating no progress – under the condition that only one value in that step was missing. If more values were missing, the corresponding subscore was also classified as missing data. Total scores on the posttest were treated the same: if more than one of the four subscores was missing, they were classified as missing value, otherwise the total was calculated over the remaining subscores.

Pretest and posttest scores

Inter-rater agreement on the scoring rubric for pre- and posttest was measured with a two-way mixed, absolute, single measure intra-class correlation and amounted to .878, indicating a reliable measure. Students scored rather low on the pretest, achieving a mean score of 41.86% (SD = 9.86). The scores varied between 18.75% and 62.5%. On the posttest, the mean score improved to 60.55% (SD = 11.16) with a range from 31.25% to 81.25%. Table 2.2 shows the mean scores per condition for the pretest and posttest. The repeated measures analysis showed that the between-subjects factor was not statistically significant: $F(3, 92) = .97$, $p = .410$, meaning that there was no effect of support and the scores did not depend on the type of support received. Indeed, the mean scores in Table 2.2 reveal that the four groups show a similar progression. The within-subjects factor did reveal a significant effect: $F(1, 92) = 187.46$, $p = .000$, $\eta^2_{\text{partial}} = .671$, indicating there was a substantial effect of training on the test scores.

Table 2.2. *Overview of scores (in percentage) per condition*

Condition	Pretest (*SD*)	Posttest (*SD*)
EM	43.75 (11.89)	63.07 (10.19)
CS	41.25 (9.02)	62.25 (11.05)
CS+EM	41.75 (8.79)	58.50 (12.48)
Control	40.89 (10.09)	58.59 (10.56)
Total	41.86 (9.86)	60.55 (11.16)

Mental effort ratings

The mental effort ratings showed a similar pattern: significant changes over learning tasks, but not between the conditions. The repeated measures analyses revealed no significant between-subjects effect: $F(3, 90) = .64$, $p = .593$, but a significant within-subjects effect: $F(3, 90) = 9.60$, $p = .000$, $\eta^2_{\text{partial}} = .100$. Contrast analysis further revealed that reported mental effort drops significantly from 5.21 (SD = 2.03) in learning task 3 to 4.36 (SD = 1.89) in learning task 4: $F(1, 90) = 18.14$, $p = .000$, $\eta^2_{\text{partial}} = .174$. Univariate ANOVAs per learning task revealed no differences between conditions. Figure 2.3 shows mean mental effort ratings for each condition and each learning task.

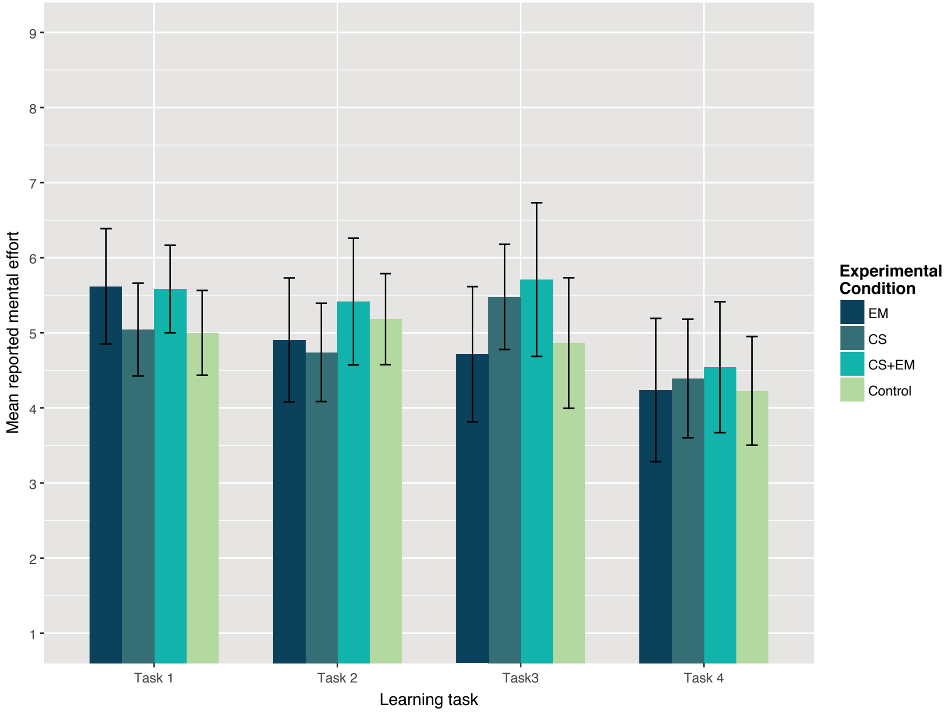


Figure 2.3. Reported mental effort per learning task for all conditions

Time pressure ratings

Analysis of time pressure showed that although scores were relatively high (all means above 5 on the 7-point scale), there were no within-subjects differences: $F(3, 89) = 1.01$, $p = .391$ or between-subjects differences: $F(3, 89) = .16$, $p = .923$. Therefore, students experienced similar time pressure in all conditions and in all learning tasks. Univariate ANOVAs per learning task confirmed this finding: on all four learning tasks, differences between conditions were not statistically significant. Figure 2.4 shows time pressure ratings for each condition and each learning task.

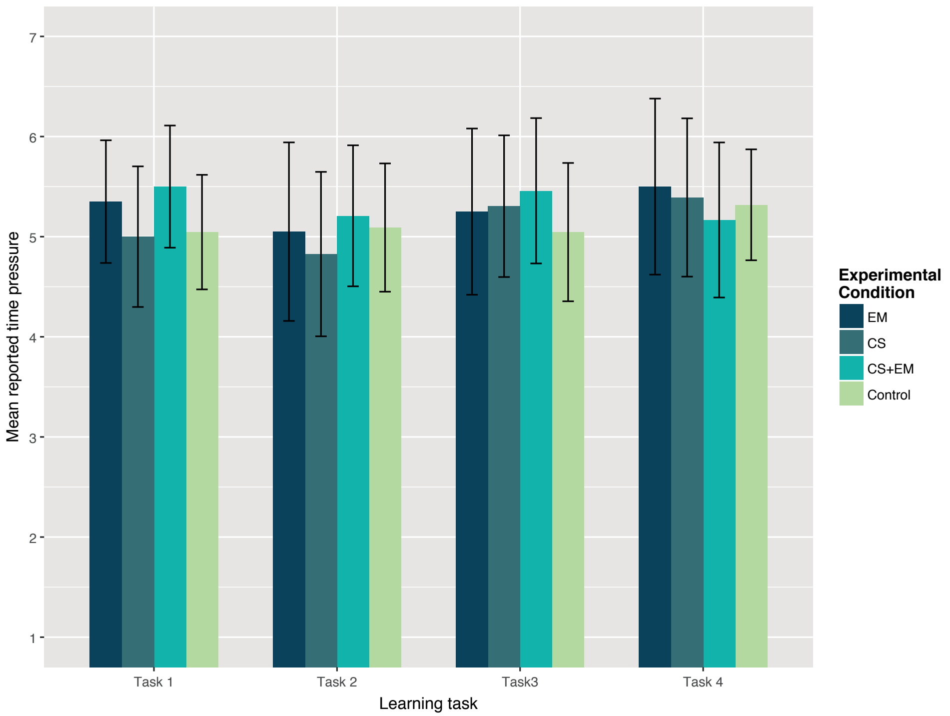


Figure 2.4. Reported time pressure per learning task for all conditions

Discussion

This experiment was designed to explore whether the acquisition of IPS skills was affected by different forms of task support. However, the results show that all groups show similar increases in skill. These findings do not provide support for the hypotheses that 1) supported students show higher learning outcomes than unsupported students, and 2) two forms of support lead to higher learning outcomes than just one form of support. As a matter of fact, the control group, which merely received conventional tasks without any built-in support, performed just as well as the three groups who received task support. There was a significant increase in scores from pretest to posttest for all conditions, showing that the intervention clearly caused a learning effect. From this finding, it can be concluded that even a short online training, much like the training sessions often offered by schools, can be effective for fostering IPS skills. While the results clearly show a short-term learning effect, it is unclear whether there is potential to achieve a long-term effect. Additionally, the different types of support might have different effects on retention, which only manifest when measured after sufficient delay, or are induced by testing (i.e., a testing effect: Dirkx, Kester, & Kirschner, 2014). No such delayed measurement was undertaken in this experiment. It would therefore be interesting to investigate delayed learning effect with a delayed posttest.

Furthermore, students reported a similar amount of required mental effort in all conditions. For these students and in this particular setting, online learning tasks with or without built-in task support, whether that is completed steps or emphasized aspects, are equally demanding in terms of mental effort. From this self-report of mental effort, it is only possible to gauge the total amount of experienced cognitive demand, but not changes in the underlying types of cognitive load. If worked-out steps reduced intrinsic cognitive load but required students to invest additional mental effort to process and self-explain the worked-out steps, it replaced intrinsic with germane cognitive load and there might be no change in the total amount of experienced cognitive load (Paas, Tuovinen, Tabbers, & van Gerven, 2003). Similarly, if prompting leads to more extraneous cognitive load and less invested energy in learning, germane load is reduced but the total amount of cognitive load remains the same. However, when intrinsic or extraneous load is replaced with germane cognitive load, this hypothetically leads to increased learning (Van Merriënboer & Ayres, 2005). It is unlikely that this has happened, because increased learning would manifest as higher scores on the skills tests, which were not found. From the current data, the only valid conclusion is that the different types of support have no effect on the total amount of experienced cognitive load as measured by reported mental effort.

The high scores on the time pressure item revealed that many students experienced time pressure to finish the experiment. In the short evaluation at the end of the training 43 out of 96 participants made a remark about experienced time pressure. From their comments, it became clear that the lack of time affected their concentration and performance during the learning tasks, or answer quality on the posttest. These students reported they took less time to think about and formulate their answers, thereby perhaps leaving out parts of the reasoning and missing points. This makes it likely that the learning outcomes are affected and possibly lowered because of time pressure. Given more time per task, students would perhaps have scored differently.

Inspection of students' solutions on the learning tasks revealed a great variation in answers. However, there was little instruction on presenting a solution incorporated in the training, so it cannot be expected that these outcomes correspond strongly to the level of their searching skills. Performance on the learning tasks was not part of the experimental design, and therefore, students' products were not scored and analyzed. For this reason, it is not possible to comment on the students' performance during the learning phase.

Experiment 2

A second experiment was conducted with the same goal as the first experiment: to investigate differences in learning outcomes due to different types of task support. The same design and conditions were used as in the first experiment, but an additional questionnaire was used and a delayed posttest was added. Some procedures were adapted to reduce time pressure.

Method

Participants

A total of 115 students between 18 and 46 years old participated in the replication (M_{age} = 20.7 years), 82 of which were female (71.3%) and 33 male (28.7%). These were all first-year Psychology students at a Dutch university. Of these 115 students, three had a Belgian nationality (2.6%) and 48 were German (41.7%). The remainder was Dutch.

Materials

Measurement of IPS skill

The same pretest and posttest were used as in Experiment 1, but a delayed posttest was added. This delayed posttest was identical to the existing pretest and posttest, but handled the topic of health benefits of red wine. Furthermore, a self-report questionnaire was added to the pretest, posttest and delayed posttest.

Self-report questionnaire

The self-report questionnaire was based on an existing questionnaire (Van Meeuwen, 2008) and contained 30 items to measure students' systematic approach and evaluation behavior; for example: "I check whether a page is up-to-date before I use its information". Students responded to these items by selecting 'Never', 'Sometimes', 'Often', or 'Always'. The questionnaire included an 'I don't know' option to reduce guessing.

Data analysis

The pretest, posttest, and delayed posttest were scored as in Experiment 1 and subjected to a repeated measures analysis of variance with *type of support* (CS+EM vs. CS vs. EM vs. control) as a between-subjects variable and *time of test* (pretest vs. posttest vs. delayed posttest) as a within-subjects variable. Mental effort and time pressure ratings were analyzed with a repeated measures analysis of variance with *learning task* as a within-subjects variable. In addition, a univariate analysis of variance was conducted on the *mental effort* and *time pressure* items per learning task to investigate differences in required mental effort between conditions.

For the self-report scale, a principle component analysis with oblimin rotation was conducted on the 30-item scale in a larger sample size (n = 250) to extract underlying clusters and form scales. A mean value was calculated for each cluster by averaging the scores on the corresponding items. The 'I don't know' answer was treated as a missing value, and averages were only calculated if there was no more than one missing value. Scores were analyzed with a repeated measures analysis of variance.

Procedure

As in Experiment 1, the training was embedded in the students' current curriculum as a practical assignment. Participation was voluntary, but strongly stimulated by granting research participation credit and informing students that the content of the training corresponded strongly to one of the course tasks about problem solving. The session was offered in eight different timeslots. Again, students were free to choose a timeslot that fit their schedule. Unlike in Experiment 1, the pretest was now administered in advance and was filled out at home, one week before the training. The delayed posttest was also filled out at home, one week after the training. The length of the training session remained two hours, which allowed students to spend approximately 20 minutes on each learning task; compared to 15 minutes in Experiment 1. Further procedures were identical to those in Experiment 1. After finishing the final evaluation, students signed a form to obtain research participation credit and were reminded to fill out the delayed posttest after one week. They were then dismissed. A debriefing followed in a lecture two weeks after the delayed posttest.

Results

As in Experiment 1, analysis of the answers on the preliminary questionnaire revealed a homogeneous group in terms of age and prior education. No notable differences arose in computer usage patterns or time spent behind the computer per day. Again, some data was missing, which was handled in the same way as in Experiment 1.

Pretest, posttest, and delayed posttest

The scores on the pretest ranged between 12.5% and 62.5% with a mean of 35.14% (SD = 11.18). For the posttest, scores ranged between 37.5% and 83.33% with a mean score of 61.58% (SD = 11.15). On the delayed posttest, the mean score was 60.6% (SD = 13.73) with a minimum score of 25% and a maximum score of 87.5%. Table 2.3 shows the mean scores per condition for the three tests. The results resemble those of the first experiment and show an increase in scores after training, but little difference between the conditions. The repeated measures analysis confirms that there was no significant difference between the groups: $F(3, 102) = 1.09$, $p = .358$ but a significant difference on the within-subjects factor: $F(2, 102) = 236.40$, $p < .001$, $\eta^2_{partial} = .699$. This confirms that there was a substantial effect of training on the test scores. A planned contrast revealed that the increase in scores from pretest to posttest was statistically significant: $F(1, 102) = 383.03$, $p < .001$, $\eta^2_{partial} = .790$, but the scores did not change significantly on the delayed posttest: $F(1, 102) = 0.72$, $p = .400$. There were no significant interaction effects.

Table 2.3. *Means and standard deviations of scores on the skills test (in percentages), systematic approach ratings (0–3), and evaluation behavior ratings (0–3) per condition on the pretest, posttest, and delayed posttest*

Condition		Pretest	Posttest	Delayed posttest
EM	Score	34.72 (11.93)	58.33 (11.88)	58.80 (12.41)
	Systematic	1.22 (.36)	1.28 (.38)	1.39 (.43)
	Evaluation	1.57 (.42)	1.76 (.33)	1.89 (.44)
CS	Score	34.25 (8.86)	63.58 (9.14)	60.50 (14.96)
	Systematic	1.08 (.41)	1.22 (.43)	1.32 (.45)
	Evaluation	1.53 (.47)	1.78 (.57)	1.94 (.42)
CS+EM	Score	36.22 (12.63)	63.06 (12.20)	64.90 (13.93)
	Systematic	1.20 (.44)	1.41 (.41)	1.32 (.40)
	Evaluation	1.74 (.50)	1.87 (.47)	1.93 (.46)
Control	Score	35.27 (12.41)	61.09 (10.79)	57.14 (14.00)
	Systematic	1.24 (.43)	1.34 (.39)	1.33 (.49)
	Evaluation	1.52 (.47)	1.95 (.35)	1.92 (.42)
Total	Score	35.14 (11.18)	61.58 (11.15)	60.60 (13.73)
	Systematic	1.19 (.39)	1.32 (.39)	1.34 (.42)
	Evaluation	1.57 (.48)	1.83 (.46)	1.93 (.42)

Self-report questionnaire

The Kaiser-Meyer-Olkin measure and sphericity measure indicated adequate sampling and sufficient correlations between items: KMO = .789, χ^2 (435) = 1544.54, p = .000. An initial analysis of eigenvalues and interpretation of the scree plot justified retaining two components for the final analysis. Table 2.4 shows the factor loadings and correlations after rotation. These loadings create two clusters that can be labeled as *systematic approach* and *source evaluation behavior*. Six items were discarded: four with both loadings below .32 and two with equal factor loadings on both components (Tabachnick & Fidell, 2007). The scales yielded reliability scores of α = .85 and α = .62 respectively. See Table 2.3 for an overview of means and standard deviations for both variables.

For the *systematic approach* data, Mauchly's test revealed that the assumption of sphericity had been violated, $\chi^2(2) = 21.19$, $p = .000$. Therefore, the Huynh-Feldt correction was applied to the degrees of freedom. The test showed a significant increase in scores: $F(1.74, 99) = 13.58$, $p = .000$, but a small effect: $\eta^2_{partial} = .125$. Subsequent contrast analysis showed that scores increased significantly from pretest to posttest: $F(1, 99) = 16.78$, $p = .000$, $\eta^2_{partial} = .150$, but did not change significantly on the delayed posttest. There were no significant differences between conditions: $F(3, 99) = .40$, $p = .756$. For the *evaluation behavior* data, the Huynh-Feldt adjustment was necessary as well: $\chi^2(2) = 8.40$, $p = .015$. Results show a significant within-subjects effect: $F(1.93, 94) = 32.98$, $p = .000$, $\eta^2_{partial} = .268$, but no significant between-subjects effect: $F(3, 94) = .44$, $p = .726$. Contrast analysis shows a strong increase in scores from pretest to posttest: $F(1, 94) = 38.79$, $p = .000$, $\eta^2_{partial} = .301$, and another small increase on the posttest. The latter just fails to reach significance: $F(1, 94) = 3.90$, $p = .051$, $\eta^2_{partial} = .024$.

Table 2.4. *Exploratory factor analysis results for the IPS self-report: factor loadings (correlations)*

Items	Systematic approach	Evaluation behavior
I work according to a predetermined plan when searching, selecting, and processing information	**.75** (.72)	-.12 (.09)
I make an overview (a list or table) of the needed information	**.72** (.68)	-.17 (.03)
I plan where I am going to search for which information	**.67** (.61)	-.23 (-.05)
I make a list of steps to follow	**.67** (.62)	-.19 (-.00)
I mostly work intuitively and do not use a predetermined plan	**.66** (.65)	-.03 (.16)
I make an overview of possible keywords	**.61** (.58)	-.12 (.05)
I just search for information without thinking about it too much	**.58** (.57)	-.05 (.11)
I make a time schedule for performing the task	**.57** (.56)	-.06 (.10)
I systematically keep track of the keywords I have used	**.51** (.52)	.04 (.18)
I regularly check whether I am searching correctly	**.46** (.49)	.12 (.25)
While searching, I try to keep an overview of the search process	**.45** (.47)	.08 (.21)
I deliberately check what I do not know yet in relation to the task	**.43** (.50)	.24 (.26)
I present the information in an organized and ordered fashion	**.42** (.47)	.16 (.28)
After visiting a site, I check which information is still needed	**.41** (.43)	.06 (.17)
At the end, I check again whether I have all the information	**.39** (.45)	.24 (.35)
I mostly work on and see how far I get	**.36** (.45)	.31 (.41)
I make sure that I organize all relevant information well	**.35** (.42)	.26 (.36)
I keep the desired end product in mind	**.33** (.40)	.25 (.34)
By looking at the URL (Uniform Resource Locator) I can see if a site is reliable	-.23 (-.06)	**.63** (.56)
To decide which site to open, I look at the URL (Uniform Resource Locator)	-.13 (.04)	**.62** (.58)
I check whether the site is up-to-date before I use the information	.01 (.17)	**.55** (.56)
I check whether information I have found overlaps with previously found information	-.05 (.10)	**.52** (.51)
Before I open a site, I check its reliability	.11(.24)	**.49** (.52)
I check whether information I have found contradicts previously found information	.09 (.22)	**.47** (.50)

Mental effort ratings

The experienced mental effort during learning tasks shows a significant within-subjects effect: $F(3, 88) = 8.31$, $p = .000$, $\eta^2_{partial} = .090$, indicating that scores change significantly over time. However, a significant interaction effect reveals that the effect depends on the type of support the student received: $F(9, 88) = 2.74$, $p = .005$, $\eta^2_{partial} = .089$. Separate repeated measures ANOVAs for each condition showed significant effects only in the EM condition: $F(3, 18) = .5.50$, $p = .002$, $\eta^2_{partial} = .244$, and in the CS condition: $F(3, 23) = 7.76$, $p = .000$, $\eta^2_{partial} = .261$. Subsequent contrast analysis indicated that the mental effort ratings in these groups only changed significantly on the fourth learning task. In the EM condition, scores dropped from 4.72 ($SD = 2.16$) on task 3 to 3.17 ($SD = 2.01$) on task 4: $F(1, 18) = 7.84$, $p = .012$, $\eta^2_{partial} = .316$. In the CS condition, scores dropped from 4.74 ($SD = 2.34$) to 2.83 ($SD = 1.64$): $F(1, 23) = 12.88$, $p = .002$, $\eta^2_{partial} = .369$. Figure 2.5 shows mental effort ratings for each condition and each learning task.

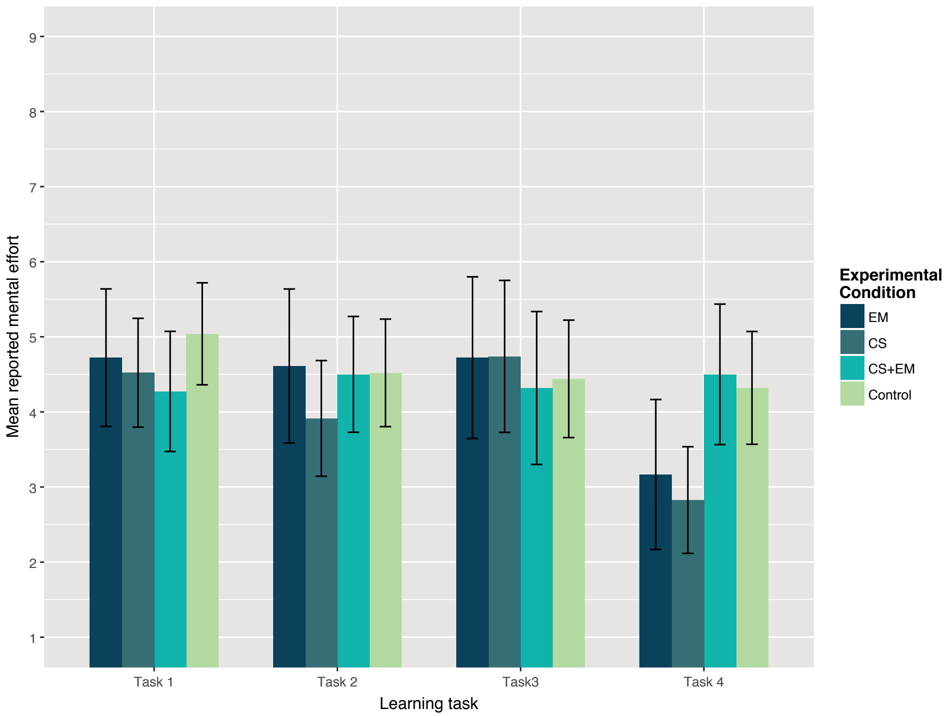


Figure 2.5. Reported mental effort per learning task for all conditions

Time pressure ratings

Analysis of time pressure ratings revealed no significant changes over time and no differences between conditions. Separate univariate ANOVAs for each learning task showed that the average amount of time pressure on each learning task was the same in each group. Figure 2.6 shows time pressure ratings for each condition and each learning task.

Discussion

The second experiment replicated the first with some improvements. First, it measured additional variables with a self-report questionnaire to achieve a more complete impression of the students' skill level. Second, it set out to reduce the experienced time pressure by administering the pretest before the training session. And finally, it included a delayed posttest to measure IPS skill one week after training. With these improvements, the findings display a similar pattern as in the first experiment. The significant increase in scores from pretest to posttest leads to the conclusion that the intervention was effective for fostering IPS skills. However, the results do not back the claim that the type of support has an effect on the learning outcomes. None of the groups that received support, whether completion strategy, emphasis manipulation, or both, outperformed the control group.

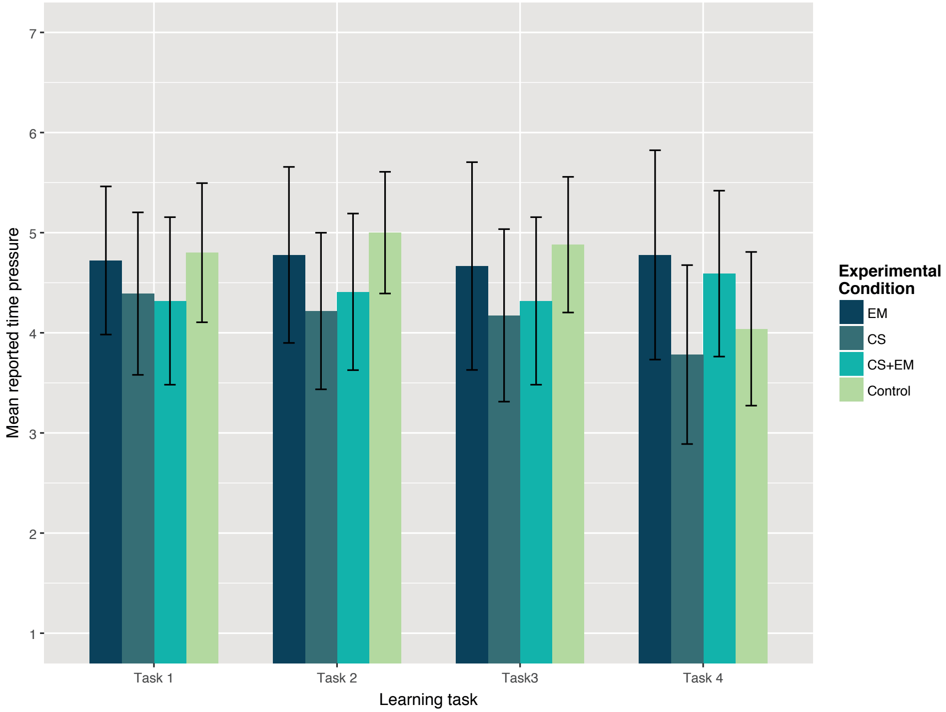


Figure 2.6. Reported time pressure per learning task for all conditions

This was also true for scores the self-report questionnaires. For systematic approach, students scored around 1.19 on the pretest, a value closer to 'Sometimes' than to 'Often', indicating that students are aware that they do not work very systematically when solving information problems. This score showed a small increase to an average of 1.32 on the posttest. While statistically significant, the effect of the training is small, and type of support again showed no effect. For evaluation behavior, a similar pattern emerges, but with larger effects. Average scores increase from 1.59 before training to 1.84 after the training, showing a large effect size. From these results, it can concluded that the training significantly improved students' scores on self-reported systematic approach and source evaluation behavior, but again, there were no significant differences between the conditions. This corroborates previous research that shows evaluation skills can be trained in classroom settings (Britt & Aglinskas, 2002; Walraven et al., 2010)

In general, scores on the delayed posttest results show a similar picture for all measured variables. While they increase from pretest to posttest, they do not change much one week later. All the differences between posttest scores and delayed posttest scores are statistically insignificant and show small effect sizes. However, some conditions show a small increase in scores after a week, while others show a decrease in scores. It would be interesting to see if this difference develops into a significant effect over a longer period of time. From these findings, it can be concluded that the learning effect caused by this intervention is sufficiently robust to last one week.

Compared to Experiment 1, the mean reported mental effort and time pressure is generally lower. This is an expected finding as students in Experiment 2 were given more time to perform the learning tasks. On the fourth learning task – a conventional problem without support or guidance – the CS and the EM conditions reported significantly less mental effort than the CS+EM and control conditions. This might be a hint that these students have become more efficient in their problem solving and require less mental effort to reach the solution. However, without performance data on the learning task, this is impossible to determine (Hoffman & Schraw, 2010). The subsequent posttest did not show any differences in performance between the conditions.

In short, Experiment 2 yielded no support for the hypothesis that supported students show higher learning outcomes than students who receive no support. Mean scores in all conditions did not differ significantly. While the EM and CS conditions reported less mental effort on a conventional learning task at the end of training, it is difficult to draw any solid conclusions from this finding.

General discussion

The experiments reported on here investigated the hypothesis that students who receive task support while acquiring IPS skills, either in the form of completion tasks or emphasis prompts, show better learning outcomes than students who do not receive task support. The findings do not support this claim. Students who receive no task support performed just as well as those who did. While Experiment 1 suffered from some methodological issues, a revised version of the experiment confirms the pattern of results and provides more confidence in this conclusion.

These findings have some implications for the domain of information problem solving. The current experiment once again confirms that IPS skills are underdeveloped, even in university-level students. Pretest scores are low in both experiments. In fact, the slightly younger group of students in the first experiment scored higher on the pretest than their counterparts in the second experiment. This difference shows a discrepancy in prior knowledge between both samples. While the exact cause of this is unclear, these differences likely originate from prior experience, practice, or instructions concerning IPS skills, such as a library training. However, the most important conclusion to draw from these findings is that this generation of first-year university students do not show very well-developed IPS skills. The scores, which lie well in the lower half of the range, can only refute claims that students are 'digital natives', a new generation technologically skilled students in need of adapted education. These findings agree with research challenging the existence of the digital native (Bennett, Maton, & Kervin, 2008; Kirschner & van Merriënboer, 2013; Smith, 2012) and underline our claim that IPS instruction in schools is a necessity.

The good news is that the current experiments show that a short online intervention can increase IPS skills. The online training session was successful, as shown by the significant increase in scores between pretest and posttest. After the training, students from both experiments scored slightly over 60% on average, which leads to two conclusions. First, a two-hour online training including an instructional video, a modeling example, and four short whole tasks can increase students' IPS skills. As shown by delayed posttest scores, this increase is maintained for at least a week. Second, effect sizes are not very large, and a 60% average score after training indicates that there is still much room to grow. However, the encouraging result of this short training indicates that a scaled-up version with more content, more task classes containing tasks of increasing complexity, offered over a longer period of time and embedded in a multitude of contexts, might prove very effective.

The findings of these experiments also lead to implications for the field of instructional design. Concerning the effect of built-in task support, the hypothesis that task support would lead to better performance was not confirmed: students who received no support showed performance equal to that of supported students. There are two possible explanations for this. First, it might be the case that both forms of support were ineffective for different reasons. Previous research has shown that completion tasks can lead to an expertise reversal effect in situations where learners have high prior knowledge (Kalyuga et al., 2003). However, this effect is less likely to occur in less structured domains (Nievelstein, van Gog, van Dijck, & Boshuizen, 2013), which, in combination with the low pretest scores, makes it unlikely that an expertise reversal effect occurred. The other method of support, prompting, can be ineffective when prompts are not used as intended, in which case they show reduced effects on learning outcomes and reported mental effort (Bannert & Reimann, 2011). Although answers to the prompts were generally short (i.e., approximately one sentence), they indicated that the prompts were used as anticipated. These findings lead to the conclusion that the task support methods were implemented correctly.

The second explanation suggests that a maximum learning effect for this setting was achieved. It could be the case that the learning effect in this experiment can be attributed to the viewing of the instructional video in combination with the modeling example including prompts. Modeling examples are very powerful learning tools when employed correctly (Bjerrum, Hilberg, van Gog, Charles, & Eika, 2013; Hoogerheide, Loyens, & van Gog, 2014b). Perhaps, after viewing both videos, students were sufficiently equipped to complete the learning tasks, and had no need of support. It follows then that the built-in support in those learning tasks had little value. A video-based modeling example is intuitively a very suitable method of instruction for teaching these skills, as most of the IPS process happens on-screen. An expert can easily record a screencast while working and reasoning through a problem and offer this as an example to students. The effects of using a modeling example for teaching IPS skills presents an interesting venue for future research.

In the context of this short online training, task support did not lead to higher learning outcomes. However, Rosman et al. (2016a) show that working memory capacity moderates the acquisition of IPS skills. In a holistic approach to learning IPS, task support, such as completion tasks or prompts for emphasis, is essential to avoid overloading the learner during complex task performance. However, in situations where the learner's skill level is sufficient or the tasks are less complex, it might have no beneficial effects. Therefore, the results of these experiments should not convince instructional designers that task support does not matter. Instead, it should stimulate them to seek closer alignment of the learner's skill level, task complexity, and built-in task support. When designing instruction for IPS on a larger scale and over a longer period of time with increasing levels of complexity, managing the cognitive load imposed on the learner remains a crucial aspect of instructional design.

Several limitations of these experiments should be regarded when interpreting and generalizing these conclusions. The IPS training was offered in a single two-hour session with learning tasks of the same type and complexity. In educational practice, students are confronted with a great variety of tasks. The current intervention did not include different task types (c.f. Gerjets & Hellenthal-Schorr, 2008), which makes it less likely that far transfer occurred. To achieve far transfer, students would benefit from more learning tasks: more practice with varying task demands and task complexity, yet without added time pressure. An embedded approach, where instructional designers combine IPS instruction with domain-specific instruction in an extensive curriculum, appears appropriate for this task (Argelagós & Pifarré, 2012; Wopereis, Brand-Gruwel, & Vermetten, 2008).

The current intervention focused on learning in an online environment without involvement of a teacher and without feedback on the learning tasks. Considering the multitude of factors that can increase task complexity and cognitive demand, personalized feedback on performance would be beneficial for students, as this allows them to learn from their mistakes. Research has shown a positive effect of feedback on development of metacognitive skills in online learning environments (Van den Boom, Paas, van Merriënboer, & van Gog, 2004) and therefore presents another interesting direction for future research. For example, interventions could be improved with the addition of a cognitive feedback element in which teachers provide students with adapted feedback on their performance (Timmers et al., 2015; Wopereis et al., 2015).

To conclude, this experiment makes clear that first-year university students are not as information literate as many assume, and that their IPS skills need to be trained. The 2-hour online intervention in this experiment shows a promising learning effect. While it was expected that different types of task support would vary in their effect on the learning outcomes, this proved not to be the case. The authors hypothesize that a powerful modeling example is most likely responsible for a large proportion of the learning effect and increased students' skill level, thereby reducing the value of the task support in the subsequent learning tasks. A follow-up study will investigate whether modeling examples are indeed a powerful learning tool for IPS.

Chapter 3

Effects of a modeling example for teaching information problem solving skills

Abstract

While students often appear to be skilled in retrieving and making use of information from the Internet, research shows that their information problem solving skills are overestimated. They show deficiencies in many of the necessary skills, such as generation of search terms, selection of sources, and critical processing of information. It is therefore necessary to design and develop effective instruction to foster information problem solving skills. Research shows that learning from examples can be an effective approach for teaching complex cognitive skills in ill-structured domains, such as writing or communicating. To explore whether this also holds for information problem solving, the current study investigates the effects of presenting a modeling example in an online information problem solving training. Results of two experiments show that viewing a modeling example, presented as a screencast of an expert thinking out loud and interspersed with cognitive prompts, leads to a higher posttest performance than performing a practice task. The effect persisted on a delayed posttest one week later. The results imply that information problem solving instruction in an online setting can benefit from employing video-based modeling examples.

This chapter is based on:

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2017). Effects of a modeling example for teaching information problem solving skills. *Manuscript submitted for publication.*

Introduction

Information problem solving (IPS) is a skill often required from students in today's educational programs, as it is common for teachers to provide assignments requiring students to search for information on the Internet. These assignments can be characterized as information problems: problems that require more information to solve than is currently available to the learner. They pose an information gap, because students must first search for the missing information and then process it in order to solve problem. Teachers might assume that searching and processing information automatically leads to learning, but such information problems are often ill-defined and present unknown or unclear task demands, goals, or solution paths. While it is tempting to regard students as 'digital natives' and expect that they automatically acquired skills to solve such problems, research shows that most students' information problem solving skills are underdeveloped. Students struggle to systematically search for information, evaluate it critically, and produce an adequate solution for an information problem (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2016; Walraven, Brand-Gruwel, & Boshuizen, 2008, 2009).

An effective approach to solving an information problem can be summarized in five steps (e.g., the IPS-I model; Brand-Gruwel, Wopereis, & Vermetten, 2005; Brand-Gruwel, Wopereis, & Walraven, 2009). First, learners build a problem representation by reviewing the task demands, activating prior knowledge on the topic, and identifying which information is needed. They form an idea of the extent and structure of the domain, and formulate a question. Research shows that this step is often neglected entirely or performed only partially (Brand-Gruwel et al., 2005). In the second step, learners determine a search strategy and start searching for information sources. In the case of an online search, they generate search terms, execute the search query in a search engine, and evaluate the search engine results page. Here, learners use ineffective strategies (Hölscher & Strube, 2000; Van Deursen & van Diepen, 2013), have problems generating relevant search terms, and formulate unproductive queries (Zhou, 2013). The third step is often executed in parallel with the second step and involves the evaluation of information and sources. In this step, learners determine whether a source is relevant, recent, and credible. This kind of critical scrutiny is essential to avoid irrelevant and unreliable sources, yet it is often lacking (Fogg et al., 2003; Gerjets, Kammerer, & Werner, 2011; Keil & Kominsky, 2013). The sources that make it through the selection process are processed in the fourth step. In this step, learners are often seen making annotations, highlights, or summaries as they critically study the contents to find similarities and differences between the sources. Research shows novices spend less time on processing the source contents than experts (Brand-Gruwel et al., 2005). In the final step, learners create a product such as an essay, presentation, or poster that integrates information from the sources in order to solve the information problem and answer the question. During these steps, it is important that learners regulate their process by monitoring their progress, gauging the needed information, and steering the process if necessary. Again, research shows novices monitor and steer their process less often than experts and pay little attention to task time constraints (Brand-Gruwel et al., 2005; Zhou, 2013).

The many problems that researchers and educators have discovered indicate that there is a need for formal instruction on IPS in schools in order to foster and improve students' IPS skills. However, as Badke (2010) illustrates, information literacy and IPS instruction is often lacking or not implemented effectively in educational programs, for a variety of reasons. Teachers may lack the necessary digital and IPS skills themselves and cannot teach them to their students, or hold a misplaced belief that such skills do not need to be trained because they develop naturally (Kirschner & van Merriënboer, 2013). And teachers who are equipped with the skills and willing to teach them they may be unaware of how to provide effective instruction and integrate it in their lessons. Reports investigating the Dutch educational context underline that there is little structural attention for the integration of digital skills, and little is known about effective implementation in practice (Platform Onderwijs2032, 2016; Thijs, Fisser, & van der Hoeven, 2014). From these findings, it becomes clear that there is a need for empirically tested instructional interventions and best practices to guide teachers and instructional designers. Fortunately, research on IPS instruction is now growing.

As a relevant example, a study by Frerejean et al. (2016) presented students with a standalone online training session, comprising an instructional video, a modeling example, and four learning tasks presenting an information problem. The study investigated which type of task support was most effective to teach IPS skills when using a whole-task approach. While the significant increase in performance after the training indicated the training was effective, no differential effects of task support (i.e., emphasis manipulation or completion strategy) were found. The authors suggested that the effectiveness of the modeling example used in all conditions could be partly responsible for the learning effect. Their findings suggest that providing demonstrations of effective IPS by experts can be an effective instructional method for teaching novices how to approach and solve information problems and they identify example-based learning in the domain of IPS as an interesting direction for future research.

Example-based learning

Example-based learning finds support in disciplines such as Bandura's *social learning theory* (Bandura, 1977) and cognitive load theory. From the perspective of Bandura's social learning theory, skills learning takes place by observing others perform the skill. Observational learning can be realized by presenting learners with modeling examples, typically showing a model performing the skill while thinking out-loud and, in contrast to traditional paper-based worked examples, providing important insight into the thought-processes and decision-making processes that otherwise remain covert. In the social learning account of observational learning, Bandura (1977) posits four processes that govern learning: attentional, retention, reproduction, and motivational. In order for modeling examples to be effective, a learner's *attention* should be focused on the essential features of the model, the actions should be stored in memory so they are *retained* and not forgotten, there should be opportunity for *reproduction* to practice the skills, and the learner should be *motivated* to display the correct behavior. Similar processes are identified in the *cognitive load* perspective on example-based learning.

In *cognitive load theory*, the *worked example effect* states that learning from fully and/or partially worked examples leads to more effective and/or efficient learning than conventional problem solving, as novice learners often lack the specific knowledge and problem-solving strategies necessary to solve problems without support (Sweller, 2006). Consequently, they mostly fall back to naïve strategies such as trial-and-error or means-ends analysis, which place a high demand on working memory and leave few mental resources to devote to learning (Sweller, 1988). Short on working memory capacity, novice learners focus primarily on irrelevant problem features and build a superficial representation of the problem. Experts, on the other hand, identify structural problem features, such as relevant domain principles to create a more elaborate problem representation (Chi, Feltovich, & Glaser, 1981; Sarsfield, 2014).

A provided worked example traditionally contains an initial problem state, a goal state, and a written account of the solution steps leading to a solution, such as a step-by-step description to solve a mathematics problem. Providing a worked example that shows the solution steps toward the goal state relieves learners of the search for a solution path and reduces the burden on working memory (Renkl, Hilbert, & Schworm, 2009). It provides an example of the correct procedure to solve a problem, which frees up cognitive resources to use activities that are germane to the construction of knowledge schema and solution procedures. Van Gog, Paas, and van Merriënboer (2004) argue that examples can be improved by providing not just a step-by-step process, but also process-oriented information. Elaborating on the rationale behind the problem solving process – the "how" and "why" - can enhance the transfer of these skills to other problem contexts (Van Gog, Paas, & van Merriënboer, 2008). In reviews on the effectiveness of example-based learning, Van Gog and Rummel (2010) and Renkl (2014) give an overview of the parallels between the social learning and the cognitive load accounts. Providing an exhaustive discussion on these accounts is outside the scope of this chapter, and therefore this section will focus on some of the factors affecting the effectiveness of examples that are relevant for the presented study.

Firstly, effectiveness of example-based learning depends on the degree to which the information in the example is processed. Learning from a model is improved when learners actively process the example by elaborating on the presented information and evaluating the process (Braaksma, Rijlaarsdam, van den Bergh, & van Hout-Wolters, 2006; Braaksma, van den Bergh, Rijlaarsdam, & Couzijn, 2001). Without performing these essential activities, learners might observe without trying to understand. Their attention might be diverted away from the relevant information or focused on less important elements in the example, with the risk of a decreased learning effect (Renkl, 1999; Stark, Mandl, Gruber, & Renkl, 2002). This can be overcome by directing learners' attention to important elements and to ensure active processing of the modeling example, for example with self-explanation prompts (Renkl, 2002; Renkl & Atkinson, 2002). Such prompts are considered an integral part of example-based learning from a cognitive load perspective (see Renkl et al., 2009), and are widely considered as an effective learning mechanism (Chi, Bassok, & Lewis, 1989; Chi, De Leeuw, Chiu, & Lavancher, 1994).

As answering such prompts requires that learners pay attention to the example and attempt to follow the solution procedure (Aleven & Koedinger, 2002), it reduces the possibility that they passively watch the example without cognitive investment (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 1997). Other types of prompts can have additional benefits, such as metacognitive prompting to stimulate metacognitive thinking (Stadtler & Bromme, 2008), or comparison prompts, asking the learner to compare and contrast their own approach to an expert's systematic approach. The latter is particularly beneficial if the learner starts out with intuitive strategies that are less effective, as such prompts can stimulate learners to think critically about the problem domain, the problem structure, and the demonstrated approach to problem solving (Van Merriënboer & Kirschner, 2013).

Secondly, example-based learning is effective if retention is ensured and the learner is able to remember and apply the observed skills in situations where the model is no longer present (Bandura, 1971). Enactment, or practice, is necessary for strengthening and automating the required skills without the presence of the model. Worked examples and modeling examples contain a high degree of guidance and support, which is beneficial for novice learners who lack domain knowledge and solution strategies. However, when progressing through the learning phase, schemas become increasingly more elaborate and more strategies are formed to cope with varying problem situations (Atkinson, Renkl, & Merrill, 2003). At some point, examples will offer little new information and much redundant information. When this occurs, learning from examples can lose its benefit over solving practice problems and induce an expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Rikers, & Paas, 2012) where providing too much support to advanced learners can be detrimental to learning. From this, it follows that examples should precede a period of practice, where learners get a chance to apply the observed knowledge and skills. This improves retention and avoids diminishing learning effects caused by the expertise reversal effect.

These findings dictate that examples are most effective for novice learners and when presented with incorporated methods to stimulate active processing, such as prompting. In addition, learners should be able to practice the observed skills after watching the examples in order to promote retention.

Instructional design for complex learning

Complex learning is defined as "the integration of knowledge, skills and attitudes; coordinating qualitatively different constituent skills; and often transferring what was learned in school or training to daily life and work" (Kirschner & van Merriënboer, 2009, p.244). The Four Component Instructional Design (4C/ID) model provides an extensive blueprint and approach for developing instruction to achieve complex learning, based on solid psychological and educational research (Van Merriënboer & Kirschner, 2013). First, the model advocates the use of authentic, whole tasks that require integration of knowledge, skills and attitudes, and coordination of constituent skills. Second, it provides guidelines to correctly provide the information needed to solve the problems: domain knowledge and a structured approach to solve the problem. Third, it advises providing just-in-time procedural information during the tasks to aid problem-solvers with routine tasks. The fourth component, part-task practice, is necessary when performance of these routine tasks needs to be automated.

This task-centered approach confronts learners with a series of whole tasks in the learner's zone of proximal development. Task complexity increases to keep up with learner progress. However, especially in the early phases of learning, tasks can be too complex for the learner, because they introduce too many interacting elements or the learner's knowledge schema are insufficiently developed. In these cases, the learner's memory system may become overloaded, which can negatively impact learning (Paas & van Merriënboer, 1994). In situations of complex learning and authentic tasks, there are many elements that potentially increase the amount of cognitive load experienced by the student. It is therefore essential that instructional designers take great care to reduce unnecessary load, while maintaining activities that induce germane load and lead to learning.

For IPS specifically, task complexity is not the only factor that influences the demands on working memory during problem solving, and in consequence, learning and instruction. Rouet (2009) summarizes additional factors in a conceptual framework comprising three dimensions: individual variables, information resources, and problem context. Instructional designers should be aware that personal factors, such as an individual's domain-specific knowledge (Monchaux, Amadieu, Chevalier, & Mariné, 2015), age (Chevalier, Dommes, & Marquié, 2015), attitudes and biases (Ford, Miller, & Moss, 2005; Van Strien, Brand-Gruwel, & Boshuizen, 2014), epistemic beliefs (Kammerer, Bråten, Gerjets, & Strømsø, 2012), and reading skills (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011) can affect the learning process and outcomes. Similarly, source factors (DeStefano & LeFevre, 2007) and task type (Wirth, Sommer, von Pape, & Karnowski, 2015) may influence variables in the learning process. While most of these factors lie outside the designer's influence, they all affect the demand imposed on working memory during the IPS process.

For situations where tasks may be too demanding for a learner to complete successfully, the problem-solving process must be supported (Van Merriënboer, 2013). The 4C/ID model stresses the importance of providing built-in task support, for example with case studies, modeling, or worked examples.

Examples in information problem solving instruction

Much of the research on example-based learning has taken place in structured, well-defined domains such as mathematics and physics where there are often fixed procedures for solving a problem, but some research exists on the effects of example-based learning in ill-defined domains. These problems cannot be solved by following a strict procedure with discrete solution steps; instead, learners will have to reason through the problem and make the right decisions relying on heuristics, strategies, and an evaluation of the currently available information. In these cases, a worked example showing only a step-by-step solution procedure is not sufficient, because learners need more information about how decisions are made and which knowledge is used to make these decisions. For ill-defined problems, modeling examples containing process information are preferred over traditional written worked examples to demonstrate how the problem solver reasons through the solution steps (Van Gog & Rummel, 2010). Seeing the solution steps being performed and hearing the reasoning behind them can help learners improve or create knowledge schemas and solution procedures.

Research has shown that modeling examples can be effectively employed to teach complex skills in unstructured domains. For example, novice psychotherapists improved their communication skills the most when watching a video of an experienced psychotherapist (Baum & Gray, 1992), students' problem-solving strategies increased by watching a teacher thinking aloud in a problem-based learning setting (Pedersen & Lui, 2003), and student designers delivered more creative work after watching videos of peers thinking aloud during a design task (Groenendijk, Janssen, Rijlaarsdam, & van den Bergh, 2013). While further research shows example-based learning is beneficial for the acquisition of complex skills, such as academic writing (Braaksma, Rijlaarsdam, van den Bergh, & van Hout-Wolters, 2004; Braaksma et al., 2001) and problem solving (Van Gog et al., 2004), no research was found investigating modeling examples in IPS instruction. From these findings, it can be expected that example-based learning in the form of modeling examples is also effective for teaching the complex skill of IPS.

In the context of IPS instruction, a modeling example might consist of an expert solving a problem while explaining the reasoning about each step and skill in the process. For example, the expert explains why a certain strategy is chosen, how search terms are generated and how the different results and sources are evaluated. A recorded screencast can show the screen and activities (i.e., clicking) of the expert while concurrently playing the expert's narration. Online learning environments provide an easy opportunity for embedding modeling examples in the form of video demonstrations. However, instructional designers should be aware that multimedia materials can easily create unwanted cognitive load, which carries the danger of impairing the learning process (Paas, Tuovinen, Tabbers, & van Gerven, 2003). It is therefore wise to follow the principles derived from research on multimedia learning to reduce hindering load on working memory and increase activities that lead to learning (Mayer, 2014).

The present study

The present experiment is a follow-up to the research by Frerejean et al. (2016) that investigated the effect of task support on acquiring IPS skills in a short online intervention. It follows up on the suggestion that modeling examples were responsible for the learning effect found in the former study, and attempts to answer the question: "What are the effects of providing a modeling example on the acquisition of IPS skills in a short online training?" Two experiments test the hypothesis that students receiving a modeling example display higher performance on an IPS test than students receiving no example, but engage in a practice task. To investigate whether and how viewing a modeling example also puts strain on working memory and affects the learning process, subjective mental effort ratings are collected during the learning phase.

Experiment 1

Method

Participants

A total of 39 first-year university students participated in the individual, computer-based online training session at a Belgian university (27 female, 12 male). All students had the Belgian nationality, and their age varied between 16 and 38 years old (M_{age} = 19.67, SD = 3.47). In the modeling example condition, 15 students were female, five were male, and the age varied between 17 and 38 years (M_{age} = 20.15, SD = 4.25). In the practice task condition, 12 students were female, seven were male, and the age varied between 16 and 22 (M_{age} = 19.16, SD = 1.8).

Design

The experiment was a pretest-posttest design with two conditions. All students received a two-hour online training in IPS, consisting of an instruction video, a modeling example for one half of the students or a practice task for the other half, and four learning tasks. Students' skill level was measured before and after the training.

Materials

Online training

The training was presented in an online learning environment and consisted of three elements. First, a 14-minute instructional video introduced a systematic approach to solving information problems, based on the IPS-I model (Brand-Gruwel et al., 2009): define the problem, search for information, select information, process information, and present the solution. The video was presented to provide students in a short amount of time the necessary domain knowledge and problem-solving approach to complete the upcoming learning tasks. Then, either the modeling example or the practice task was presented, further explained below. Finally, the students received four learning tasks, each consisting of a problem description and a textbox to enter an answer. Students had to search the web for information to reach a solution. The learning tasks contained no further support or guidance. The presented problem descriptions handled disputed socio-scientific topics: the effect of stretching before sports, the dangers of electromagnetic radiation from cell phones, the consequences of violence in videogames, and the influence of using media devices on sleep quality.

Modeling example

The modeling example was presented as a 10-minute screencast in which a fictitious expert demonstrated how to solve an information problem about the effect of GPS navigation systems on traffic safety. The model was a 23-year old Dutch female speaking in a standard-accented voice. While earlier research suggests a speaker/gender effect, stating that learning outcomes from video modeling examples are higher when narration is presented by a female speaker rather than a male speaker (e.g., Linek, Gerjets, & Scheiter, 2010), more recent research finds gender has no beneficial effects on learning outcomes, though it may influence affective aspects of learning (Hoogerheide, Loyens, & van Gog, 2015). The model in this example was not visible to the viewer, but narrated the actions on-screen by thinking aloud and explaining her reasoning behind each decision. The modeling example was split into four short fragments and interspersed with prompts. Before viewing each fragment, students first activated their prior knowledge by answering the prompt “Where will you focus your attention while executing the next step?” This prompt served as a method to activate the relevant principles and strategies pertaining to that step before the student watches the model.

The first fragment showed the expert reasoning about the problem description and generating a brief and clear problem statement. The fragment ended with a prompt that included the questions “What do you think of the actions of the expert?” and “How does this differ from your current approach?” These questions were intended to stimulate comparisons of solution procedures between the student and the expert and an active processing of the example. Students entered the answers to these questions in a textbox before clicking through to the next fragment.

The second fragment demonstrated how the expert chose search terms and entered them into the Google™ search engine. The subsequent results page was analyzed by thinking aloud while showing relevant on-screen elements with the cursor. The fragment ended with the questions “What do you think of the actions of the expert?”, “Would you have chosen the same keywords?”, and “Do you agree with the evaluation of the search results?”

The third fragment started with a short reflection by the expert on her reasons for selecting a particular website. These additional comments served as a feedback component, so students could compare their answers to the expert’s reasoning. The fragment continued with a demonstration on how to quickly scan and evaluate a source. The information in the source was deemed relevant and reliable and subsequently added to the bookmarks. The expert noted that the information was a bit outdated, so she returned to the search results to find a more recent source. After evaluating and saving a second source, the expert made some changes to the keywords and evaluated two additional pages. The fragment concluded with the following prompts: “What do you think of the actions of the expert?”, “Would you have done the same?”, and “What would you do differently?”

The final fragment showed the expert’s formulated answer to the problem. Students were advised to pause the video to read the answer in their own pace. Afterwards, they were prompted with the questions: “What do you think of the expert’s answer?”, and “Would you have given a similar answer?”

The screencast was a complete yet condensed application of the five steps of the systematic approach introduced in the instruction video. The design of the video followed several instructional design principles: Schema construction was promoted by activating the learner's knowledge prior to each fragment, and active processing was promoted by adding prompts after viewing each fragment. In addition, students were stimulated to compare the expert approach to their own. At the beginning of the third fragment, the modeling example included some general reflection remarks that serve as feedback on the students' answers to the prompts. In addition, care was taken to design the modeling example following principles for effective multimedia learning (Mayer, 2014). Appendix 2 gives an overview of these principles and how they were applied to the screencast. After the example, students received four learning tasks to practice the demonstrated approach. These aspects (prompting, segmenting, etc.) are considered integral parts of a well-designed modeling example and are therefore implemented and analyzed as one intervention.

Practice task

The practice task contained the same problem description as the modeling example and a textbox to enter an answer. Students were asked to spend approximately 15 minutes searching the web for information before formulating a short answer; comparable to the amount of time it took the other students to process the modeling example. After the practice task, students received the same four learning tasks for more practice.

Preliminary questionnaire

To collect demographic data, including age, nationality and prior education, a short questionnaire was administered before the pretest. The questionnaire also included items about the amount and pattern of Internet and computer usage. Students were asked to indicate their perceived level of competence in solving information problems on a scale of 1 to 10.

Measurement of information problem solving skill

IPS skill was measured using a situational judgment instrument developed by Frerejean et al. (2016). The online measure consisted of seven fabricated situations that occur during IPS and asked students to describe how they would act in the presented situation. Table 3.1 provides a schematic overview of the seven questions in these skills tests. To ensure content and face validity, the items correspond to the skills and subskills in the IPS model by Brand-Gruwel et al. (2009). For example, to measure the skill 'selecting information', a fabricated search engine results page (SERP) was presented and students were asked to select three results and give reasons for their selection. The answers were scored blindly using the scoring rubric in Appendix 1. Students could obtain four points for each of the four skills: defining the problem, searching information, selecting information, and processing information. The skill *presenting information* was not included in the tests for two reasons. First, presenting can be done in countless ways and concerns a multifaceted skill that is difficult to measure in a short timeframe. Second, the training presented little instruction on presenting information, so little improvement is expected. The four scores were then averaged to obtain the total test score and expressed in a percentage for ease of interpretation. The items on the posttest were identical to those on the pretest, but on a different problem domain. In the pretest gender-specific education was used as a problem domain, while the malleability of intelligence was used in the posttest. A second rater rescored 20 randomly selected cases to allow inter-rater reliability analysis. The two-way, mixed, absolute, single-measure intra-class correlation of .878 indicated a high inter-rater agreement and therefore a reliable measurement.

Table 3.1. *Items in the pretest and posttest*

Item	Skill	Subskill	Given	Question
1	Defining the problem	Problem orientation	A problem description	How would you start this task? What is your first step and why?
2	Defining the problem	Formulating a problem statement	A problem description	Which problem statements would you formulate? Why do you choose these?
3	Searching information	Generating search terms	A problem description	Which search query would you type into Google? Formulate two alternative search queries.
4	Selecting information	Evaluating search results	A fabricated SERP	Which three websites would you select? Why did you select these websites?
5	Processing information	Scanning a source	A screenshot of a text-rich website, zoomed so the text is unreadable	What do you do when you visit a text-rich website and want to find out if it contains relevant information? How do you proceed?
6	Processing information	Evaluating information	A short text fragment containing an argument given by an expert	Which criteria do you use to determine whether information is useful for your task? What are your conditions for use?
7	Processing information	Dealing with conflicting information	Two short, contradicting arguments	How do you deal with contradicting information? How does this affect your solution? Explain.

Mental effort

Solving an information problem is a complex task imposing a high cognitive demand, especially when the required skills are insufficiently developed. To investigate whether viewing a modeling example alters the experienced cognitive demand during practice, mental effort was measured four times during the training phase. At the end of each learning task, students answered the item: *How much effort did it take to perform this task?* on a 9-point scale (Paas, 1992).

Data analysis

An analysis of covariance was conducted on the posttest scores with *modeling example* (yes vs. no) as a between-subjects factor and the pretest score as a covariate. A repeated measures analysis of variance was conducted on the mental effort ratings, with *learning task* as a within-subjects variable and *modeling example* (yes vs. no) as a between-subjects variable.

Procedure

Students participated in the experiment as a practical assignment in their curriculum. The two-hour session took place in the university computer room where students received log-in credentials to access the online experimental environment. They received instructions to work individually through the tasks and to spend approximately 15 minutes on each learning task, which is a realistic time limit for finding information online (Lazonder, Biemans, & Wopereis, 2000). They were further informed that their screen content could differ from that of the other students and then presented with the preliminary questionnaire. After filling out the questionnaire, students were automatically randomly assigned to one of the two conditions. They then received the pretest and the instructional video. Half of the students received the modeling example and half received the practice task. Afterwards, students could practice the skills in four learning tasks, followed by the posttest. After the posttest, students signed a form to obtain course credit and were subsequently dismissed. A debriefing followed eight weeks later.

Results

Preliminary analysis

Analysis of the demographic data revealed no significant differences on any of the variables measured, such as age and computer use. Students reported they use the Internet for 2.72 hours per day (SD = 1.32) and estimated their own IPS ability with a 6.28 out of 10 (range = 4-8). A missing data analysis showed some incomplete data caused by students skipping questions and some incidental technical issues. For the pretest and posttest this meant that some scores that make up each of the four sub-scores were missing. In the cases where there was only one missing value, the posttest score was replaced with the corresponding pretest score. If more than one score was missing, the sub-score on that skill was classified as missing and the subsequent total average was calculated over the remaining three sub-scores. If more than one of the four sub-scores was missing, no total average was calculated and the test score was classified as missing.

Information problem solving skill tests

Both conditions obtained higher scores on the posttest than on the pretest, indicating that learning took place. As shown in Table 3.2, students receiving a modeling example scored higher on the posttest than those receiving a practice task. The difference between the conditions was statistically significant when controlling for pretest scores: $F(1, 39) = 5.64$, $p = .023$, $\eta^2_{partial} = .135$.

Table 3.2. *Mean scores (in percentages) and standard deviations on pretest and posttest*

Condition	Pretest (*SD*)	Posttest (*SD*)
Modeling example	38.44 (10.97)	55.94 (10.03)
Practice task	39.47 (11.23)	49.34 (6.88)

Mental effort

The ratings on the 9-point mental effort scale collected after each of the four learning tasks are displayed in Figure 3.1. The repeated measures analysis shows that reported mental effort changed significantly over time: $F(3, 37) = 3.01$, $p = .033$, but with a small effect size: $\eta^2_{partial} = .079$. Subsequent contrast analysis indicates only a significant decline from learning task 3 to 4: $F(1, 37) = 6.68$, $p = .014$, $\eta^2_{partial} = .160$. There was no difference between the conditions: $F(1, 39) = .22$, $p = .645$.

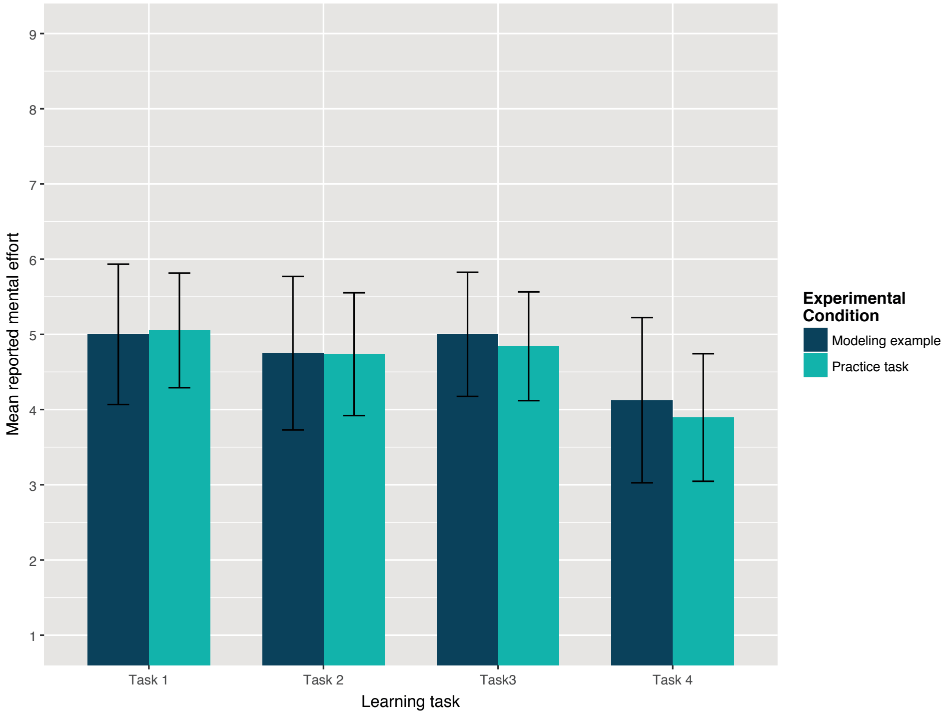


Figure 3.1. Reported mental effort per learning task for both conditions

Conclusion

These findings provide support for the hypothesis that students receiving a modeling example achieve higher learning outcomes than students receiving a practice task. While both groups show improved scores on the posttest, the group receiving a modeling example increased nearly twice as much as the practice task group. This shows a higher learning effect for students receiving a modeling example. While this provides an answer to the research question, it was attempted to replicate and further investigate the learning effect. Firstly, it would be interesting to investigate transfer over time. A delayed posttest would reveal if modeling examples have potential for robust learning. Furthermore, a larger sample size would increase confidence in the findings. For these reasons, a replication experiment was conducted in a slightly larger student sample and with an added delayed posttest.

Experiment 2

The design and materials were identical to those in the first experiment, but included a delayed posttest to measure the delayed learning effect. Additionally, the pretest was administered at home in the week before the training session, reducing possible priming effects on the learning phase.

Method

Participants

A total of 60 first-year Psychology students from a Dutch university participated in the replication (41 female, 19 male). Two students had the Belgian nationality, and 24 had the German nationality. The remainder was Dutch. The age ranged from 18 to 32 (M_{age} = 20.63, SD = 2.14). The modeling example condition contained 21 female students and eight male students with an age range of 18 through 24 years (M_{age} = 20.43, SD = 1.61), of which one was Belgian, 13 were German, and 17 were Dutch. The practice task condition consisted of 20 female students and 11 male students with an age between 18 and 32 years (M_{age} = 20.83, SD = 2.57), of which one was Belgian, 11 were German, and 17 were Dutch. Participation was voluntary, but strongly stimulated by granting research participation credit and informing students that the content of the training was relevant for the current topic in their curriculum (problem solving). Students could choose one of eight different timeslots. Furthermore, students were informed that an online pretest and delayed posttest had to be filled out in their own time. A debriefing followed in a lecture two weeks after the delayed posttest.

Materials

Measurement of information problem solving skill

A delayed posttest was added after the posttest. It was identical to the existing pre- and posttests, but handled the topic of health benefits of drinking red wine. The pretest and posttest were the same as in Experiment 1.

Data analysis

The pretest, posttest, and delayed posttest were scored as in Experiment 1. An analysis of covariance was conducted on the posttest scores with *modeling example* (yes vs. no) as a between-subjects factor and the pretest score as a covariate. This analysis was repeated on the delayed posttest scores. A repeated measures analysis of variance was conducted on the mental effort ratings, with *learning task* as a within-subjects variable and *modeling example* (yes vs. no) as a between-subjects variable.

Procedure

The procedure and design were largely identical to the procedure of the first experiment, with the exception that the pretest was filled out at home in the week before the training and the delayed posttest was filled out at home, one week after the training. Because it was known that a large proportion of students were German, the online environment was programmed to divide students in conditions on a random basis, yet to stratify for nationality. This was done as a precaution in case the German students' performance suffered because the materials were all in Dutch. This resulted in conditions containing approximately the same proportion of Dutch and German speaking students. Before starting the training session, the experimenter stimulated students to spend approximately 20 minutes on each learning task. After finishing the final evaluation, students signed an informed consent form and obtained research participation credit. They were reminded to fill out the delayed posttest one week later and were then dismissed. The same conditions were used as in the first experiment.

Results

Preliminary analysis

No differences arose on any of the variables in the demographic questionnaire. Students reported they use the Internet for 4.40 hours per day ($SD = 1.95$) and estimate their IPS ability with a 6.32 out of 10 (range = 3-8). Missing data was handled the same way as in Experiment 1.

Information problem solving skill tests

On the pretest, students obtained an average score of 35.02% ($SD = 11.45$) which increased to 57.90% ($SD = 10.04$) on the posttest. Table 3.3 shows an overview of scores per condition. The analysis revealed a significantly higher posttest score in the modeling example group when controlling for pretest scores: $F(1, 55) = 4.46$, $p = .040$, $\eta^2_{partial} = .079$. Running the same analysis on the delayed posttest scores indicated that the effect of modeling example remains significant: $F(1, 54) = 5.51$, $p = .023$, $\eta^2_{partial} = .097$.

Table 3.3. *Means and standard deviations of scores on the skills test (in percentages) per condition*

Condition	Pretest (*SD*)	Posttest (*SD*)	Delayed posttest (*SD*)
Modeling example	34.15 (11.35)	60.71 (9.45)	58.93 (15.72)
Practice task	36.00 (11.73)	54.75 (9.93)	52.42 (16.59)

Mental effort

Reported mental effort ratings are displayed in Figure 3.2. The repeated measures analysis showed that reported mental effort changes significantly over time: $F(3, 49) = 2.76$, $p = .045$, but with a small effect size: $\eta^2_{partial} = .055$. As in Experiment 1, subsequent contrast analysis indicated a significant decline from learning task 3 to 4: $F(1, 49) = 6.58$, $p = .014$, $\eta^2_{partial} = .123$. While mean mental effort scores were higher on each learning task in the modeling example condition, there was no significant difference between the conditions: $F(1, 49) = 2.79$, $p = .102$.

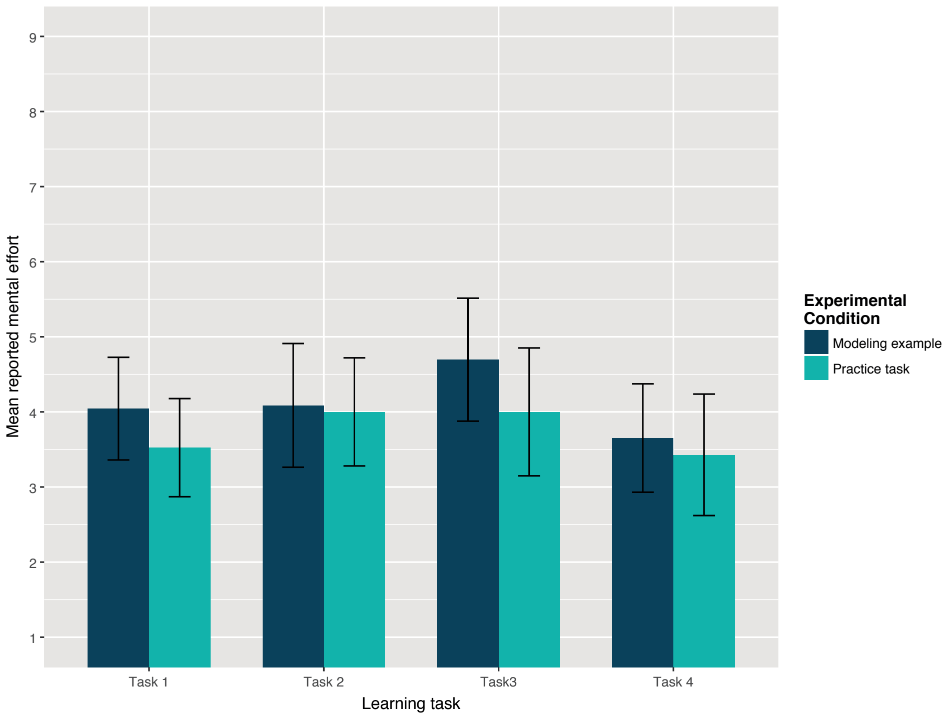


Figure 3.2. Reported mental effort per learning task for both conditions

Conclusion

The findings in this replication study resemble those of the first experiment. Students receiving a modeling example achieve higher learning outcomes (i.e., scores on the skills test) than students receiving a practice task instead. This supports the hypothesis that modeling examples, in which students actively process an example of problem solving, are more effective for teaching IPS skills than practice tasks in which students practice the newly acquired knowledge by themselves. The analysis of delayed posttest scores reveals that the learning effect of the modeling example persists at least one week after the training.

Students in the first experiment reported mental effort scores around or slightly below the midpoint on the 9-point scale. In the second experiment, these scores are lower and are scattered around the 4-point mark, yet they follow the same pattern as in the first experiment declining significantly in the final learning task. The lack of a difference between both conditions indicated that receiving a modeling example does not alter the amount of reported mental effort during the learning phase.

General discussion

These experiments were designed to investigate the effect of learning from a modeling example in which learners see the application of a solution procedure accompanied by additional procedural information ("how") and application of domain-specific knowledge ("why"), on the acquisition of IPS skills. Experiment 1 showed that students who receive a modeling example significantly outperform students who receive a practice task. Experiment 2 showed similar findings in a larger yet comparable sample, and revealed that this effect persisted after one week. Compared to a practice task, a single modeling example was found more effective for the formation of cognitive models and strategies needed for IPS performance. The results in this study also illustrate the low level of performance of first-year university students on IPS tasks. Untrained, they obtain average scores of under 40% on the skills tests. This observation directly opposes claims that students are digitally native and naturally develop the necessary skills to deal with information technologies (Prensky, 2001; Tapscott, 1999). The ease with which they seemingly manage to retrieve information online seems to mislead those who present these claims. These results once again underline the necessity for formal training in the area of IPS (Bennett, Maton, & Kervin, 2008; Kirschner & van Merriënboer, 2013; Smith, 2012)

For teachers and researchers in the domain of information literacy or IPS, this study shows that modeling examples are effective strategies to employ when training students to solve information problems. While beneficial effects of modeling examples were found in other domains, no research was found confirming the same for the domain of IPS. This research fills that gap. It shows that a modeling example, designed according to prevailing insights and principles derived from research in cognitive load theory and social learning theory, is effective for the development of an important 21st century skill such as IPS. Teachers are advised to consider using modeling examples during IPS instruction (Hilbert, Renkl, Schworm, Kessler, & Reiss, 2008). For the domain of instructional design, the results of this research imply that example-based learning can be effective in ill-structured domains, such as IPS. Video-based modeling examples can form an effective element of instruction to foster a complex skill such as solving information problems. The findings add a data point to the body of research showing that modeling examples are effective for teaching complex skills in ill-structured domains (e.g., Braaksma et al., 2004; Van Gog et al., 2004). They show that an instructional design containing an instruction video, a single modeling example followed by a period of practice leads to higher skill acquisition than an instruction video followed by mere problem solving. More specifically, they show that a 15-minute video of a modeling example, segmented and interspersed with cognitive prompts, taking multimedia principles into account and followed by four learning tasks was able to achieve a higher learning effect than instruction followed by mere practice. Teachers, instructional designers, or researchers interested in developing effective IPS instruction are therefore advised to consider including well-designed modeling examples.

In addition, the results show that viewing a modeling example did not affect reported mental effort during the practice phase. Solving information problems effectively and efficiently requires the integration of knowledge, skills, and attitudes and the coordination of several constituent skills. Because these experiments focused on the effects of the modeling example, the learning tasks were intentionally stripped of all support and guidance – such as worked-out steps or prompts – to avoid confounding effects. For novices, solving information problems without receiving any form of built-in task support should be cognitively demanding. Yet, average experienced mental effort ratings were scattered around the 4-point mark on the scale, which corresponds to *rather low mental effort*. Students in the second experiment had approximately five minutes longer to complete each learning task and reported less mental effort than students in the first experiment. This is likely causal: more time means less time pressure which means lower cognitive demand (Frerejean et al., 2016). Not much research exists on experienced mental effort during search tasks, making it difficult to compare these ratings, but they seem to be slightly lower than in other studies (Kim & Rieh, 2005; Rieh, Kim, & Markey, 2012).

These low mental effort ratings might indicate low investment. It may be the case that students regarded the tasks as simple teacher-imposed obligations with little relevance, which lowered their motivation and lead them to invest little energy in performing the tasks (De Vries, van der Meij, & Lazonder, 2008; Russell & Grimes, 2007). While effort was made to create learning tasks on socio-scientific topics with relevance to the study domain of the students, they were not topics that were integrated in the curriculum outside of the presented IPS training session. While some students informally expressed they perceived the training as somewhat boring and long, motivation and perceived relevance were not measured in the study, making it difficult to draw any solid conclusions from these statements.

The pattern of reported mental effort was similar in both experiments: it remained stable over the first three tasks in both conditions, then significantly dropped in the final learning task. Perhaps working on several conventional tasks in a row might have demotivated students, making them decide to rush through the final task to end the session. As motivation is one of the four governing processes as identified by Bandura (1977), one might expect that task content more aligned to the students' curriculum might increase learning effects. However, when interpreting mental effort ratings, it is important to remember that without knowing whether the cognitive demand refers to load that leads to learning (i.e., germane load) or hinders learning (i.e., extraneous load), one cannot explain effects on learning outcomes (Van Merriënboer & Ayres, 2005). No data was collected on students' motivation, so these findings merely warrant a suspicion that students have experienced these short, conventional problems as uninteresting and therefore invested less energy in performing the task to their best abilities.

As an alternative explanation, the low mental effort ratings can be caused by overestimation. The ease with which students find information online might lead them to overestimate their ability to solve information problems in an effective and efficient way. A Dunning-Kruger effect can occur, where the unskilled learners are unable to assess their own level of competence and consider themselves more skilled than they are (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger & Dunning, 1999). Indeed, students' perceptions of their own competence were higher than their objective scores in the pretest: an estimation of 6.28 on a scale of 10 compared to a score of 38.94% in Experiment 1 and 6.32 on the same scale compared to 35.02% in Experiment 2. This contrast between skills perception and actual performance points in the direction of a Dunning-Kruger effect (Kruger & Dunning, 1999). After the training, approximately one-third of the students informally stated they already knew much of what was taught in the training, showing that students might think of themselves as competent, while in reality their scores after the training are still below 60%. Students are apparently unable to correctly judge their IPS performance.

Several general limitations should be considered when interpreting these results. First, due to time constrains, the training session could only include one modeling example. Results from previous research suggest using multiple examples allows students to detect structural and surface features (Atkinson et al., 2000; Renkl, 2014). Multiple examples can improve the abstraction of knowledge schema because students have more opportunities for encoding information from examples and comparing their schema to the expert performance (Alfieri, Nokes-Malach, & Schunn, 2013; Gerjets, Scheiter, & Schuh, 2008). Additionally, only one type of search task was included: a simple information collection task using a general search engine. This prevents any conclusions about transfer and generalizability to tasks with a different level of complexity (Becerril & Badia, 2015), such as tasks that require specific information (e.g., academic articles) or specific strategies (e.g., using an academic literature database). Researchers and instructional designers need to further investigate how employing sequences of examples can lead to transfer and contribute to teaching skills in a way that allows students to apply them in different contexts (Fyfe, McNeil, Son, & Goldstone, 2014; Johnson, Reisslein, & Reisslein, 2014).

Second, while this study investigated effects of an integral modeling example, variations in the design of that example can impact those effects. Instructional designers can make a myriad of design choices concerning length, visual design, application of multimedia principles, method of presentation, and information provided in each example (Hoogerheide, Loyens, & van Gog, 2014; Van Gog, Verveer, & Verveer, 2014a). The investigated modeling example was optimized to achieve maximum effect in the ecologically valid setting of the current study, based on known best practices in instructional design. For that reason, it is not possible to attribute the learning effect to one of the design choices or the application of a specific principle (i.e., segmenting, prompting, etc.). Further research is necessary to disentangle these effects and detect which design choices are most effective. To achieve this, researchers need to measure process variables such as mental effort and attention continually during the processing of the example, in addition to learning (Spanjers, Wouters, van Gog, & van Merriënboer, 2011). Such methodology is already employed, for example in research by Kammerer et al. (2012), which combines eye tracking methodology, process logging, and verbal protocols. With information on learning processes that occur during example processing, and by comparing different designs, conclusions can be drawn about the application of effects of individual instructional principles and design choices.

To conclude, the intervention in the current study is a short, one-shot, standalone training and yields only small effect sizes, yet it shows a promising result: modeling examples are effective tools for fostering IPS skills. Based on the findings it can be predicted that these skills can be developed with a well-designed training program including modeling examples and providing sufficient time for practice. A longitudinal approach, where IPS instruction is embedded in a curriculum and combined with domain-specific instruction might be a fruitful design to achieve this challenging goal (Argelagós & Pifarré, 2012; Rosman, Mayer, & Krampen, 2016a; Wopereis, Brand-Gruwel, & Vermetten, 2008).

Chapter 4

Embedded instruction to learn information problem solving

Effects of a whole task approach

Abstract

In contemporary education, students often need to use the Internet to find information for solving a problem and completing a learning task. Teachers assume that students are sufficiently skilled to do so, but research shows the skills necessary for effective information problem solving (IPS) are more often than not underdeveloped. This chapter presents a study on embedded IPS training consisting of whole IPS tasks integrated in a 20-week course on vocabulary development, and its effects on student teachers' IPS skills. Skill measurements show that student teachers receiving the training search and select information more systematically, but their search queries, sources, and solutions are not of significantly higher quality than those of student teachers who received the regular course without IPS training. The training therefore succeeded in developing cognitive strategies for approaching an information problem, but did not create lasting improvements in all aspects of the IPS skill. Methodological and practical implications are discussed.

This chapter is based on:

Frerejean, J., Velthorst, G., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2017). Embedded instruction to learn information problem solving: Effects of a whole task approach. *Manuscript submitted for publication.*

INTRODUCTION

Contemporary education increasingly relies on tasks that confront students with a problem, but contain insufficient information to reach the solution. Students are then required to undertake a search operation to collect the necessary information from reliable sources and combine this information to formulate a complete and correct solution. Effectively and efficiently solving such information problems requires a set of complex skills (Brand-Gruwel, Wopereis, & Vermetten, 2005; Rosman, Mayer, & Krampen, 2016b), domain-specific knowledge (Lucassen & Schraagen, 2013; Salmerón, Kammerer, & García-Carrión, 2013), and a critical attitude to correctly judge the relevance and quality of information sources (Kammerer, Bråten, Gerjets, & Strømsø, 2012; Walraven, Brand-Gruwel, & Boshuizen, 2013). Observing the ease with which students navigate the Internet and uncover information, it is tempting to believe they automatically develop these requirements without any explicit instruction, but this is not the case (Kirschner & van Merriënboer, 2013). While some aspects of information literacy, such as operational skills (Van Deursen & van Dijk, 2009), might indeed develop quickly, research shows that information problem solving (IPS) skills do not develop sufficiently from mere exposure to online search tasks, making formal training necessary (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2016; Van Deursen & van Diepen, 2013; Walraven, Brand-Gruwel, & Boshuizen, 2008). This study presents such formal training in the form of whole-task IPS instruction embedded in a content knowledge domain, and reports its effects on students' IPS skills.

INFORMATION PROBLEM SOLVING

Many educational programs that adopt resource-based, inquiry-based, or problem-based approaches, confront students with information problems and expect them to find the necessary information and construct knowledge on their own (Hill & Hannafin, 2001). To function in such environments, students need sufficiently developed IPS skills. Effective and efficient skill execution requires the coordination of several constituent skills and application of knowledge and attitudes (Kirschner & van Merriënboer, 2009). Figure 4.1 presents a five-step approach to solving information problems based on a decomposition of the skill into constituent skills. When confronted with an information problem, the student has to define the problem, determine what information is needed, and formulate a clear and concise question. This question often contains the core concepts that can subsequently be used as search terms in the search engine. On the search engine results page (SERP), critical evaluation of the results is necessary to determine which sources appear relevant and reliable. By judging the source's relevance and trustworthiness, the student determines if the information is useful. Useful information is then processed more deeply and contrasted with own knowledge and information from other sources. When sufficient information is processed, the student integrates the selected information to formulate an answer to the question and presents the solution to the problem.

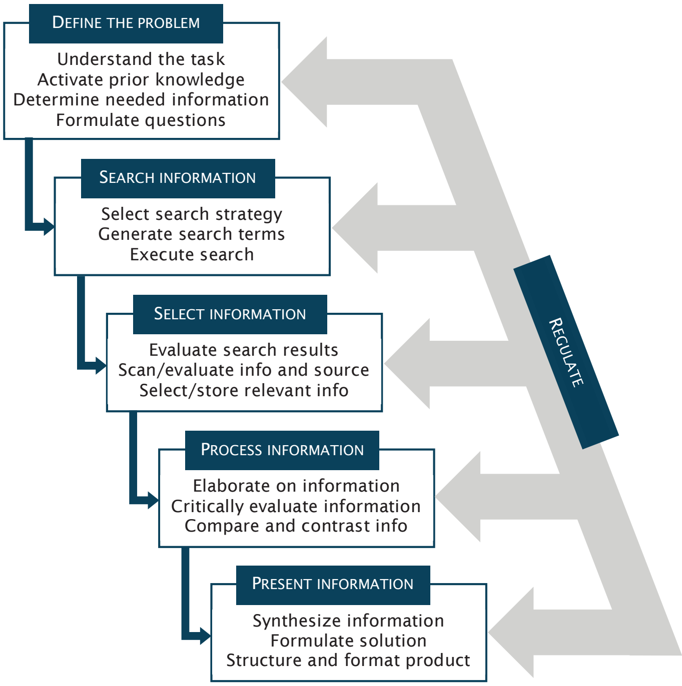


Figure 4.1. Five-step systematic approach to information problem solving (based on Brand-Gruwel et al., 2005; Frerejean et al., 2016)

Research shows that at all levels of education these IPS skills are underdeveloped and overestimated, underlining the necessity for IPS instruction (Frerejean et al., 2016; Kirschner & van Merriënboer, 2013; Miller & Bartlett, 2012; Rosman, Mayer, & Krampen, 2014; Walraven et al., 2008). While this need is often acknowledged, schools and teachers are poorly equipped to systematically integrate IPS in the educational program (Badke, 2010), often resulting in short classroom sessions or in one-shot library training sessions (Derakhshan & Singh, 2011; Probert, 2009). While these sessions can have positive effects on skill acquisition, research on contextualization shows that integrating instruction within a meaningful context and presenting it simultaneously with domain-specific instruction can lead to deeper learning and improved transfer (Perin, 2011). Prior research indeed found benefits of embedded IPS instruction. For example, in primary education, Kuiper, Volman, and Terwel (2008) showed that fifth graders benefit from embedded instruction on search and evaluation skills in a course focusing on healthy food. Argelagós and Pifarré (2012) showed that in secondary education, students receiving IPS instruction embedded in a two-year curriculum, outperform students following the regular curriculum. Similarly, Squibb and Mikkelsen (2016) showed that university students' information literacy skills improved most after following IPS instruction integrated in a writing curriculum, when compared to standalone library training or no training at all. These findings indicate that contextualized presentation of IPS instruction is preferred above standalone, one-shot instruction sessions that do not focus on domain-specific content.

Other instructional interventions focus only on improving specific aspects of the IPS skill, such as source evaluation skills (Britt & Aglinskas, 2002; Gerjets, Kammerer, & Werner, 2011). In these fragmented approaches, constituent skills are taught separately, out of context of the whole skill. While such programs can also be effective, they present little whole-task practice for integrating and coordinating all the constituent skills. This makes it difficult to transfer and apply these skills when needed in research projects or when writing theses (Van Merriënboer & Kirschner, 2013). Indeed, when Walraven et al. (2013) integrated instruction on source evaluation in a history program for ninth graders, they found students' evaluation skills significantly improved compared to a control group, but the training did not lead to transfer.

The prevailing view is that whole-task instruction is more effective to teach complex skills such as IPS than fragmented, part-task instruction (Lim, Reiser, & Olina, 2009; Van Merriënboer, Kirschner, & Kester, 2003).Whole-task approaches for IPS require the student to solve information problems from beginning to end, thereby performing and practicing all of the constituent skills of the IPS process. A study by Frerejean et al. (2016) evaluated a two-hour standalone online IPS training adopting a whole-task approach for first-year university students. The training offered authentic search tasks and significantly improved students' IPS skills, albeit with small effects. Wopereis, Frerejean, and Brand-Gruwel (2015) evaluated a standalone university-level IPS course using a similar design. Again, results showed that a holistic approach using a series of varied learning tasks was effective to improve students' IPS skills.

To design instruction for complex skills with a whole task approach, the four-component instructional design (4C/ID) model (Van Merriënboer, Clark, & de Croock, 2002; Van Merriënboer & Kirschner, 2013) presents a useful blueprint for the design of instruction using four components (see Figure 4.2): (1) *Learning tasks* form the backbone of the instructional blueprint. Authentic learning tasks mimic real-life situations that are encountered in practice, but with sufficient built-in task support and guidance to assist the learner. Examples of task support mechanisms are the *completion strategy* (Van Merriënboer, 1990; Van Merriënboer & de Croock, 1995), which uses completely worked-out problems at the start of the training and removes parts of the worked-out solution as training progresses, or *emphasis manipulation* (Gopher, 2007; Gopher, Weil, & Siegel, 1989), which reduces cognitive demand by emphasizing one aspect of the skill in a learning task. (2) *Supportive information* is presented to develop a cognitive models and strategies necessary to complete the learning tasks. (3) *Procedural information* is included by providing step-by-step instruction at the moment the learner performs recurrent and procedural aspects of the skill (Van Merriënboer, 2013). For online IPS, instrumental skills such as using a browser, mouse, and keyboard are examples of required recurrent skills, but these are often already acquired by the time IPS instruction starts and therefore do not need to be taught (Van Deursen & van Dijk, 2009). (4) *Part-task practice* can be included to provide repeated practice for recurrent skills. For IPS, there are no recurring aspects requiring a high degree of automaticity. Instruction developed according to 4C/ID principles was found effective for the development of skills in domains of technical expertise (Sarfo & Elen, 2007), communication (Susilo, van Merriënboer, van Dalen, Claramita, & Scherpbier, 2013), electrical skills (Melo & Miranda, 2015), and medical education (Vandewaetere et al., 2015).

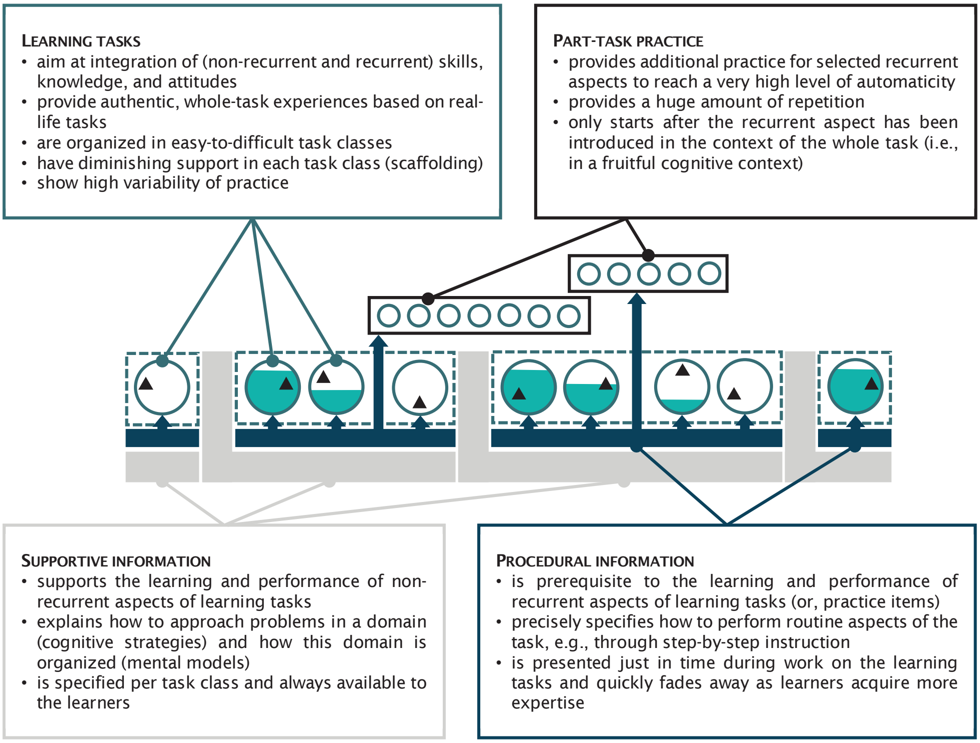


Figure 4.2. Four–component instructional design (4C/ID) model (based on Van Merriënboer & Kirschner, 2013)

From the results above it can be expected that IPS instruction is most effective when it is presented in the context of domain-specific instruction and it uses whole tasks that require integration of knowledge, skills, and attitudes. Brand-Gruwel and Wopereis (2006) and Wopereis, Brand-Gruwel, and Vermetten (2008) attempted to combine these two guidelines and implement whole tasks as embedded IPS instruction. They found an embedded whole-task approach lead to more explicit execution of IPS skills. Students receiving the instruction performed many of the IPS skills longer and more often when compared to untrained students. Although these studies were done in small groups with limited instruction and found small effects, the results nonetheless encourage further research.

THE PRESENT STUDY

The present study attempts to further investigate the effects of embedded and whole-task instruction in an ecologically valid setting, using more detailed measurements and in a larger sample. The study aims to answer the question *What are the effects of embedded IPS instruction on development of each of the five key constituent skills in IPS?* To answer this question, an existing educational program in first-year teacher training was redesigned to incorporate whole-task IPS instruction according to 4C/ID principles, resulting in a blended course that makes use of face-to-face workgroups and an online learning environment containing IPS tasks. In a quasi-experimental design, IPS skills of students following the regular curriculum were compared to those of students following the redesigned curriculum with embedded IPS training. Based on the literature discussed above, it can be predicted that this embedded, whole-task approach is effective to develop all constituent skills needed for IPS. More specifically, it is expected that students receiving embedded training display better performance on each of the five key skills of the IPS process than their counterparts who receive the regular curriculum.

- H1: students receiving IPS instruction will display more problem definition activities, such as actively determining the needed information or formulating a question.
- H2: students receiving IPS instruction will use more relevant search queries and will display a more systematic approach in their search process.
- H3: students receiving IPS instruction will select more relevant and more trustworthy sources.
- H4: students receiving IPS instruction will spend more time processing high quality sources than low quality sources.
- H5: students receiving IPS instruction will produce a better product, as measured by the number of relevant concepts presented within.

METHOD

PARTICIPANTS

A total of 155 student teachers enrolled in a Dutch teacher-training program preparing them for teaching at primary school level with an average age of 19.1 years (*SD* = 1.64) participated in this study. The sample comprised four classes in the first year (N = 75, M_{age} = 19.0, *SD* = 1.56, 27 male, 48 female) and four classes in the year thereafter (N = 80, M_{age} = 19.2, *SD* = 1.72, 20 male, 60 female). In the remainder of this chapter, the term *students* refers to the student teachers who participated in this study, while the term *teachers* refers to the teaching staff at the teacher training institute.

Materials

Regular course

The targeted 20-week course with a study load of 112 hours focused on language education for primary school children with an emphasis on vocabulary development. The project-based course revolved around a research project in which students conducted a small classroom intervention to develop primary school students' vocabulary at a partnering primary school. Students collaborated in groups of four to describe the effects of their intervention in a report which was submitted for assessment and grading by the teachers. A sufficient grade (higher than 5.5 out of 10) was necessary to pass the course.

During the course, students followed six classroom sessions (i.e., workgroups): two dealing with domain-specific knowledge about language learning and vocabulary development, and four in which teachers guided students through the research project (i.e., conducting a literature study, designing the intervention, analyzing its results and writing the report). A team of five teachers were active in the course, supervising the groups of students and running the workgroups. Students received a template document for their research proposal in which they recorded the research question, strategy for searching literature, and search terms. The teachers then provided feedback on this document. Other than that, there was no explicit instruction presented to teach students how to perform a literature research on the Internet, or on any of the other aspects of IPS process, such as formulating a research question or generating relevant search terms. Teachers operated under the assumption that students either already possessed these skills or possessed sufficient self-regulatory skills to develop them during the course with the limited feedback provided.

Embedded IPS instruction

To address the lack of explicit IPS instruction, the course was redesigned into a blended course of 112 study hours where IPS skills training (approx. 15 study hours) was embedded. The course was identical in terms of the number and content of the workgroups, learning materials, and assessment methods. One of the five teachers was replaced with a new teacher. An online learning environment allowed students to perform IPS learning tasks at home in their own time. The environment contained materials aimed at developing the necessary domain knowledge and solution strategies for solving the information problems. This supportive information preceded the five learning tasks. During these tasks, which can be characterized as *evaluation tasks* (Wirth, Sommer, von Pape, & Karnowski, 2015) or *interpretation tasks* (Becerril & Badia, 2015), on-screen instructions systematically guided students through the necessary steps. The learning tasks follow a completion strategy and contain decreasing amounts of built-in task support (i.e., scaffolding). In addition, emphasis manipulation is employed by applying a *prompt triad:* an approach that emphasizes parts of task execution with anticipative, instructional, and reflective prompts (for a description of the prompt triad, see Frerejean et al., 2016). In addition to the online activities, one classroom session was added halfway through the semester in which two researchers provided cognitive feedback on students' performance on the learning tasks. Table 4.1 shows an overview of the design of the online IPS training. Table 4.2 shows a timeline of the semester in both years and displays the differences and similarities.

Table 4.1: Overview of the IPS training

Training element	Description
Supportive information	Seven instructional videos (total: 32 minutes), giving an overview of the five-step approach and how to regulate the process.
Learning task 1	Video of a model demonstrating how a competent information problem solver approaches and solves an information problem, taking a recent newspaper article as a starting point for a web search on the effects of reading aloud on children's vocabulary.
Learning task 2	Completion task on the question: *How do teachers provide good vocabulary education in primary schools?* Based on a worked-out research question and a set of given search terms, students search for information and produce a presentation.
Learning task 3	Completion task on the question: *How does vocabulary size affect school success in primary school students?* Based on a worked-out research question the students searched and collected information to produce a presentation.
Feedback session	A face-to-face session where two researchers provide feedback on frequently observed errors and inefficiencies displayed in students' learning task performance.
Learning task 4	Conventional task in which students formulate a question about *assessment of classroom interventions*, search for information online and summarize their answer to the question in 200 words.
Learning task 5	Conventional task in which students collect and process sources to produce an outline for the theoretical framework of their research report. Formulating a research question and generating search terms was done in groups of four, but the search was performed individually.

Measurement of IPS skills

IPS skills were tested with a pretest, posttest and delayed posttest, performed in the same online environment. The tests presented an authentic information problem containing a problem description and instruction to collect relevant information to solve the problem and present this information in a mind map. The topics of the tests were *effects of mandatory school uniforms on bullying*, *effects of late-night media usage on sleep*, and *effects of GPS navigation on traffic safety*, respectively. The test topics and difficulty were determined by the researchers and presented to the teachers for assessment. After deliberation, the tasks were determined to be of comparable relevance and difficulty for the student group.

Table 4.2. *Course timeline*

Week	Regular course	Course with embedded IPS
1	Pretest	Pretest
2	Research workgroup (1) focusing on research in education. Studying a worked-out research report.	Research workgroup (1) focusing on research in education. Studying a worked-out research report.
3	Vocabulary workgroup (1) focusing on vocabulary instruction in primary schools.	Vocabulary workgroup (1) focusing on vocabulary instruction in primary schools.
4 & 5	Students pick a research topic and conduct a literature study, using the provided template document	Students work in the online environment, study supportive information and perform tasks 1-3.
6	Vocabulary workgroup (2): students report on literature study and practice reading aloud.	Vocabulary workgroup (2): students give presentations and practice reading aloud. Cognitive feedback session lead by researchers: reflecting on IPS performance
7	Vacation	Vacation
8	Students conduct intervention at partnering primary school	Students conduct intervention at partnering primary school
9	Research workgroup (2) focusing on writing a research report	Research workgroup (2) focusing on writing a research report
9 & 10	Students write their research report	Students work in the online environment and perform task 4 Students write their research report
11 & 12	Research workgroup (3) focusing on feedback on the research report	Research workgroup (3) focusing on feedback on the research report Students work in the online environment and perform task 5
13	Research workgroup (4) focusing on feedback on the research report	Research workgroup (4) focusing on feedback on the research report
14	Posttest	Posttest
15	Vacation	Vacation
16	Students work on research report	Students work on research report
17	Submission deadline for report	Submission deadline for report
18	No activities	No activities
19	Delayed posttest	Delayed posttest
20	End of semester	End of semester

Procedure

At the start of the semester, students were informed that they would participate in a research project. The pretest was administered in the first week and took place in the computer rooms under supervision of the researchers. Students were informed that participation in the pretest, posttest, and delayed posttest was not mandatory, but were kindly requested to participate voluntarily, as the course itself revolved around learning about educational interventions. At the time of the pretest, students received an introduction on the test procedure and built-in functionalities of the online environment (e.g., mind mapping). They then received a problem description and were given four minutes to create a mind map of their prior knowledge, without consulting online sources. Students were then given 20 minutes to collect information and create the mind map, during which all measurable browser actions were recorded with a Firefox® browser plugin and stored in a log file for each student. Five minutes after starting the task, students received an on-screen prompt to report what they did during or in the past five minutes. This information was used to assess whether skills concerning *define the problem* were performed (i.e., determining the needed information or formulating a question). This prompt, combined with a full log of the search process provided researchers with sufficient information to assess the key aspects in the IPS process: *problem definition, search process, selection of sources, processing of information,* and *the solution*. Assessment of these skills and calculation of scores is explained in the section Data analysis. The procedure for the posttest and delayed posttest was identical.

Data analysis

Log file parsing

The log files obtained during the tests were parsed to obtain a chronological overview of the students' actions, and an overview of the queries and sources. This information was combined with the mind maps and the answer to the five-minute mark prompt. Table 4.3 presents an overview of all variables in the current study.

Assessing prior knowledge

Previous research shows that prior domain knowledge is an important factor affecting multiple aspects of the IPS process (MaKinster, Beghetto, & Plucker, 2002), such as query formulation (Monchaux, Amadieu, Chevalier, & Mariné, 2015) and source evaluation (Salmerón et al., 2013). It was therefore included as a covariate in several analyses in the current study. To assess prior knowledge, the number of relevant idea units in the prior knowledge mind map was assessed by counting the unique idea units and comparing them to the maximum number of idea units for the respective task. *Prior knowledge* was therefore expressed as the percentage of idea units in the mind map, compared to all possible idea units. Two researchers scored 10% of the mind maps to obtain interrater agreement. The mixed model, absolute, single-measure intra-class correlation was .989, indicating high agreement.

Table 4.3. *Overview of variables and their measurement*

Variable	Measurement
Prior knowledge	The number of idea units included in the mind map at the start of the test, as percentage of the maximum number of idea units in the respective task.
Problem definition activities	Inspection of answer to prompt. "Yes" if students reported performing problem definition activities, "no" otherwise.
Number of queries	The number of unique queries used by a student.
Query relevance	Average relevance of students' search queries, in percentage, assessed by the researchers.
Query: systematic approach	Researchers' assessment of systematic approach to the search process, expressed in a percentage score.
Number of sources	The number of unique sources visited by a student.
Average source trustworthiness	Average trustworthiness of students' selected sources [0-3]
Average source coverage	Average number of idea units in students' selected sources, as percentage of the maximum number of idea units in the respective task.
Selection: systematic approach	Researchers' assessment of systematic approach to selecting sources, expressed in a percentage score.
Solution score	The number of idea units included in the mind map at the end of the task, as percentage of the maximum number of idea units.

Assessing problem definition

To test the first hypothesis that trained students perform more problem definition activities than untrained students, the number of problem definition activities reported in the answers to the prompt were counted. To ensure that students understood the problem, they were allowed to ask questions before start of the test. In addition, all tests started with measurement of prior knowledge. Because the subskills *understand the task* and *activate prior knowledge* were performed as part of the test, assessment focused only the other two subskills of problem definition: *determine needed information*, and *formulating question(s)*. Answers to the prompts were often no longer than one sentence, so a score of 1 was awarded if the student reported any of these *problem definition* activities, or a 0 if none of these activities were mentioned. Two researchers collaborated to score all cases. A Chi-square test was performed on these scores to detect differences.

Assessing the search process

To assess the *search process*, two key aspects were considered: *query relevance* and *systematic approach*. To assess *query relevance*, a score was awarded for how relevant the chosen keywords were to the respective problem. A scoring matrix was produced, where each unique term received score between 0 (irrelevant) and 3 points (highly relevant). As an effective query generally contains three terms: the two key concepts and the relationship between them, each query received a total score between 0 and 9 (3 terms each worth maximally 3 points). For each student, the average query relevance was then calculated. To assess the *systematic approach* during the search process, the researchers used a scoring sheet to assess how systematically students worked on the respective task, according to what was taught in the training. The assessment included the scope of the first query, logical and systematic adjustments based on this first query, the total number of queries, and the correct use of Boolean operators. This resulted in a score between 0 and 100. The assessment procedure for search skills is further detailed in Appendix 3.

Two researchers scored 150 of the 1451 queries allowing the calculation of an interrater reliability coefficient for *query relevance*. For *systematic approach*, 15 students were scored by two researchers. The intra-class correlation was .873 for *query relevance* and .956 for *systematic approach*. One researcher scored the remaining queries and students. To investigate Hypothesis 2 that trained students would display better search processes, MANCOVAs were performed on all three tests including *query relevance* and *systematic approach* as dependent variables, *training (yes vs. no)* as independent variable, and *prior knowledge* as a covariate.

Assessing source selection

To investigate Hypothesis 3 that trained students select sources of higher quality, researchers scored each of the approximately 1500 unique sources that were found on two dimensions. *Coverage* is defined as the number of unique idea units relevant to the task as a percentage of the combined number of unique idea units relevant to this task, from all sources (Wirth et al., 2015). *Trustworthiness* indicates the quality of the source as either *very trustworthy* (e.g., scientific reports), *trustworthy* (e.g., news articles from national news outlets), *questionable* (e.g., personal blogs), or *untrustworthy* (e.g., anonymous opinions on discussion forums), judged by aspects such as author reputation, goal of the text, and source of publication (Walraven, Brand-Gruwel, & Boshuizen, 2009). For each student, the average coverage and trustworthiness scores were complemented with a score for *systematic approach*, much like the assessment of the search process. Using a scoring sheet, a score was given on a scale of 0 to 100 by assessing the number of sources found, the variation of sources, persistence in accessing and processing relevant sources until the end of the task, and time spent on low and high-quality sources. These procedures are further detailed in Appendix 4.

Approximately 10% of all sources were scored by two raters, obtaining an intra-class correlation of .935 for trustworthiness and .989 for coverage. The interrater agreement for the *systematic approach* score was determined by double scoring 15 students, and amounted to .755. After further deliberation, one researcher rated the remaining cases. Differences between the conditions were investigated with a MANCOVA using *average coverage*, *average trust*, and *systematic approach* as dependent variables, *training (yes vs. no)* as the independent variable, and *prior knowledge* as a covariate.

Assessing processing of information

For assessing the processing of information, the *time spent on a source* was investigated. Hypothesis 4 states that trained students spend more time on trustworthy and relevant sources and less on irrelevant and untrustworthy sources. To assess this aspect, a multiple linear regression analysis was conducted on a dataset that contained all page visits, sources, durations, and the trustworthiness and coverage scores. The regression used *coverage* and *trust* as predictor variables, and *duration* as an outcome variable. On each of the three tests, the regression models were compared between the two conditions by including *training (yes vs. no)* as a predictor.

Assessing the solution

To assess the solution, the total number of idea units in the mind map was counted, identical to the assessment of the prior knowledge mind map. The *solution score* was expressed as the percentage of idea units in the mind map, compared to all possible idea units. This indicates the amount of information processed by the student to reach the solution. To investigate Hypothesis 5 that trained students show more relevant information in the solution, an ANCOVA was conducted on the *solution scores* using *condition* as an independent variable and *prior knowledge* as a covariate.

Results

Table 4.4 presents an overview of the means and standard deviations of the variables collected in this study.

Preliminary analysis

Students in the trained group and the untrained group showed no statistically significant differences in age (19.1 years, SD = 1.64), amount of Internet use (4.5 hours per day SD = 3.02), or prior knowledge on any of the three tests. Therefore, the groups can be considered comparable. Seven students did not complete the online training, and only their pretest data was retained for analysis. Out of 155 participants, data was complete for 147 on the pretest, 132 on the posttest and 115 on the delayed posttest, due to dropout and absence. Some technical issues with the mind mapping functionality led to some additional missing data in the outcome measures at the pretest, which explains why some of the statistical analyses are conducted on slightly smaller datasets.

Problem definition

To assess the skill *problem definition*, students' answers to the five-minute prompt were analyzed and scored if statements occurred reflecting either the determining of needed information or formulation of questions. The frequency of such statements was very low, occurring only three times on the pretest (twice for students in the trained group, once for students in the control group), eight times on the posttest (four times in both conditions), and six times on the retentiontest (three times in both conditions). Therefore, Fisher's exact test was used, which yielded insignificant results on the pretest: p = .480, posttest: p = .633, and delayed posttest: p = .660. Based on these results, the hypothesis that trained students display more activities concerning problem definition (H1) is rejected.

Searching for information

Before addressing Hypothesis 2, the number of used queries was explored. For the *number of queries* on the pretest, an ANCOVA using *prior knowledge* as covariate showed no statistically significant difference between trained students and untrained students: $F(1, 133) = .792$, $p = .375$. The same analysis on the posttest showed that after the training, trained students used significantly more queries than untrained students, when controlling for *prior knowledge*: $F(1, 119) = 41.499$, $p < .001$, $\eta^2_{partial} = .259$. On the delayed posttest, trained students did not use more or fewer queries than untrained students: $F(1, 93) = 1.357$, $p = .247$. The covariate *prior knowledge* displayed no statistically significant influence in any of the analyses.

Analysis of the skill *searching information* was performed by conducting a MANCOVA with *query relevance* and *systematic approach* as dependent variables, *training* as an independent variable, and *prior knowledge* as covariate. On the pretest, this analysis revealed no significant difference on query relevance and systematic approach between groups: $F(2, 132) = .764$, $p = .468$. On the posttest, a significant difference was found between trained and untrained students: $F(2, 117) = 16.177$, $p < .001$, $\eta^2_{partial} = .217$. Subsequent univariate analyses showed no significant difference on *query relevance*: $F(1, 118) = .077$, $p = .782$, but a significant difference on *systematic approach* $F(1, 118) = 12.856$, $p < .001$, $\eta^2_{partial} = .098$. The trained students achieved an average score of 47.17% while untrained students scored 31.95%, constituting a difference of 15.22, but a small effect size. On the delayed posttest these differences disappeared, and both groups of student showed similar scores: $F(2, 92) = 1.735$, $p = .182$. With these results, Hypothesis 2 is partially confirmed. Trained students showed a more systematic approach to searching information, but did not use more relevant queries. On the delayed posttest, both groups of students performed equally.

Selecting information

Before investigating Hypothesis 3 (i.e., trained students select more trustworthy and relevant sources than untrained students), first the *number of used sources* was analyzed. On the pretest, an ANCOVA using *prior knowledge* as covariate yielded no significant results, indicating both groups used a similar number of sources: $F(1, 132) = 2.061$, $p = .153$. On the posttest, trained students used significantly more sources than untrained students: $F(1, 119) = 15.199$, $p < .001$, $\eta^2_{partial} = .113$. However, on the delayed posttest, this difference is no longer present: $F(1, 93) = .488$, $p = .487$. The covariate shows a significant influence on the number of sources in the delayed posttest: $F(1, 93) = 4.407$, $p = .039$, $\eta^2_{partial} = .045$, yet the effect size is very small.

Table 4.4. *Mean and standard deviations of collected data, per test, by condition*

	Pretest		Posttest		Delayed posttest	
	Intervention	Control	Intervention	Control	Intervention	Control
Prior knowledge	8.13 (4.05)	6.96 (3.87)	10.48 (7.80)	8.42 (6.75)	9.29 (7.33)	7.84 (8.03)
Number of queries	4.98 (2.74)	4.65 (2.45)	5.75 (3.06)	2.82 (1.85)	7.83 (3.96)	6.90 (4.59)
Query relevance	32.03 (14.76)	33.51 (11.25)	36.71 (19.18)	37.44 (17.51)	33.67 (18.81)	30.32 (14.23)
Query: systematic approach	32.49 (15.31)	31.90 (13.83)	47.17 (24.18)	31.95 (19.93)	43.23 (21.77)	35.63 (18.63)
Number of sources	8.75 (4.12)	7.85 (2.95)	11.12 (3.59)	8.28 (3.91)	10.02 (4.44)	11.12 (5.61)
Average source trustworthiness	2.34 (0.48)	2.51 (0.61)	2.61 (0.33)	2.86 (0.31)	2.80 (0.47)	2.98 (0.44)
Average source coverage	18.90 (6.02)	19.00 (5.65)	12.73 (4.66)	16.14 (5.19)	13.19 (8.23)	13.01 (9.12)
Selection: systematic approach	42.63 (19.81)	41.31 (19.68)	56.72 (15.62)	40.89 (18.37)	42.77 (18.39)	46.07 (18.71)
Solution score	19.78 (5.13)	20.61 (6.38)	27.96 (9.05)	27.94 (10.37)	31.13 (11.60)	32.78 (12.05)

On the pretest, a MANCOVA using *average trustworthiness*, *average coverage*, and *systematic approach* as dependent variables, *training (yes vs. no)* as an independent variable, and *prior knowledge* as a covariate, yielded no significant differences: $F(3, 126) = .893$, $p = .447$. On the posttest, the difference was significant: $F(3, 116) = 18.482$, $p < .001$, $\eta^2_{partial} = .323$, which indicates that the scores compositing *selection of sources* differ between trained and untrained students. Further univariate analyses reveal that untrained students show higher *coverage* scores $F(1, 118) = 14.765$, $p < .001$, $\eta^2_{partial} = .111$, as well as *trustworthiness* scores: $F(1, 118) = 17.422$, $p < .001$, $\eta^2_{partial} = .129$. For *systematic approach*, the effect is reversed and trained students show significant higher scores than untrained students: $F(1, 118) = 22.712$, $p < .001$, $\eta^2_{partial} = .161$. Furthermore, the covariate *prior knowledge* appeared to have a small yet significant influence on *systematic approach* $F(1, 118) = 7.963$, $p = .006$, $\eta^2_{partial} = .063$. On the delayed posttest, all differences disappeared: $F(3, 92) = 1.739$, $p = .165$. Considering these results, the hypothesis that trained students show higher competence in selecting sources (H3) can only be partially confirmed for a systematic approach. However, for coverage and trustworthiness, untrained students score higher than trained students. All effects disappear on the delayed posttest.

The finding that untrained students select sources of higher trustworthiness and coverage on the posttest was unexpected and therefore warranted further investigation. On the posttest, 572 unique sources were visited in total by all students. To ease inspection, this dataset was first limited to only sources visited by more than one student. Trained students showed 409 page visits across 82 unique sources, and untrained students showed 379 visits across 61 sources. Analysis of these sources showed that untrained students made 50 visits to eight sources that had publication dates later than the date at which the trained students were posttested. Furthermore, the average trustworthiness and coverage of those eight sources (2.12 and 21.12%) was much higher than the average coverage and trustworthiness of the remaining 53 sources (1.74 and 16.51%) and the sources used by trained students (1.41 and 14.78%). This showed that the untrained students had made 50 visits to eight sources with above average coverage and trustworthiness that were unavailable to the trained students. The same investigation was carried out on the pretest and the delayed posttest data. On the pretest, only four newer sources were used by untrained students, and on the delayed posttest only two out of 51 sources were newer. No discrepancies in coverage and trustworthiness were found.

Processing information

To assess whether trained students spent more time on trustworthy and relevant sources, a multiple regression analysis was conducted to investigate if source trustworthiness and source coverage significantly predict the number of seconds students spend on a source (model 1). *Training* was added as a predictor, scored 1 for trained students, 0 for untrained students, to investigate whether a model including *training* (model 2) better predicted duration than the model with only coverage and trustworthiness (model 1). This was only the case on the posttest. Table 4.5 provides an overview of the results of the multiple regression. On the pretest, model 1 explained 16.5% of variance (R^2_{adj} = .164, $F(2, 1057) = 104.597$, $p < .001$). Including *training* as a predictor did not improve the model ($\Delta R^2 = .000$, $F_{change}(1, 1056) = .159$, $p = .690$). This indicates that coverage and trustworthiness predict duration similarly in both the trained group and the untrained group. In model 1, coverage significantly predicted duration: $t(1057) = 13.542$, $p < .001$, as did trustworthiness: $t(1057) = 4.913$, $p < .001$.

On the posttest, model 1 was not as strong as on the pretest, and explained less variance: 9.1% (R^2_{adj} = .09, $F(2, 1190) = 59.676$, $p < .001$). Including *training* as a predictor marginally improved the model ($\Delta R^2 = .003$, $F_{change}(1, 1189) = 3.593$, $p = .058$). In this second model, coverage was again a significant predictor: $t(1190) = 10.247$, $p < .001$, as was trustworthiness: $t(1190) = 2.476$, $p = .013$. Training, however, just failed to reach statistical significance as a predictor: $t(1190) = -1.896$, $p = .058$. The regression coefficient for *training* is negative, as can be seen in Table 4.5, which indicates that trained students spent approximately 5.5 seconds less on a source than untrained students. On the delayed posttest, model 1 explained 43.5% of variance (R^2_{adj} = .434, $F(2, 964) = 371.082$, $p < .001$). Including *training* as a predictor did not improve the model ($\Delta R^2 = .001$, $F_{change}(1, 963) = 1.781$, $p = .182$). Again, coverage formed a significant predictor $t(964) = 24.756$, $p < .001$, as did trustworthiness $t(964) = 5.308$, $p < .001$. After analysis, the models were checked for influential outliers, multicollinearity, and homoscedasticity to determine whether the regression models met all relevant assumptions. No violations of assumptions were found, but residuals in the pretest and delayed posttest show heteroscedasticity, which means generalization of the model is problematic and requires further investigation and replication.

Table 4.5. *Outcomes of multiple linear regression predicting duration in seconds spent on page*

	Pretest			Posttest			Delayed posttest		
	B	SE B	β	B	SE B	β	B	SE B	β
Model 1									
Constant	15.60	3.32		27.88	3.77		3.83	3.64	
Coverage	172.75**	12.76	.381	128.68**	12.16	.293	276.69**	11.18	.616
Trustworthiness	6.65**	1.35	.138	4.88**	1.78	.076	9.30**	1.75	.132
Model 2									
Constant	14.93	3.72		31.96	4.34		1.35	4.08	
Coverage	173.04**	12.78	.381	125.57**	12.25	.285	275.53**	11.21	.614
Trustworthiness	6.66**	1.35	.138	4.43*	1.79	.069	9.48**	1.76	.135
Training	1.25	3.13	.011	-5.48	2.89	-.053	4.70	3.52	.032

** significant at the .01 level; * significant at the .05 level

Solution

Even though students were not required to formulate a solution to the problem presented in the task, the mind map that was produced gave insight into the amount of information they believed relevant for their solution. The ANCOVA tests using *prior knowledge* as a covariate showed no significant differences on the pretest: $F(1, 105) = .456$, $p = .501$, posttest: $F(1, 126) = .091$, $p = .763$, or delayed posttest: $F(1, 100) = 1.083$, $p = .301$. The covariate *prior knowledge* showed significant effects on solution scores in the posttest: $F(1, 126) = 4.992$, $p = .027$, $\eta^2_{partial} = .038$ and in the delayed posttest: $F(1, 100) = 9.440$, $p = .003$, $\eta^2_{partial} = .086$. The hypothesis that trained students provide more relevant information in their solution (H5) is rejected on the basis of these results.

Discussion

This study investigated effects of a curriculum containing embedded IPS training, compared to a curriculum without explicit IPS instruction. This study distinguishes itself from other IPS studies by integrating whole-task IPS instruction within domain-specific instruction and by investigating IPS competence in an ecologically valid setting. In other studies, task performance is often constrained, for example by providing fabricated SERPs or a limited list of sources (e.g., Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017). The current study put few constraints on task performance, letting students work on realistic tasks in a natural environment, inducing a more realistic application of knowledge, skills, and attitudes. In addition, a novel method of data collection was applied. Automatic logging of all browser actions provided the researchers with rich data files containing thousands of data points and allowing for various analyses. While not without drawbacks, this research design and method of data collection delivered a detailed view on the five key skills in IPS performance, and how they were affected by embedded IPS training.

Results show that activities pertaining to *defining the problem* were scarce, which is common for novices (Brand-Gruwel et al., 2005). However, this persisted even after instruction. There are three possible reasons for this behavior. First, students might have received too few opportunities to practice problem definition skills, as they were mostly presented as a worked-out step in the current training design. Second, students might have decided that elaborate consideration of the problem was not necessary for these tasks because reading the task and activating their prior knowledge was already part of the test. The problems in the test were smaller and shorter than the learning tasks, making this explanation plausible. Third, the nature of the prompt might not have triggered students to report every action, limiting themselves to the most recent ones such as formulating search queries. In future research, different measurements should be employed to record problem definition activities more effectively.

Turning to *search skills*, results show that trained students do not formulate more relevant search queries. Inspection of the queries leads the researchers to believe students might have reverted to a data-driven approach, simply using the most salient or common search terms in the problem description. However, the lack of a training effect might also be a caused by students failing to transfer their acquired skills to a new situation, because the pretest, posttest, and delayed posttest were all tests on other topics than vocabulary development. Trained students did work more systematically while searching and showed a more logical progression of queries, making small changes instead of using a more trial-and-error approach where completely new queries are used repeatedly. The results therefore indicate that the training succeeded in developing a systematic approach that students could apply in the test setting.

Trained students also exhibited a more logical approach during the *selection* of their sources. They did not limit themselves to 'hits' at the top of the SERP, were more persistent in their source selection and used a greater variation of sources to gather the necessary information. They also used more sources on average than untrained students did, but the selected sources were not more trustworthy or relevant than those selected by untrained students. These findings can be explained by the fact that the untrained students had access to several very trustworthy and relevant sources that were unavailable at the time when trained students performed the posttest. This might also explain why the untrained group used fewer search queries and fewer sources than trained students. If those few queries already lead to good sources containing sufficient information, the need to use more queries or sources quickly diminishes. Also, because this set of sources was of high quality, the untrained group reaches similar average trustworthiness and coverage scores as the trained group. Finally, it might explain why trained students spent less time on relevant and trustworthy sources, although this effect was only marginally significant. If the untrained students used fewer sources, it follows logically that the average duration of a source visit is higher than that of the trained group.

Concerning *outcomes*, there were no differences in scores of outcomes between the two conditions. While striving for whole-task instruction, little attention was focused on presenting skills in the IPS training for two reasons. *Presenting* is a complex skill itself that can be done in a myriad of ways, and it is a time-consuming aspect of the IPS skill. Training presentation in a whole-task approach would require a large time investment, as would its assessment. Therefore, the finding that both groups perform equally on this aspect of the skill is unsurprising. There might possibly be differences in the *quality* of the collected information in the products, but it was not possible to retrace where students retrieved the information reported in the mind maps. Future research on IPS should therefore include measures that show where certain information was found.

In summary, the embedded instruction in this study was effective to develop several aspects of the skill, particularly pertaining to systematic approaches (i.e., systematic searching and selecting of sources), but induced no improvements or a showed lack of transfer for other aspects (e.g., query relevance, source selection). The lack of strong learning effects may partially be attributed to the quasi-experimental design. Testing two separate cohorts of students eliminates random assignment of participants that is necessary to ensure that control group and intervention group do not differ systematically. However, both groups appear to be comparable, as the preliminary questionnaire and the pretest indicated no significant differences. In addition, care was taken to provide both groups with a highly similar instructional sequence – apart from the added IPS. Despite these efforts, the rapid development of the Internet induced a biasing effect on the posttest, where untrained students had access to more recent and high-quality sources that were not available to the trained group.

An alternative hypothesis is that the untrained students improved their IPS skills without explicit instruction. In research by Rosman, Mayer, and Krampen (2016a), a comparable sample of students participating in a curriculum requiring information-seeking skills developed some level of IPS skills without explicit instruction and by self-regulated learning. The untrained students in this study might have similarly developed some strategies to improve their IPS skills. In any case, trained and untrained students perform equally five weeks after training, indicating that performance has regressed.

Implications

For the domain of IPS instruction, this study shows that students tend to spend little time on problem definition, which is in line with previous research (Brand-Gruwel, Wopereis, & Walraven, 2009; Walraven et al., 2008). For IPS teachers, this implies that problem definition skills should be strongly emphasized in IPS education to teach students the importance of understanding the task and the benefits of exploring the problem space before attempting a targeted search for information (Argelagós & Pifarré, 2016; Brand-Gruwel et al., 2005). Steering students toward a goal-driven approach instead of a data-driven approach avoids fragmented understanding (Land & Greene, 2000). Problem definition activities present a particularly interesting venue for further research. Research by Sarsfield (2014) showed that domain experts in professional domains generate complex, detailed problem representations, while novices form broad and superficial representations. More research is needed to investigate how learners perceive the problem at the start of the task, and whether this perception changes throughout the problem-solving process. Defining a problem might constitute an iterative process in itself, which might have implications for existing problem-solving models, such as the IPS-I model. Further research on this topic is warranted.

In general, this study shows that integrating whole-task IPS practice in domain-specific instruction can potentially be effective for the development of abstract knowledge structures and cognitive strategies necessary for IPS. However, the results also show that the effect quickly fades when practice is stopped. Therefore, to achieve and maintain the desired improvements, an educational program encompassing more opportunities for practice over a longer period (i.e., more deliberate practice, see Ericsson, Krampe, & Tesch-Römer, 1993) embedded in multiple domains might be more fruitful. This study further demonstrated an application of well-established instructional principles to design task-centered instruction incorporating scaffolding, examples, cognitive feedback, and blended delivery of instructional materials. Unfortunately, this arrangement did not lead to lasting improvements in all aspects of the IPS skill, either due to insufficient development of the skill, or lack of transfer to the testing domain.

For transfer of learning to occur, it is necessary to develop abstract or generalized knowledge, usually from dealing with a variety of specific problems (Kalyuga & Hanham, 2011). Variability of practice is one of the factors affecting transfer of learning (Van Merriënboer, Kester, & Paas, 2006), but learning tasks used in this study were all of the same type, in the same domain – vocabulary instruction – and required the same strategy to complete. Exposing students to problems with different surface features and structural features leads to formation of abstract knowledge that allows them to think more creatively when confronted with newer problems. Therefore, instructional designers who aim to adopt a whole-task approach to develop abstract cognitive schemas and strategies for performing higher-order skills in multiple domains are advised to incorporate more variation of problems in their educational program.

For researchers, the methodology of assessment adopted in this research provides a basis to develop a more detailed view of IPS performance. While log file analysis is often used in research on usability of information retrieval systems or search engines (Agosti, Crivellari, & Di Nunzio, 2012), it is not often used to investigate the search process from the searcher's point of view. This research has made clear that meticulous logging of activities during IPS performance on naturalistic search tasks provides a wealth of information, allowing a detailed view of the searcher's activities, choices, and strategies. However, it does not tell the whole story. By looking at objective measures, it is difficult to draw conclusions about some of the cognitive aspects of the task. Future research adopting a similar approach would benefit from additional qualitative data, such as thinking aloud protocols, interviews, or focus groups to investigate cognitive processes during phases of problem definition, search term formulation, or source evaluation (Brand-Gruwel et al., 2017; Gerjets et al., 2011; Van Gog, Paas, van Merriënboer, & Witte, 2005).

To conclude, this study showed that online, embedded, whole-task IPS instruction shows potential for developing IPS skills, and identifies areas where such instruction can be improved. In the end, the goal of developing IPS skills is to foster the ability in learners to find learning materials and effectively solve information problems in order to advance their domain-specific expertise. This notion is in agreement with Rieh, Collins-Thompson, Hansen, and Lee (2016), who suggest future research should adopt a broader framework, where objective search process characteristics stemming from log file analysis are linked to aspects such as learner intent, motivation, task complexity, and growth of domain-specific knowledge to paint a more complete picture of searching as a learning process.

CHAPTER 5

EXPLORING SOURCE SELECTION DURING WEB SEARCH

LOG ANALYSIS OF STUDENTS' INFORMATION PROBLEM SOLVING SESSIONS

ABSTRACT

Information problem solving is an essential skill to acquire for students in contemporary education, yet research consistently shows such skills are underdeveloped and students struggle with all aspects of the process. One of the crucial constituent skills for information problem solving is selecting relevant and trustworthy sources. Research on these skills is often conducted in controlled settings, where source evaluation is investigated out of context of a whole search task, and fabricated lists of search results and sources are given. This study investigates source evaluation in a realistic setting, where 135 students received an information problem and could freely search the web for sources. Results show that in a typical 20-minute search task, students used five queries, eight sources, and spend 65 seconds on a source on average. The relevance of the search query is positively associated with the amount of relevant information in a source, but negatively associated with its trustworthiness. While sources with more relevant information generally reside higher in the search results, this pattern is not found for the trustworthiness of sources. Students spend more time on sources that are relevant and trustworthy. Finally, the amount of relevant information and the source's trustworthiness appear to be negatively associated. Implications of these findings for the design of instructional interventions are presented.

Introduction

Often included under the umbrella of 21st century skills, information problem solving (IPS) is recognized by researchers and educators as an essential skill to teach in contemporary education (Van Laar, van Deursen, van Dijk, & de Haan, 2017). Many educational approaches, such as resource-based or problem-based learning often rely on students' IPS skills for gathering information to complete a task. Whether it is a primary school student writing a report on bears or a PhD student in biomedicine conducting a literature study for a research paper, both are likely to gather information from online sources. Starting with an initial question, the learner generally enters search terms into a search engine and subsequently receives a list of sources that may or may not be useful. With the information gathered from the selected sources, the learner formulates an answer and, if necessary, presents it – in written or oral form – to an audience.

Analysis of novice and expert IPS processes led to the development of a descriptive model of IPS using the Internet, the IPS-I model (Brand-Gruwel, Wopereis, & Vermetten, 2005; Brand-Gruwel, Wopereis, & Walraven, 2009). The IPS-I model distinguishes five phases of problem-solving and describes the skills necessary in each phase: definition of information problem, search for information, scan information, process information, organize and present information. It also includes a regulation component to indicate continuous orientation and monitoring activities during the problem-solving process. While more models of information search and retrieval exist (for an overview see: Wilson, 1999; Xie, 2011), they collectively demonstrate that IPS is complex and an efficient and effective search process requires integration and coordination of knowledge, skills, and attitudes.

Especially skills pertaining to source selection have received much attention in the research literature. To best solve an information problem, a student should make use of information that is relevant (i.e., contains information that is necessary to formulate a solution to the problem), and trustworthy (i.e., the information in the source is reliable, as is the source itself, and of high quality). Yet online sources show a great variation in relevance and trustworthiness, and can provide diverse and contradictory information. When searching information online, learners engage in a cognitive process to compare sources, often on the search engine results page (SERP), with the goal to select (i.e., click on) the link to the source they expect is most helpful to solve their problem. The evaluation process then continues while the information in the source is being processed and the learners are evaluating the source's characteristics such as publication date, quality, author reputation, etc. Research indicates, however, that these processes do not occur spontaneously.

Gerjets, Kammerer, and Werner (2011) show that students who are asked to think aloud during source evaluation express few verbal utterances concerning evaluation criteria. Students who are explicitly instructed to evaluate SERP results and sources talk more about evaluation criteria and select higher quality sources. Walraven, Brand-Gruwel, and Boshuizen (2009) report similar findings and show that students often do not use evaluation criteria when judging sources. When they do, they primarily focus on the source's relevance to the task. These findings are in accordance with other research on evaluation behavior showing that students struggle with source evaluation, and if they evaluate, they neglect quality aspects such as publication date or the author's authority on the subject (Walraven, Brand-Gruwel, & Boshuizen, 2008). More recent publications provide nice summaries of other research illustrating how students engage in source selection and evaluation and reach similar conclusions: students give limited justifications for their source selection, and students base their selection mostly on the source's relevance (e.g., Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; List, Grossnickle, & Alexander, 2016). This makes clear that evaluation skills essential for effective and efficient IPS are underdeveloped and need training.

Studies on evaluation skills often adopt a research design that investigates source evaluation in isolation, out of context of an information problem, and providing prefabricated and manipulated materials. These experimental designs create a context that is not comparable to an authentic IPS task. In such studies, evaluation behavior is not observed in a natural environment, because the controlled settings present a simplified version of the context in which source evaluation naturally occurs. Sometimes, SERPs are prefabricated (e.g., Brand-Gruwel et al., 2017; Kammerer, Bråten, Gerjets, & Strømsø, 2012), or a limited number of sources is given (e.g., Lucassen, Muilwijk, Noordzij, & Schraagen, 2013; Lucassen & Schraagen, 2013). In these cases, the researchers control the properties of the material by balancing or restricting the variation in quality that students are confronted with. In a realistic task, students might encounter SERPs that contain no relevant or trustworthy sources at all, or SERPs displaying an unwieldy large set of potential sources of varied and unknown quality. How students deal with such cognitively more demanding circumstances in a realistic setting remains unknown.

In addition, these studies isolate source evaluation as one fragmented skill and neglect the preceding action of formulating a search query. Search queries affect the composition of the SERP and thereby subsequent evaluation behaviors. IPS is an iterative process in which students switch between phases. At any point in the process they engage in concurrent cognitive processes such as deciding when to attempt a different search strategy, deciding on new search terms, reflecting on how much information is still needed, or deciding which information needs to be stored. For example, a student using irrelevant search queries is likely presented with a SERP containing few relevant sources. The student must then decide whether to continue exploring the SERP, choose a different search strategy (e.g., go to the library), try other search terms, switch search engines, review the task demands, etc. All these activities may impose additional cognitive demands affecting spontaneous evaluation processes. This implies that evaluation processes should be measured taking into account multiple iterations, and combined with other process data, such as the query quality. One method to achieve holistic data collection on IPS processes is using log files. In such cases, all user actions, such as entering search terms or following of hyperlinks, are automatically tracked, timestamped and stored in a log file for further inspection. For example, Bulger, Mayer, and Metzger (2014) have used log files to retrieve detailed information on the number of clicks and keystrokes, the number of times students visited a previously visited site, and the number of copy and paste actions. Such data collection methods lead to rich datasets that provide the necessary context to reconstruct an individual search process and study the student's evaluation behavior.

The process of solving an information problem, and more specifically *source evaluation*, is affected by many influencing factors, among which are factors associated with the student, the query, and the source. Examples of student-level factors are *prior knowledge*, *systematic searching*, and *systematic selecting*. An important query-level factor is the *query relevance*, and on the source-level, factors such as *SERP rank*, *time-on-source*, *source coverage*, and *source trustworthiness* are of interest. Of the factors associated with the student, domain-specific *prior knowledge* appears to be one of the most important characteristics affecting IPS performance (Monchaux, Amadieu, Chevalier, & Mariné, 2015). Searchers with little background knowledge in the domain in question appear to use less appropriate search queries and reformulate their queries more often, making small and ineffective changes (Hölscher & Strube, 2000; Monchaux et al., 2015). Those searchers also struggle more with evaluating sources (Salmerón, Kammerer, & García-Carrión, 2013), which leads to the selection of more irrelevant sources and a less optimal search path (i.e., more backtracking) (Hölscher & Strube, 2000). Brand-Gruwel et al. (2017) found that domain experts select more reliable sources. Based on these studies, a positive association is expected between prior knowledge and query relevance, a more systematic search process, and a better selection of sources in terms of relevance and trustworthiness.

Two other student-level aspects, *systematic searching* and *systematic selecting* were investigated in a study by Frerejean, Velthorst, van Strien, Kirschner, and Brand-Gruwel (2017). In this study, students received a blended IPS training integrated into the school curriculum. The training succeeded in making the students work more systematically when searching and when selecting sources. Students who searched systematically used sufficient, narrowly-scoped and relevant queries, and made logical adjustments during the process. Also, students who selected systematically persistently used sources of varying nature (i.e., pros, cons, different viewpoints and authors), and spent more time on trustworthy sources than untrustworthy ones. It can therefore be expected that students who search and select in a more systematic way, select sources that are more relevant and more trustworthy.

On the level of the query, research shows that students with more domain knowledge often create more *relevant queries* (Monchaux et al., 2015), because the student's mental model contains a more extensive list of relevant concepts and their synonyms. A student using the search terms 'operant' and 'conditioning' is more likely to find sources that are more relevant for a behavioral psychology task than a less knowledgeable searcher using the search terms 'learning', 'reward' and 'punishment', because the number of online sources containing the combined words 'operant' and 'conditioning' is smaller and those sources are probably more useful for the searcher than sources containing the words 'learning', 'reward', and 'punishment'. The latter query will lead to more sources, many of which will not deal with conditioning at all.

It follows that formulating relevant queries is an important aspect of effective IPS, and presumably leads to more relevant sources.

After the query is executed, a list of sources appears on the SERP. On the source-level, first, its *rank on the SERP* is of interest (Walraven et al., 2009). Search engines usually rank results based on the source's relevance, which is determined by a complex algorithm that scans the contents of each source and then matches this to the entered search terms. Previous research has shown that students come to expect that more relevant sources reside at the top of the SERP, and pay more attention to them than to sources lower on the SERP (Salmerón et al., 2013).

The source's relevance is the second variable of interest. In this study, relevance is defined as the amount of meaningful and useful information that is covered in the source, in the context of the current task. This is expressed as *coverage* and is determined by counting the idea units on the task topic. For example, an article on teenagers' sleeping habits would have a high coverage (i.e., many relevant idea units) for a task on the effects of media on sleep, but a low coverage for a task on the effect of violent videogames on aggressive behavior.

Third, the source's coverage can be distinguished from its *trustworthiness*. The degree of trustworthiness of a source indicates the quality or reliability of the information and of the site on which it resides, and depends on the assessment of several characteristics, including the expertise of the author, its publication date, whether the source was reviewed or edited, whether it contains references to other sources, and the quality of the argumentation and logic (Bråten, Strømsø, & Britt, 2009). While sources on the SERP are usually ranked by relevance, whether a source's trustworthiness is related to its ranking is an open question.

The final variable of interest at the source-level is the *time* a student spends on a source. In most authentic settings, students have limited time to solve the information problem and are stimulated to quickly decide whether a source is useful enough to invest time on. This should lead to higher time-on-source for high quality sources and lower time-on-source for low quality sources. However, it is unclear how much time students need (and take) to evaluate a source, and whether longer time-on-source is associated with a higher quality selection.

The present study

The present study is a first step towards exploring the characteristics of a typical authentic search process in a realistic setting where students are free to navigate the whole world wide web, and are required to perform their own search queries instead of receiving prefabricated SERPs and sources. While omitting instructions that interfere with students' natural behavior and by meticulously logging all their search activities, the collected data are used to model the students' source selection behavior. This study focuses on several aspects of the IPS process on the student level, query level, and source level: Student-level aspects consist of prior knowledge, systematic searching, and systematic selection. The query-level aspect consists of query relevance. Source-level aspects consist of the SERP rank, source coverage, source trustworthiness, and time-on-source. A detailed view of how these aspects relate to one another can provide a better understanding of the search process in a realistic setting, which can lead to valuable insights for the design of effective instruction or the design of assessment instruments.

The general objective of the study is to investigate how these student-level, query-level, and source-level characteristics are associated, and how they predict the selection of relevant and trustworthy sources. Three research questions are:

- RQ1: What are the general characteristics of a typical search process of (untrained) searchers, and how are they correlated?
- RQ2: How well do prior knowledge, systematic searching ability, systematic selecting ability, query relevance, SERP rank, time-on-source, and source trustworthiness predict the coverage (relevance) of a selected source?
- RQ3: How well do prior knowledge, systematic searching ability, systematic selecting ability, query relevance, SERP rank, time-on-source, and source coverage predict the trustworthiness of a selected source?

METHOD

PARTICIPANTS

For this study, 135 first-year student teachers in a Dutch teacher-training program served as participants. They received no prior training in IPS skills. Their average age was 19.01 years (SD = 1.65); 39 (29%) male, 96 (71%) female. In the remainder of this chapter, the term students refers to the student teachers who participated in this study.

PROCEDURE

Students participated in a 20-minute web search task in the computer rooms at their institute. The researchers explained the test procedure, instructed students to log in on an online environment, and demonstrated built-in functionalities such as mind mapping. After filling out a short demographic questionnaire, students received a problem description about the effects of mandatory school uniforms on bullying behavior. This task was chosen as it mimics a real-life IPS task students could receive during the educational program. In addition, the topic of the task matches their educational context, which should help motivate them and make the task more relevant. In the next four minutes, students created an online mind map of their prior knowledge, during which they were not allowed to consult online sources. Thereafter, the main assignment was explained. Students were given 20 minutes to collect information online that helped them solve the problem explained in the problem description, and store this information in a new, blank mind map. During this task, all measurable browser actions were recorded with a Firefox® browser plugin and stored in a log file for each student. After 20 minutes, the researcher indicated the end of the session, and instructed students to store their final mind map. This procedure, including the data collection and calculation of variables, was performed as a part of the study by Frerejean et al. (2017).

Variables

The data stored in the log files lead to the computation of eight variables. The researchers assessed the amount of *prior knowledge* and the students' *systematic searching* and *systematic selection* (i.e., student-level variables). For each generated query, the researchers assessed the *query relevance* (i.e., the query-level variable). For each visited source, the *SERP rank* was stored, the *time-on-source* in seconds was calculated, and researchers assessed the source's *coverage* and *trustworthiness* (i.e., the source-level variables).

Student–level variables

Prior knowledge was calculated based on the mind map at the start of the task. The researchers identified the number of task-relevant idea units in the map. This was then expressed as the percentage of the maximum number of idea units in the respective task. Interrater reliability as measured by intra-class correlation amounted to .989. *Systematic searching of information* was assessed by looking at the student's search process. Indicators of a good, systematic search process are: starting with a narrowly-scoped, relevant query, making small, logical adjustments to subsequent queries, using sufficient relevant queries to cover the complete problem, and correct use of Boolean operators. Based on these aspects, the researchers awarded a score between 0% and 100% with an intraclass correlation of .956. For assessing the *systematic selecting of sources*, a similar score between 0% and 100% was awarded, based on the indicators: using a sufficient number of sources (compared to the average of all the students), exploring more than just the top hits on the SERPs, persistently selecting high-quality sources, and the relative amount of time spent on high-quality sources. The intraclass correlation was .755.

Query–level variable

Query relevance was expressed as a score indicating how relevant the chosen terms in the search query were to the task. A matrix of terms was created based on an inductive-deductive process in which the researchers first formulated relevant queries of their own, and then amended this matrix with terms used by the students. Each relevant term in a query was awarded a score between 0 and 3, with a maximum of 9 points for the complete query. See Appendix 3 for the matrix. Intraclass correlation yielded an interrater reliability of .873. Some queries were discarded, for example when students made use of the Google™ spelling correction and no sources were visited, or when queries were task-irrelevant, such as entering 'google' in the browser quick search bar.

Source–level variables

SERP rank indicates the rank of the respective source on the list of search results. A standard Google™ SERP shows ten results on each page. A SERP rank of 1 indicates the student chose the top hit in the list, while a SERP rank between 11 and 20 indicates the student selected the source from the second page of the search results. *Time-on-*

source is the number of seconds the student dwells on the current source. A student might return several times to the same source, but time is only added when the source is open and active. No time is added when the source is open but inactive, for example in the background. The total time-on-source for each student differs, because students spend different amounts of time creating their mind map. Therefore, time-on-source was converted to a percentage score based on the total number of seconds the respective student spends on sources.

Source coverage is a measure indicating the amount of relevant information in the source regarding the current task. This is calculated by reading the source and counting the number of relevant idea units that appear in it (Wirth, Sommer, von Pape, & Karnowski, 2015). The total number of relevant idea units that were identified across all sources was 40. The coverage score was expressed as percentage of this maximum number of idea units. Interrater reliability was .989. A source with five relevant idea units therefore received a coverage score of five out of 40, or 12.5%. *Source trustworthiness* is the researchers' assessment of the trustworthiness of a source. The researchers analyzed the following aspects of a source: the author, the quality of the argumentation, the motive or goal of the source, and its layout, format and language, and used a scoring sheet (see Appendix 4) to classify each of the sources as untrustworthy, questionable, trustworthy, or very trustworthy. Interrater reliability amounted to .935.

Data analysis

The data collected for this study contains a clear hierarchical structure. Starting at the bottom of the hierarchy, a total of 247 unique sources were visited. Each of those page visits followed a query that was entered in the Google™ search engine. Therefore, each page visit is nested in one of the 366 unique queries, which form the second level of the hierarchy. In turn, each query is generated by one of the 135 students that participated in the study, which constitutes the top level of the hierarchy. Figure 5.1 shows an overview of this structure and displays on which levels the variables were measured. To take this hierarchy into account in the statistical analysis, three-level multilevel analyses were conducted to investigate the effects of the independent variables on the outcome variables. To answer research question 2 on the predictors of source coverage, source coverage served as the dependent variable in the analysis and all other variables as independent variables. To answer research question 3 on source trustworthiness, source trustworthiness served as the dependent variable and all other variables as independent variables. The analyses were conducted using R version 3.4.0 and the packages *nlme* version 3.1-131, *lme4* version 1.1-13, and *ordinal* version 2015.6-8 following guidelines published by Holmes Finch, Bolin, and Kelley (2014).

The conceptual framework in Figure 5.1 guided the testing of four separate models for the prediction of source coverage and source trustworthiness. First, an unconditional null model was tested and compared to a single-level linear regression model to investigate the need for multilevel modeling. Then, the following three models respectively added the student-level predictors, query-level predictors, and source-level predictors. To answer research question 3, which uses trustworthiness as an ordinal dependent variable, the same approach was followed, but with generalized linear modeling.

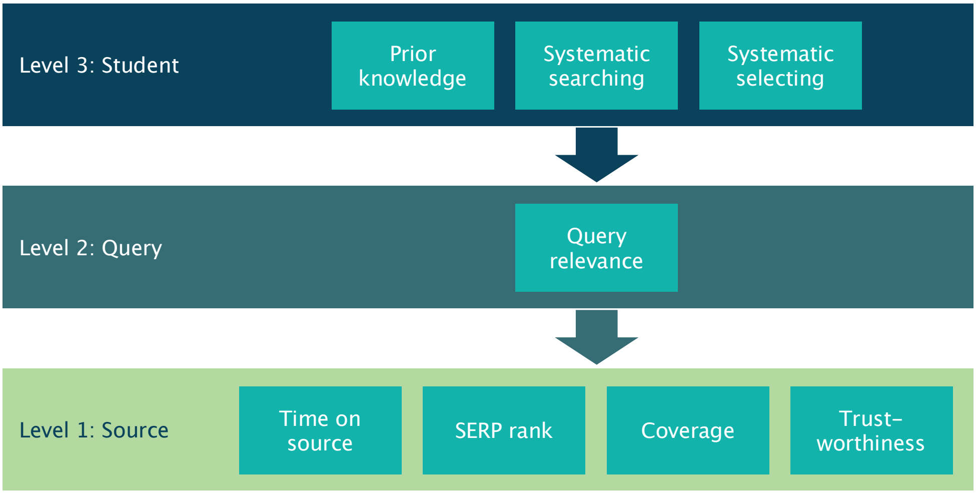


Figure 5.1. Overview of the hierarchy in the data, and the corresponding variables

RESULTS

GENERAL CHARACTERISTICS AND RELATIONSHIPS

The first research question explores the general characteristics and relationships between the measured variables. In total, the students produced 629 queries, of which 365 were unique. On average, a student produced 4.66 queries (SD = 2.46). These queries ultimately lead to 1061 page visits, on 247 unique sources. The average coverage of these unique sources was 11.29% (SD = 10.91). Their trustworthiness was distributed as follows: 98 sources (42.24%) were labeled as *untrustworthy*, 30 (12.93%) were labeled as *questionable*, 72 (31.03%) as *trustworthy*, and 32 sources (13.79%) as *very trustworthy*. For the remaining 15 sources, researchers were unable to visit the source to assess its trustworthiness. Students visited 7.8 sources on average (SD = 3.47), and lingered on a source for 65.75 seconds (SD = 40.83).

Table 5.1 reports means and standard deviations *per student* for the student-level variables, *per query* for the query-level variables, and *per visit* for the source-level variables (i.e., the 1061 page visits reported above). Table 5.1 also includes bivariate correlations of all variables measured on the source level. Because not all data were distributed normally, a nonparametric Spearman's rho is reported to express the correlations, which allows for direct comparison. For source trustworthiness, an ordinal variable, no mean and standard deviation is reported. Its distribution is: 334 visits (32.46%) to *untrustworthy* sources, 159 visits (15.45%) to *questionable* sources, 426 visits (41.40%) to *trustworthy* sources, and 110 visits (10.69%) to *very trustworthy* sources.

Table 5.1. *Descriptives and correlations*

		M	*SD*	1	2	3	4	5	6	7	8
	Student level										
1	Prior knowledge	7.78%	3.91	1							
2	Systematic searching	32.15%	14.51	-.06	1						
3	Systematic selecting	42%	19.53	.08*	.04	1					
	Query level										
4	Query relevance	30.67%	25.04	-.05	.41**	-.08**	1				
	Source level										
5	SERP rank	5.83	4.65	.04	-.07*	.19**	-.15**	1			
6	Time-on-source	12.28%	12.24	-.02	-.03	-.24**	.12**	-.21**	1		
7	Coverage	18.53%	12.16	.02	.04	-.11**	.17**	-.24**	.36**	1	
8	Trustworthiness	-	-	-.07*	-.01	.02	-.12**	.04	.09**	-.10**	1

** significant at the .01 level; * significant at the .05 level

The correlations show that prior knowledge is not associated with the other variables, apart from a positive correlation with systematic selecting and a negative correlation with trustworthiness. However, while these two correlations show statistical significance, the associations are minimal (all < .1). Systematic searching correlates strongly and positively with query relevance, which indicates that students who are good at searching systematically also generate better queries. It also correlates negatively with SERP rank, which indicates systematic searchers select sources at the top of the SERP hit list, but this is also a weak association. Students who select more systematically use queries that are less relevant, choose sources lower on the SERP, spend less time on sources, and select sources with lower coverage. Both systematic approaches do not correlate, indicating these are two distinct skills.

Query relevance correlates with all source-level variables. It is negatively associated with SERP rank, indicating that students who use better queries often do not explore more than the top hits on the SERP. The relationship between query relevance and time-on-source is positive, meaning that students spend more time on sources following relevant queries. Furthermore, more-relevant queries generally lead to sources with a higher coverage, but also to sources with a lower trustworthiness. The SERP rank correlates negatively with time-on-source and source coverage, indicating that relevant sources are generally at the top of the SERP and students spend more time on higher ranked hits. Interestingly, this is not true for trustworthiness, which does not correlate significantly to SERP rank. Sources with high coverage and trustworthiness scores are generally visited longer, but this association is much more pronounced for coverage. Finally, the negative correlation between coverage and trustworthiness indicates that more trustworthy sources generally contain less idea units relevant to the task.

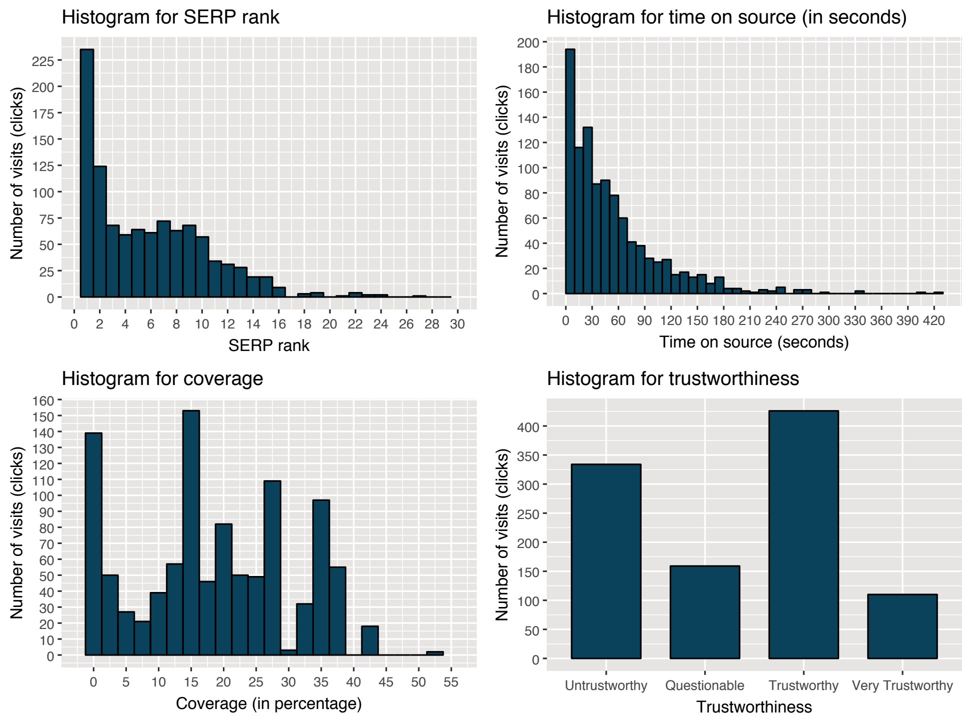


Figure 5.2. Histograms showing the distribution of page visits by (a) SERP rank, (b) coverage, (c) time-on-source, and (d) trustworthiness

To gain more insight in the characteristics of the sources and the pattern in page visits, Figure 5.2 displays histograms showing the number of page visits plotted against SERP rank, time-on-source, coverage, and trustworthiness. Figure 5.2a displays the number of page visits plotted against SERP rank, and reveals that students generally click the top hit in a SERP almost twice as often as the second hit. Subsequently, the second hit is clicked almost twice as often as hits 3 through 10. The number of clicks on hits on the second SERP page then quickly trails off. Figure 5.2b displays the page visits plotted against time-on-source. This plot shows that page visits of 30 seconds or less occur most often. The graph quickly declines, indicating that page visits of more than a minute do not occur very often, and page visits longer than 180 seconds are exceptional. Figure 5.2c displays the distribution of coverage, which shows no clear pattern. Most page visits occur on sources that are completely irrelevant and have a coverage score of zero, or on sources that are somewhat relevant and have 15% coverage score. Apart from a few exceptions, page visits in this task generally did not reveal more than 40% of relevant information. Figure 5.2d shows the distribution of trust, and indicates that students visited approximately as much untrustworthy and questionable sources as they visited trustworthy and very trustworthy sources.

PREDICTING SOURCE COVERAGE

NULL MODEL

The analysis began with a single level null model containing no predictors for coverage. The intercept of this model therefore corresponds to the overall mean score of coverage ($M = 18.53$). Comparing this model to a three-level random intercept model shows that the latter produces a significantly better fit to the data (see Table 5.2). Furthermore, an intraclass correlation of .22 indicates that 22% of variance is explained by the nested structure of queries in students. These findings warrant the use of multilevel modeling.

Table 5.2. *Comparing a single-level model to a three-level random intercept model*

Model	df	AIC	BIC	logLik	L.Ratio	p
Single level	2	8065.14	8075.01	-4030.57		
Three-level random intercept	4	8062.88	8082.63	-4027.44	6.2569	.04

MODEL 1

In the first step, student-level variables were included in the model as predictors. Source coverage was therefore predicted by prior knowledge, systematic searching, and systematic selecting. The intercept of 20.44 represents the mean coverage score of sources selected with average prior knowledge, systematic searching, and systematic selecting scores (see Table 5.3). Only systematic selecting is a statistically significant predictor in this model. Its negative coefficient of -0.09 indicates that higher scores on systematic selecting ability are associated with selection of sources with a lower coverage score. While the predictor is significant, its low value indicates the association is weak. Prior knowledge and systematic searching are not associated with source coverage. This model fits the data significantly better than the previous model.

MODEL 2

In the second step, the single query-level predictor is added to the model. Query relevance also contributes significantly to the model, with a small but positive slope. This indicates that relevant queries are associated with a small increase in source coverage. While the coefficients of the student-level variables change slightly, there is no change in the overall pattern. The comparison shows that model 2 fits the data significantly better than model 1.

MODEL 3

In the final step, source-level variables are added to the model: SERP rank, time-on-source, and trustworthiness. The negative coefficient of SERP rank indicates that sources with higher coverage scores are associated with higher places on the SERP. The positive association between coverage and time-on-source indicates relevant sources are visited longer. Trustworthiness was added as a dummy variable with the *untrustworthy* category as the baseline category. The results show that for the first two categories (i.e., *untrustworthy vs. questionable* and *untrustworthy vs. trustworthy*), there is no statistically significant association with coverage, meaning that choosing an *untrustworthy*, *questionable,* or *trustworthy* source is unrelated to its coverage score. However, compared to an *untrustworthy* source, a *very trustworthy* source is associated with a decrease in coverage. While query relevance remains a weak but significant predictor of coverage, systematic selecting is no longer significant in this full model. When compared to model 2, model 3 produces a significantly better fit to the data.

Table 5.3. *Estimates and standard errors from the random intercept model*

	Model 0	Model 1	Model 2	Model 3
Fixed effects				
Intercept (SE)	18.70 (0.43)	20.44 (1.64)	18.59 (1.62)	13.61 (1.68)
Student level				
Prior knowledge	-	0.12 (0.10)	0.13 (0.10)	0.12 (0.09)
Systematic searching	-	0.05 (0.03)	-0.02 (0.03)	0.01 (0.03)
Systematic selecting	-	-0.09 (0.02)***	-0.08 (0.02)***	0.02 (0.02)
Query level				
Query relevance	-	-	0.10 (0.02)***	0.06 (0.02)***
Source level				
SERP rank	-	-	-	-0.48 (0.08)***
Time-on-source	-	-	-	0.31 (0.03)***
Trustworthiness: questionable	-	-	-	1.89 (1.06)
Trustworthiness: trustworthy	-	-	-	-0.07 (0.82)
Trustworthiness: very trustworthy	-	-	-	-4.88 (1.21)***
Random effects				
Student	1.71	1.48	1.45	0.00
Query in Student	3.31	2.95	2.09	2.40
Residual	11.58	11.58	11.59	10.61
Model fit				
df	4	7	8	13
AIC	8062.9	8050.8	8024.9	7855.5
BIC	8082.6	8085.3	8064.4	7919.7
logLik	-4027.4	-4018.4	-4004.5	-3914.8
Deviance	8054.9	8036.8	8008.9	7829.5
p	-	<.001	<.001	<.001
ICC	.22	.20	.15	.18

*** significant at the .001 level; ** significant at the 0.01 level; * significant at the 0.05 level

Based on these findings, model 3 appears to provide the best fit to the data. In this model, none of the student level variables contribute significantly to predicting source coverage. Query relevance still significantly predicts source coverage, but weakly. When all other predictors are kept equal, a 10% increase in query relevance is associated with a 0.6% increase in coverage. As for SERP rank, results show selecting a source two spots lower on the SERP is associated with almost 1% decrease in coverage. Picking an untrustworthy, questionable, or trustworthy source is not significantly associated with coverage. However, picking a very trustworthy source instead of an untrustworthy source will likely yield a decrease in coverage of almost 5%. Further inspection of the model showed no significant outliers and influential cases or clusters. Even though some predictors (e.g., query relevance and systematic searching) were correlated, there were no issues detected with multicollinearity or heteroscedasticity.

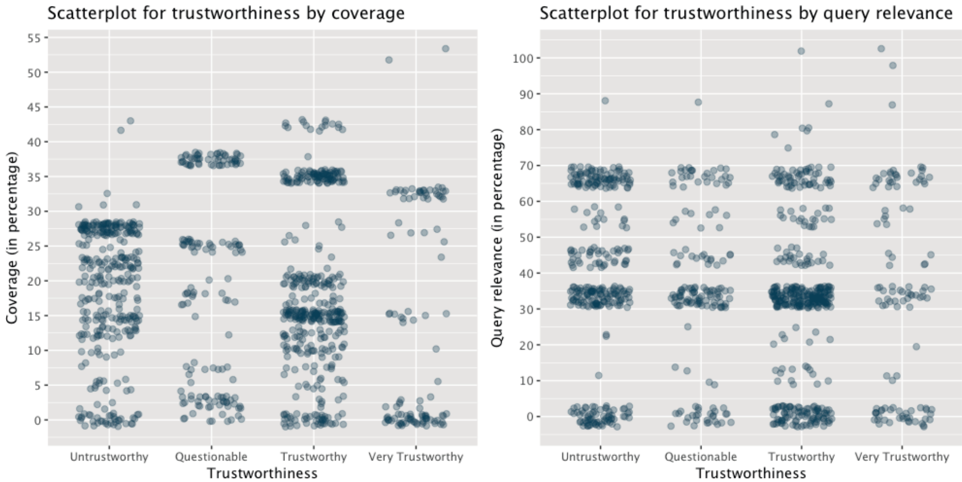


Figure 5.3. Scatterplots showing the relationship between coverage and trustworthiness (a), and query relevance and trustworthiness (b)

Figure 5.3a displays a scatterplot to shed more light on the relationship between coverage and trustworthiness. For untrustworthy, questionable, and trustworthy sources, coverage shows a similar variation between 0 and 45%. However, for very trustworthy sources, an almost dichotomous relationship appears: the source likely contains either quite some relevant information or it contains no relevant information. This effect may occur because a portion of these sources consists of abstracts of dissertations or theses, and only have a low number of idea units due to the short texts. In addition, some full texts, such as a research report on uniform use, are much longer and subsequently contain a higher number of relevant idea units, while others are more specialized and focus on a very narrow topic, leading to low coverage.

Predicting source trustworthiness

Null model

As trustworthiness is an ordinal variable with four categories, a multinomial logistic regression analysis was conducted. Following the same approach as with coverage, the analysis started with a single level null model containing no predictors for trustworthiness, and compared this model to a three-level random intercept model. While the intraclass correlation of .04 would not lead to suspect large grouping effects of queries and students, the three-level random intercept model did produce a significantly better fit to the data (see Table 5.4). This led the researchers to decide to use multilevel modeling, as was done in the previous analysis.

Table 5.4. *Comparing a single-level model to a three-level random intercept model*

Model	df	AIC	logLik	L.Ratio	p
Single level	3	2594.8	-1294.4		
Three-level random intercept	5	2582.2	-1286.1	16.562	<.01

Model 1

In the first model, the student-level predictors are included. Table 5.5 shows the estimates and their exponents, the odds ratios. In the case of logistic regression, these values indicate the change in the odds of the outcome occurring, associated with a unit change in the predictor. Odds ratios greater than 1 indicate that when the predictor increases, the odds of the outcome occurring also increase. Conversely, odds ratios less than 1 indicate that the odds of the outcome occurring decrease when the predictor increases. Prior knowledge, systematic searching, and systematic selecting appear to have no significant contribution to the prediction of trustworthiness. In addition, this first model provides no statistically significant better fit to the data than the null model.

Model 2

In the second model, query relevance is added as a predictor. The relevance of the query makes a significant contribution to the model. Here, the odds ratio is smaller than 1, indicating that a more relevant query *decreases* the odds of finding a source with higher trustworthiness. The inclusion of query relevance as a predictor in the model significantly improves the fit when compared to model 1.

Model 3

The final model includes all predictors, including the source-level predictors. SERP rank shows no significant association with the trustworthiness of a source. Time-on-source is a significant predictor of trustworthiness. Its odds ratio of 1.03 indicates that if time-on-source increases with 1%, the odds of finding a *questionable, trustworthy,* or *very trustworthy* source, are 1.03 times greater than finding an *untrustworthy* source. For this predictor, the correct interpretation is clearly reversed, as spending more time on a source does not make it more likely that it becomes more trustworthy.

Instead, the higher the trustworthiness of the source, the higher the odds that the students spend more time on the source. Coverage shows an odds ratio of .98, indicating that if coverage increases with 1%, the odds of finding a source in a trustworthiness category higher than the current one decrease with 2%.

Binomial regressions

The multinomial regression analysis reported in Table 5.5 provides a single coefficient for each of the predictors, which indicates the change in odds of reaching a higher category on the outcome. For further inspection, three consecutive binomial regressions were performed, comparing the baseline outcome category (i.e., *untrustworthy* source) with each of the other categories (i.e., *questionable, trustworthy,* and *very trustworthy*, respectively). These additional analyses provide a more detailed view of the relationship between the predictors and the outcome for each of the separate categories of sources. This was done following the same procedure as reported above: a first model was created with only student-level predictors, then the query-level predictor was added in a second model, and a third model incorporated all predictors. These full models are reported in Table 5.6.

When comparing *untrustworthy* sources with *questionable* sources, none of the predictors appear to contribute significantly, apart from time-on-source. This indicates that students spend slightly more time on questionable sources than on untrustworthy sources. Interestingly, effects of query relevance and source coverage do not yet manifest in this comparison.

More significant relationships become visible when comparing the two categories of sources that received the most page visits: *untrustworthy* sources and *trustworthy* sources. Again, time-on-source contributes significantly to the model, showing that students spend more time on the more trustworthy sources. But in this comparison, systematic selecting is now also a significant predictor. The odds ratio of 1.011 indicates that when the student has a higher score on systematic selecting, the chance of selecting trustworthy sources increases. Query relevance is another significant predictor, but as in the multinomial regression, it shows a negative relationship: searching with better queries actually slightly decreases the odds of selecting a *trustworthy* source instead of an *untrustworthy* source. Two other predictors are of interest. First, SERP rank shows a slight positive relationship with source trustworthiness, indicating that trustworthy sources are generally lower on the SERP than untrustworthy sources. However, this relationship did not reach statistical significance: $p = .063$. Second, prior knowledge shows a slight negative relationship with the outcome measure, indicating that students with more prior knowledge are slightly more likely to choose untrustworthy sources than trustworthy sources. Again, this finding did not reach statistical significance: $p = .058$.

Table 5.5. *Estimates, standard errors, odds ratios (OR) and confidence intervals (CI) from the random intercept model*

	Model 0	Model 1		Model 2		Model 3	
Student level			OR [CI]		OR [CI]		OR [CI]
Prior knowledge	-	-0.030 (0.019)	0.970 [0.934; 1.007]	-0.029 (0.019)	0.971 [0.936; 1.007]	-0.031 (0.019)	0.970 [0.935; 1.006]
Systematic searching	-	0.002 (0.005)	1.002 [0.991; 1.012]	0.008 (0.006)	1.009 [0.998; 1.020]	0.011 (0.006)	1.011 [1.000; 1.022]
Systematic selecting	-	0.002 (0.004)	1.002 [0.994; 1.010]	-0.000 (0.004)	1.000 [0.992; 1.008]	0.006 (0.004)	1.006 [0.997; 1.014]
Query level							
Query relevance	-	-	-	-0.011 (0.003)***	0.989 [0.983; 0.995]	-0.011 (0.003)***	0.989 [0.983; 1.006]
Source level							
SERP rank	-	-	-	-	-	0.016 (0.014)	1.015 [0.988; 1.045]
Time-on-source	-	-	-	-	-	0.029 (0.006)***	1.030 [1.018; 1.041]
Coverage	-	-	-	-	-	-0.015 (0.006)**	0.985 [0.974; 0.996]
Random effects							
Student	0.499		0.483		0.456		0.439
Query in Student	0.368		0.377		0.348		0.347
Model fit							
df	5		8		9		12
AIC	2582.2		2585.4		2574.4		2551.6
logLik	-1286.1		-1284.7		-1278.2		-1263.8
Likelihood ratio	16.562		2.759		13.008		28.871
p	<.01		.43		<.001		<.001
ICC	.04		.04		.04		.04

*** significant at the .001 level; ** significant at the 0.01 level; * significant at the 0.05 level

Table 5.6. *Estimates, standard errors, odds ratios (OR) and confidence intervals (CI) from the binomial regressions*

	Untrustworthy vs. Questionable		Untrustworthy vs. Trustworthy		Untrustworthy vs. Very trustworthy	
Student level		OR [CI]		OR [CI]		OR [CI]
Prior knowledge	-0.012 (0.026)	0.988 [0.938; 1.040]	-0.039 (0.021)	0.962 [0.924; 1.001]	-0.047 (0.048)	0.954 [0.869; 1.046]
Systematic searching	-0.009 (0.008)	0.991 [0.976; 1.006]	0.009 (0.006)	1.009 [0.996; 1.022]	0.025 (0.014)	1.026 [0.998; 1.054]
Systematic selecting	0.009 (0.006)	1.009 [0.997; 1.020]	0.011 (0.005)*	1.011 [1.001; 1.020]	-0.002 (0.011)	0.998 [0.977; 1.019]
Query level						
Query relevance	-0.007 (0.005)	0.993 [0.984; 1.002]	-0.152 (0.004)***	0.985 [0.977; 0.992]	-0.013 (0.007)	0.987 [0.973; 1.001]
Source level						
SERP rank	0.023 (0.023)	1.023 [0.978; 1.071]	0.030 (0.018)	1.030 [0.995; 1.067]	0.041 (0.031)	1.041 [0.980; 1.107]
Time-on-source	0.027 (0.010)**	1.028 [1.007; 1.049]	0.043 (0.008)***	1.044 [1.028; 1.061]	0.047 (0.012)**	1.048 [1.024; 1.073]
Coverage	0.018 (0.010)	1.018 [0.998; 1.038]	0.000 (0.007)	1.000 [0.986; 1.016]	-0.063 (0.012)***	0.939 [0.919; 0.961]
Random effects						
Student		0		0.187		0.912
Query in Student		0		0.438		0.306
Model fit						
df		10		10		10
AIC		619.72		1004.1		475.23
logLik		-299.86		-492.05		-227.62
ICC		0		.05		.03

*** significant at the .001 level; ** significant at the 0.01 level; * significant at the 0.05 level

When comparing the two endpoints of the scale: *untrustworthy* sources with *very trustworthy* sources, time-on-source remains a significant predictor, as was the case in the previous models. The effects of systematic selecting and query relevance no longer appear, but coverage is now a significant predictor with a negative coefficient. Sources with higher coverage scores are therefore slightly more likely to be *untrustworthy* than *trustworthy*. This finding is expected in the light of our previous discussion of the relationship between coverage and very trustworthy sources (also see Figure 5.3).

To conclude, source trustworthiness here is only associated with the relevance of the query, time spent on the source, and source's coverage. When comparing untrustworthy with very trustworthy sources, systematic selecting also significantly predicted trustworthiness. All relationships are quite weak, as indicated by odds ratios of approximately 1. Of note is that query relevance showed a negative relationship with trustworthiness. The bivariate relationship between these two variables is plotted in Figure 5.3b. While the graph does not directly show a negative relationship, two specific observations stand out: Firstly, the collection of data points in the top left indicates students clicked many *untrustworthy* sources after using relevant search queries. Conversely, some *trustworthy* sources were clicked after using irrelevant queries. This might explain why these two variables are negatively associated. As concluded earlier, in the case of *very trustworthy* sources, there appears to be a quite dichotomous relationship with coverage: the source is either contains much relevant information, or very little. The latter might be the result of sources that contain excerpts such as abstracts, and do not constitute full-texts. Further inspection showed no significant outliers and influential cases or clusters.

Discussion

This study explored the characteristics of a typical authentic search process in a realistic setting while logging all search activities. The collected data were used to investigate how student-level, query-level, and source-level aspects relate to one another to provide a better understanding of evaluation behavior in a typical search process. In addition, it was investigated how well prior knowledge, systematic searching ability, systematic selecting ability, query relevance, SERP rank, time-on-source, and source trustworthiness predict the coverage (relevance) of a selected source, and how well prior knowledge, systematic searching ability, systematic selecting ability, query relevance, SERP rank, time-on-source, and source coverage predict the trustworthiness of a selected source.

Typical search process

The results indicate that in this 20-minute task, students use approximately five queries on average. Students visit a source on average for 65 seconds, and use about eight sources, which translates as spending approximately 40% of their total time on processing and evaluating sources. The remainder of the time was spent on the SERP or creating their product. One may wonder how much deep processing of these sources takes place. It appears that the students were mostly concerned with data *collection* instead of processing, meaning that they appear to scan the text to retrieve elements that can be copied for storage (i.e., the mind map). Students did not synthesize this information and formulate an answer to the question, which are essential cognitive processes for domain-specific learning to occur.

Furthermore, students are unlikely to visit sources not on the first SERP page, which is in line with findings by Wu and Kelly (2014). An average source contains a little more than 10% of the total information (i.e., idea units) and the results show that students uncover approximately 18% of information per page visit, meaning that novices were able to select sources that were relevant to the task. Considering that many sources contain identical or similar idea units and only a small amount of unique information, consulting multiple sources is needed to retrieve sufficient information for a complete and well-informed answer to the problem at hand. To illustrate further, there was only one out of the 247 sources with just over 50% of the total number of idea units, showing that there is no perfect source with all needed information. Incidentally, approximately 50% of the selected sources are either untrustworthy or questionable. This leaves much to be desired in terms of source selection strategies. In addition, students in this sample have a very low prior knowledge, and do not innately work very systematically.

Predicting source coverage and trustworthiness

The finding that *prior knowledge* does not predict the amount of relevant information in a selected source, and did not correlate to any of the other predictors contradicts previous research that has established a positive relationship between prior knowledge and source relevance (Brand-Gruwel et al., 2017; Monchaux et al., 2015; Salmerón et al., 2013). Two lines of reasoning can explain this contradiction. First, as indicated above, students' average prior knowledge was very low in general, and such a small bandwidth of data created a restriction of range, preventing any solid findings. Second, the nature of the task is perhaps simple enough that prior domain-specific knowledge is not essential. The task might not provide a context sufficiently complex to require domain-specific terminology or domain models to generate relevant search terms or to understand and judge information in online sources.

The two other student-level predictors, systematic searching and systematic selecting, also do not predict any of the dependent variables. Systematic searching was positively correlated to query relevance, but this is to be expected as the scoring sheet for systematic searching partly assesses the logic in subsequent query formulations (see Appendix 3). Students with higher systematic selecting scores choose more sources lower on the SERP and spend less time on each source, which is possibly explained by the fact that they simply select *more* sources. Both systematic searching and selecting constitute indicators describing properties of the student and are therefore perhaps more associated with the efficiency rather than the quality of the selection of each individual source.

Query relevance shows to be a predictor of source coverage, indicating that better queries generally lead to a selection of sources with more information that is relevant to the task. The small effect size can be explained by the fact that this is an indirect effect. Good queries do not directly lead to sources with a higher coverage, they only lead to different SERPs. It is up to the student to select the relevant sources on that SERP. In addition, there is again some restriction of range, as average query relevance is only 30%. It would be interesting to see whether more pronounced effects occur with a wider range of queries, including more queries of higher quality. As for trustworthiness, a similarly small yet negative effect is present, indicating that a more relevant query does not lead to more trustworthy sources. This effect mainly occurs between *untrustworthy* and *trustworthy* sources, and might be explained by the fact that coverage and trustworthiness show a mutual negative correlation: better queries lead to sources that contain more information relevant to the task, but those sources are also generally less trustworthy.

SERP rank is a good predictor of source coverage, indicating that sources on the SERP are often ranked on the amount of relevant information with the most relevant sources on top (Kammerer & Gerjets, 2012). However, this is not the case for trustworthiness. The trustworthiness can vary even if a source with much relevant information is selected. In fact, while not significant, the pattern trends towards more trustworthy sources lower on the SERP. The correlation between query relevance and SERP rank indicates that students who use better queries generally choose sources located at the top of the SERP (i.e., the more relevant sources).

Time-on-source is one of the most important predictors in our analyses. It correlates positively with source coverage and source trustworthiness, and constitutes a significant predictor of both outcome variables. This implies that even though these students are novices, they recognize relevant and trustworthy sources and once they arrive on a such a source, they generally spend more time there. However, as time-on-source is clearly affected by the student's reading speed and the length of the source, and both variables were not measured in this study, caution is warranted when drawing conclusions from these findings.

Finally, there exists an interesting relation between a source's coverage and its trustworthiness. Especially very trustworthy sources appear to be hit or miss; they contain either much relevant information, or not much at all. This may be confounded by the accessibility of such sources, as students in the current study were sometimes limited to only abstracts or excepts, explaining some low coverage scores for those sources. Full-texts were not always available.

Implications and limitations

This study has several implications for practitioners, such as teachers or instructional designers of IPS training. First, the study reveals a significant distinction between relevance of a source and its trustworthiness. These are different things and students should be aware that sources with much relevant information can still vary greatly in trustworthiness. In fact, the findings could indicate that different strategies may be required depending on the goal of the search task. When it is essential to collect a broad range of content, it might be advisable to focus on relevant search queries and simply work down the SERP. However, when source quality is essential, different search strategies may be more useful, including different search engines, more elaborate evaluation processes, and more source exploration to identify the trustworthy ones. Students should learn that sources higher on the SERP generally contain more information that is relevant, but are not necessarily more trustworthy. Second, while this research did not focus on predictors of query quality, the results imply that query formulation is an important aspect in IPS and warrants training. Different queries affect the composition of the SERP, and thereby also the potential sources. Further research is needed to untangle these effects and investigate which characteristics of queries affect SERP content and source quality.

Third, though teaching students to work systematically might lead to a more efficient process, this study shows no association between systematic work and good source selection. It is therefore also important that teachers focus on teaching the correct cognitive strategies and heuristics for good IPS and for good source selection in particular. While research shows that evaluation behavior can be improved with training and digital tools (Mason, Junyent, & Tornatora, 2014; Stadtler & Bromme, 2008; Walraven, Brand-Gruwel, & Boshuizen, 2010, 2013), many approaches simply provide students with checklists and criteria to use while evaluating sources (Meola, 2004). Deeper insight into the search process can be beneficial to develop more specific guidelines for evaluation and selection strategies. For example Metzger (2007) proposed a sliding scale approach to teaching critical evaluation skills, meaning that teachers present a variety of approaches to source assessment fitting specific search tasks or situations. For example, when the quality of information is essential, students may adopt more elaborate and time-consuming strategies, such as using checklists or contextual models. When the task is more time critical and the quality of information is less important, some simple heuristics to determine source quality might suffice. Of course, learning when to use which strategy should be included in the instruction. Further research should indicate how different strategies can be optimized to finding relevant sources and to finding trustworthy sources.

Fourth, as teachers typically provide IPS tasks to students with the goal to acquire domain-specific knowledge from external sources, it might be useful to adopt a framework of searching as a learning process (Rieh, Collins-Thompson, Hansen, & Lee, 2016). In this study, the low amount of time students spend on an average source makes it unlikely that students learned much from those sources. When learning is the goal, it is essential that students are given sufficient time and opportunity to study and process their sources and to learn from this information. This is challenging when students lack the required search skills, as they are unlikely to find sufficient or correct information. In such cases, monitoring and providing cognitive feedback on students' search strategies will be necessary so the desired quality of information is collected. Discussion of collected information and generated solutions with peers and teachers can then stimulate critical thinking, processing, and evaluation, leading to the necessary construction and development of cognitive schema (Kong, 2014; Loo et al., 2016).

For researchers, the current study provides a good example of the breadth and depth of information stemming from using log files. They provide a rich source of data and can be used to answer many more research questions other than the ones asked in this study. Fine-grained logging leads to large datasets that, when analyzed with the right tools, can provide answers to a myriad of research questions. In theory, further development of this approach might lead to solutions that provide real-time logging of search processes, which can subsequently be presented on a dashboard visible to teachers. A live overview of the used search engines, queries, selected sources, and time-on-source can provide future teachers with sufficient information to quickly diagnose when and why students struggle, and make swift remediation possible. Secondly, it is advisable that researchers provide a clear definition of what constitutes a high-quality source. As this study shows, relevant sources are not necessarily trustworthy sources and vice versa. Relevance and trustworthiness appear to be two separate aspects of a source, making it important to underline which aspects of the source are investigated.

Finally, researchers should investigate aspects of IPS in the context of the whole task, and not limit investigations to merely providing fabricated SERPs or sources. Such part-task settings do not require the participants to apply their skills in a natural context, which can possibly bias the research results. Future research may include aspects of the IPS process that remain uninvestigated in the present study, such as problem definition skills, and solution formulation. In addition, prior research shows IPS processes are also affected by other factors, such as prior attitudes (Van Strien, Brand-Gruwel, & Boshuizen, 2014), epistemic beliefs (Kammerer et al., 2012; Scheiter, Gerjets, Vollmann, & Catrambone, 2009), and working memory capacity (Rosman, Mayer, & Krampen, 2016a). Future researchers are advised to include these measurements to complete the picture and correct for confounding aspects. While log files provide much information, they are limited to overt actions. To paint a more complete picture, covert actions and decisions should also be addressed, such as why some sources are *not* selected. Eye tracking and cued retrospective reporting may be used to answer this question, although there are drawbacks (Gerjets et al., 2011; Van Gog, Paas, van Merriënboer, & Witte, 2005)

The limitations described above prevent the generalization of these findings to other tasks, settings, subjects, and warrant further investigations. Future research should focus on further exploring the factors that influence the search process and evaluation behavior in particular, possibly generating new and improved models to show how these student-level, query-level, and source-level variables interact. While current IPS models can inform instructional design, they often provide high-level steps or phases to solve information problems, and describe a sequence of skills and activities that – when performed correctly – should benefit systematic task performance (Brand-Gruwel & Wopereis, 2006; Wopereis, Brand-Gruwel, & Vermetten, 2008). However, to optimize learning environments and maximize learning, educators need a more detailed view of what happens, or should happen, during the problem-solving process. The quality of the solution depends largely on making smart decisions on a more specific level, such as carefully choosing the most relevant queries for the specific task and thoroughly judging multiple aspects of a source to determine its usefulness.

Generating more specific strategies, approaches, and heuristics based on a solid understanding of the factors at play during source selection can further help teachers and instructional designers support students with query generation strategies (Hsu, Tsai, Hou, & Tsai, 2014; Lin & Xie, 2013), source selection strategies (Brand-Gruwel et al., 2017; Walraven et al., 2009), debiasing strategies (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012), and during task performance (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2016). This study has taken the first step to explore evaluation behavior in the context of the whole task. The next important step is to widen and deepen our understanding of this complex skill.

Chapter 6

General discussion

Students nowadays are increasingly being asked to self-regulate and self-direct their own learning, which typically includes searching and collecting online information sources to use in their study. For students who are unskilled in information problem solving (IPS), this is an inefficient and frustrating process as they have problems identifying sources of sufficient quality, which significantly hinders their learning. Research shows students of all ages struggle with IPS (Walraven, Brand-Gruwel, & Boshuizen, 2008), and it is therefore essential that IPS skills are incorporated in their education. This proves to be a major challenge, and IPS does not appear to be implemented sufficiently in most educational institutions in the Netherlands (Platform Onderwijs2032, 2016; Thijs, Fisser, & van der Hoeven, 2014) as well as abroad (Badke, 2010; Derakhshan & Singh, 2011; Probert, 2009). Practical guidelines for the design and implementation of effective IPS instruction are needed. Based on solid theoretical frameworks as the IPS-I model (Brand-Gruwel, Wopereis, & Vermetten, 2005; Brand-Gruwel, Wopereis, & Walraven, 2009) and the 4C/ID model (Van Merriënboer, Clark, & de Croock, 2002; Van Merriënboer & Kirschner, 2018), the studies in this dissertation demonstrate the application of instructional design principles by Merrill (2002) and investigate the effects of instructional interventions in realistic settings with the goal to arrive at best practices for the design of IPS instruction.

MAIN FINDINGS

The study presented in *Chapter 2* focused on the principle of application, which states that learning is promoted when learners are required to apply their knowledge and skills to solve problems (Merrill, 2002). More specifically, the study focused on methods of support during task performance. A two-hour online intervention was presented to first-year university students with the goal to investigate effects of different types of built-in task support. A training design employing the *completion strategy* was compared to a training design employing *emphasis manipulation*. The completion strategy entails a progression from completely worked-out problems, via intermediate completion problems (e.g., partially worked-out problems in which a part of the solution is given and a part is missing), to conventional problems containing no worked-out parts and which the student must carry out alone (Van Merriënboer, 1990; Van Merriënboer & de Croock, 1995). The completion strategy approach was compared to emphasis manipulation, an approach that places instructional emphasis on a different aspect of the complex skill in each learning task (Gopher, 2007; Gopher, Weil, & Siegel, 1989). The results of the experiments showed no clear benefit of either of these approaches, preventing any conclusions about the effectiveness of those types of task support. In fact, none of the conditions outperformed the control condition, who received no additional task support apart from being guided through the problem-solving phases. In addition, mental effort measurements during the learning phase revealed no differences in experienced mental effort. However, test scores increased significantly from pretest to posttest for all conditions, showing that a learning sequence consisting of a short instruction video, a modeling example, and four whole tasks is effective to foster IPS skills in the target group. It was hypothesized that the modeling example was responsible for a large part of the learning effect for the setting at hand. To verify this hypothesis, a follow-up study was conducted on the effects of the modeling example for teaching IPS skills.

Chapter 3 presents this follow-up study. Research shows that learning from examples can be an effective approach for teaching complex cognitive skills in ill-structured domains, such as academic writing (Braaksma, Rijlaarsdam, van den Bergh, & van Hout-Wolters, 2004). To explore whether this also holds for IPS, two more experiments were conducted using the same learning environment as in Chapter 2. To isolate the effects of the modeling example, all task support (i.e., emphasis manipulation, completion strategy, and guidance through the problem-solving phases) was stripped from the learning tasks. Results of these two experiments show that viewing a modeling example, presented as a screencast of an expert thinking out loud and interspersed with prompts, leads to a higher posttest performance than performing a practice task. The effect persisted on a delayed posttest one week later. Interestingly, posttest scores of the group receiving the modeling example approximate the posttest scores in the study reported in Chapter 2. Taking into account that all built-in task support was removed, the increase in scores can be explained by the instructional sequence consisting of an instructional video, a modeling example, and authentic learning tasks. The results make clear that a modeling example is an effective instructional method in this context, and illustrate that Merrill's (2002) principle of demonstration holds true for online IPS instruction. IPS instruction in an online setting can therefore benefit from employing video-based modeling examples.

Chapter 4 presented a study on embedded IPS training consisting of whole IPS tasks integrated in a 20-week course on vocabulary development, and its effects on student teachers' IPS skills. Skill measurements show that student teachers receiving the training search and select information more systematically, but their search queries, sources, and solutions are not of significantly higher quality than those of student teachers who received the regular course without IPS training. The training, thus, succeeded in developing cognitive strategies for solving an information problem, but did not improve all skills relevant to the IPS process. In addition, a delayed posttest showed the learning effects dissipated five weeks later. In conclusion, it appears the embedded, blended instruction using whole tasks to supplement domain-specific instruction shows potential, but did only create short-term learning effects. It was hypothesized that an educational program encompassing more opportunities for practice, using more varied learning tasks, is necessary to obtain the desired long-term results.

The study reported in Chapter 4 adopted a novel approach to measuring IPS skills in an ecologically valid setting. By employing log-file analysis, a fine-grained view of all actions during the IPS process was obtained. Instead of investigating an instructional principle, the study in *Chapter 5* attempted to obtain a state-of-the-art of students' search process using this innovative measurement method. Here, in a typical 20-minute search task, students used five queries, eight sources, and spend 65 seconds on a source. Approximately half of the visited sources were not trustworthy or were at best questionable, and half were trustworthy or very trustworthy. In addition, a selected source contained 18% relevant information. In an attempt to uncover predictors of source selection, the conducted analysis showed that query relevance was weakly but positively associated with the amount of relevant information in a source, but it was weakly and negatively related to the trustworthiness of a source. Sources with much relevant information generally reside at the top of the SERP, but this is not the case for trustworthy sources. Furthermore, there was a positive relationship between the time a student spent on a source and its relevance and trustworthiness. This study primarily showed students' search skills show much room for improvement, and that instructional interventions should incorporate well-tested strategies for generating effective search queries. In addition, it suggests the formation of different strategies for selecting relevant sources and trustworthy sources.

Before discussing the implications of these findings, two important general observations arise from these collective results. First, the studies in this dissertation provide examples of how online or blended whole-task instruction for IPS can be designed, developed, and integrated by applying well-established instructional principles. They show the principle of problem-centered instruction is valid for IPS instruction: acquiring the skill is promoted when learners work on realistic whole tasks that require integration and coordination of constituent skills. They also show the principle of demonstration is valid: demonstration by means of a video-based modeling example is useful for fostering IPS skills. In addition, the principle of application was valid and showed clear learning effects in the standalone studies, but could only create temporary learning effects in the embedded setting. Second, all student samples in these studies showed underdeveloped IPS skills. While these students, and also their teachers, might believe they are well-versed in finding information online, these beliefs do not correspond to their level of baseline performance. Students clearly overestimate their own IPS competence. This overestimation is mirrored in academia, where students are sometimes called *digital natives* and ascribed specific characteristics and sophisticated technological skills for which current education is unprepared (see: Prensky, 2001; Tapscott, 1999). Fortunately, this myth is now debunked (Kirschner & De Bruyckere, 2017; Kirschner & van Merriënboer, 2013; Margaryan, Littlejohn, & Vojt, 2011), and this dissertation adds strength to the claim that even the generations of students born in the nineties are not particularly well-skilled in finding and evaluating online sources for their own learning. Formal IPS instruction remains indispensable.

Implications

The research conducted for this dissertation specifically addressed the following four research questions:

- What are the effects of built-in task support (e.g., completion tasks, emphasis manipulation) on the acquisition of IPS skills?
- What are the effects of a modeling example on the acquisition of IPS skills?
- What are the effects of embedded IPS instruction on the acquisition of IPS skills?
- What are the general characteristics of novices' IPS process, and how do source, query, and student characteristics predict the selection of relevant and trustworthy sources?

Effects of built-in task support

The study described in Chapter 2 unfortunately does not provide a clear answer to the first research question concerning the differential effects of completion tasks, emphasis manipulation, or a combination of both. In this study, the *completion strategy* showed no benefits over applying *emphasis manipulation* or vice versa. This prevents formulation of useful guidelines detailing which form of task support should be adopted in practice and why. However, the study presented in Chapter 4, where a combination of both methods of task support were used in an embedded intervention, might reveal a disadvantage of the completion strategy. The results of the study showed that students hardly performed problem-definition activities. It is suggested that this may be due to the fact they had little opportunity to practice these skills. In the case where constituent skills show a temporal relationship and situations depend on previous actions (such as a SERP that depends on the search query, search terms depend on the question, etc.), providing a partly worked-out problem always means the earlier phases in the process need to be worked out. And while these worked-out phases have to be studied and understood before proceeding, they are not actively practiced. This creates an imbalance in the amount of practice dedicated to each of the constituent skills. While one may expect that students still learn from these worked-out problem definitions, research should investigate whether the lack of practice is actually decreasing the learning effect. For IPS, application of the completion strategy entails creating a series of learning tasks with *fading*, where the first step (i.e., problem definition) is almost always worked out, and later steps (i.e., presenting the solution), are almost always performed by the learner. While this approach has been effective in short interventions with only a handful of learning tasks (see Wopereis, Frerejean, & Brand-Gruwel, 2015, 2016), research on its effects in longer curricula is lacking.

Effects of a modeling example

The second research question involved the effects of a modeling example on the acquisition of IPS skills. As previous research in other domains has shown, learning from examples can be very powerful, even in ill-structured domains such as problem-solving (Pedersen & Lui, 2003) and academic writing (Braaksma et al., 2004). In the study presented in Chapter 3, a video modeling example proved more effective to foster IPS skills than a practice task. It is important to note, however, that the modeling example in question contained some built-in features to maximize its effectiveness. First, the example was designed as a screencast showing the screen of while an expert solved an information problem. When designing such videos as learning resources, care must be taken to incorporate relevant principles of multimedia learning to avoid any detrimental effects (Mayer, 2014). For example, the redundancy principle recommends avoiding presentation of on-screen text and narration simultaneously. A learner attending to the narration and written text simultaneously must coordinate both sources. This coordination unnecessarily stresses working memory capacity, which may hinder learning. When only one form is presented, there is no need for coordination, and more working memory capacity is freed to use for learning. Chapter 3 describes other principles of multimedia learning and how they were applied to the example.

Second, research on example-based learning has shown that in order for learning to occur, instructors should stimulate active processing of the example. This can be done by elaboration on the presented material, evaluating the process, or providing prompts that steer the learner's attention (Atkinson, Derry, Renkl, & Wortham, 2000; Braaksma, Rijlaarsdam, van den Bergh, & van Hout-Wolters, 2006; Renkl, 1999). The modeling example in question was split into fragments that were preceded by a prompt to activate prior knowledge, and followed by a *comparison prompt* stimulating learners to compare the demonstrated approach to their own approach.

Third, the modeling example was followed by a series of learning tasks in which learners had the opportunity to freely practice their newly learned skills. The results make clear that this carefully arranged instructional sequence was effective to foster IPS skills. This implies that educators willing to create modeling examples are advised to incorporate these additional activities to enhance learning. Further research is needed to disentangle this rich intervention to learn how much of the learning effect can be attributed to the example itself, the application of principles of multimedia learning, the prompting, or the practice. What remains clear, is that a 10-minute screencast of an expert solving an information problem following the IPS-I model (Brand-Gruwel et al., 2009), designed according to multimedia principles (Mayer, 2014), interspersed with prompts to activate prior knowledge and stimulate comparison of approaches, and followed by practice, is effective for fostering IPS skills.

Effects of embedded instruction

The research presented in Chapter 4 investigated IPS instruction embedded in a 20-week semester at a teacher training institute. An existing curriculum was redesigned to incorporate blended IPS instruction, comprising several IPS tasks in an online learning environment, discussions in face-to-face sessions, and a researcher-led session aimed at providing cognitive feedback on the IPS skills. Results show that the embedded IPS training was successful in developing students' systematic approach to solving information problems, but did not succeed in improving the separate IPS skills, such as query formulation or source selection. On a delayed posttest, no differences in performance were detected between the group with the embedded instruction and those following the regular curriculum. Therefore, the study did not show strong learning effects, which may be partially attributed to its quasi-experimental design in an ecologically valid setting, which caused some methodological issues. It did however demonstrate the application of well-established instructional principles to design task-centered instruction incorporating scaffolding, examples, cognitive feedback, and blended delivery of instructional materials. As such, it provides a case study of integrated *second-order scaffolding*.

While first-order scaffolding refers to the decreasing amount of support and guidance for learning the domain-specific content, Van Merriënboer and Kirschner (2017) refer to second-order scaffolding as decreasing amount of support and guidance for acquiring self-directed learning skills:

> For teaching self-directed learning skills, we speak about second-order scaffolding because it does not pertain to the complex cognitive skill that is being taught but to the self-directed learning skills intertwined with it. Basically, second-order scaffolding involves a gradual transition from teacher/system control to learner control, thus from adaptive learning to on-demand education, from planned information presentation to resource-based learning, from unsolicited to solicited information presentation, and from dependent to independent part-task practice. (p. 32)

As 21st century skills and self-directed learning skills are becoming increasingly more important, educational institutions should think about sustainable ways to incorporate these across their domain-specific courses and programs. As an example, longitudinal application of second-order scaffolding in a teacher training institute would imply that beginning students start with learning tasks and learning resources (e.g., lectures, text books) provided by the teacher, but as they progress, they gradually transition to settings in which the teacher no longer dictates the instructional sequence and they must select their own learning tasks (e.g., courses, workshops, activities) and find their own learning resources (e.g., trustworthy online materials, talk to expert teachers, etc.). It is essential that the institute provides sufficient support and guidance during this transition so that domain-specific skills and knowledge as well as the self-directed learning skills can be developed simultaneously.

Characteristics of a typical IPS process and evaluation behavior

In the study reported in Chapter 5, a novel assessment method was used to gain insight in students' IPS process, mainly focusing on evaluation behavior. It shows that while some IPS models provide a high-level overview of how to approach a problem and which steps and skills are required for an efficient and effective IPS process, a detailed view of the process reveals how aspects of the process, such as query relevance and the selection of relevant sources, are associated. This can inform further modeling of the process. More specifically, the study implies a source with a high *relevance* (i.e., its connection to the task) is not automatically highly *trustworthy*. In fact, in the study these two aspects displayed a negative correlation, indicating relevant sources were often less trustworthy and vice versa. While it is difficult to generalize from these findings, they might imply that instructors should consider developing more evaluation strategies beyond simply providing evaluation criteria and checklists (Meola, 2004; Metzger, 2007). Should future research confirm these findings, then it might be useful to think about devising a collection of source evaluation strategies tuned to specific task goals. For example, if gathering much information is the goal (i.e., source *relevance* is most important), different strategies may be preferred than when *quality* of information is most important. In addition, the findings of the study hint at the importance of query relevance for source evaluation. While it is often ignored, some research on query formulation and search skills is emerging (e.g., Smith, 2015, 2017).

Another interesting finding is that prior knowledge in this study was not associated with any of the other IPS aspects, while previous research shows it plays an important role and affects query generation and source evaluation (Monchaux, Amadieu, Chevalier, & Mariné, 2015; Salmerón, Kammerer, & García-Carrión, 2013). While there might be a restriction of range in prior knowledge scores preventing any effects in the presented study, it might also be the case that the effect of prior knowledge is attenuated by other factors outside the scope of this study. Future research should investigate whether task characteristics or other factors interact with the degree of prior knowledge and its effects on for query generation and source evaluation.

LIMITATIONS AND FURTHER RESEARCH

In each of the chapters, specific limitations and suggestions for future research are reported for the respective studies. In this section, general limitations are described for the dissertation as a whole.

TASK–CENTERED INSTRUCTION

The research in this dissertation employs task-centered instruction in all studies. The prevailing view is that instruction that presents learners with authentic and relevant tasks that require the integration and coordination of skills, knowledge and attitudes is most effective for teaching complex skills (Francom & Gardner, 2014; Merrill, 2002; Van Merriënboer & Kirschner, 2018; Wopereis et al., 2015). However, the effectiveness of task-centered instruction is not guaranteed by simply letting learners work on authentic tasks. To maximize learning, tasks and their goals, properties, and complexity, must be carefully selected or designed and sequenced. Research shows that search behavior, such as query length, query composition, time-on-task and time-on-source depend on the goals associated with the task (Athukorala, Głowacka, Jacucci, Oulasvirta, & Vreeken, 2016; Russell & Grimes, 2007). This indicates that students use structural problem features to select which strategies they employ for solving the problem. Presenting sufficient tasks with different goals and characteristics allows students to discriminate between surface and structural features and subsequently differentiate between strategies that are most effective for the problem type at hand. It allows for inductive learning, in which students generalize from concrete experiences. Such concrete experiences (i.e., the learning tasks) should vary on dimensions in which search tasks differ in practice. Providing sufficient opportunities for practice with sufficient variability in practice tasks is a necessary ingredient to achieve *transfer of learning*.

However, in the presented studies, all tasks were evaluation tasks of similar level of complexity. While tasks differed on surface features, they did not differ on structural features or in complexity. While variation of practice is essential, it requires time and exposure to many different problems. The presented instructional interventions were short in terms of time and number of learning tasks, which prevented the required variability in tasks and complexity to achieve long-term learning and transfer to other tasks. Although the current studies likely did not provide sufficient variability of practice, based on previous research it can be expected that a more longitudinal approach using more varied learning tasks would improve learning effects. This implies that teaching IPS and other self-directed learning skills should not be restricted to only a few tasks or courses, but preferably be a curriculum-wide approach. Instructional designers should keep in mind that different task characteristics can impose different amounts of cognitive load, and it is important that task complexity is carefully managed to avoid overloading the learners. Sequencing tasks from simple to complex is therefore essential.

Feedback

Another limitation in the presented studies is that the instruction adopted a one-size-fits-all approach. No attempts were made to tailor instruction and feedback to the specific needs of individual students. The 4C/ID model prescribes the presentation of *cognitive* and *corrective* feedback. The former is aimed towards enhancing learners' systematic approaches and cognitive strategies for execution of nonrecurrent skills, while the latter is aimed towards improving procedures for recurrent skills. Both types of feedback were not systematically incorporated in the interventions in this dissertation. Only the study in Chapter 4 included one cognitive feedback session. In this session, the researchers had evaluated students' performance during the first three IPS learning tasks, and discussed their findings with the group of students during a face-to-face session. This feedback session was an integral part of the instructional intervention, making its individual effects unclear. The lack of an essential instructional activity as feedback presents a limitation to these studies, and it is highly likely that inclusion of feedback on student performance during the IPS instruction would have had positive effects on learning outcomes (Timmers & Veldkamp, 2011; Timmers, Walraven, & Veldkamp, 2015). Future research should investigate how feedback on IPS skills can be optimized in blended or online environments.

Assessment

Assessment of IPS skills presents a real challenge. As it constitutes a higher-order skill, it is always performed in the context of the problem domain. Many aspects of such skills are therefore contextualized and do not necessarily carry over to other domains. For example, a competent information problem-solver in the medical domain might not be as competent when required to find and evaluate legal documents. Assessing IPS competence is, therefore, always contextualized. While some of those contextualized instruments exist (e.g., Rosman, Mayer, & Krampen, 2015), none fit the requirements in the presented studies, and new assessment instruments were developed as a result. These instruments were not submitted to rigorous validation studies, which presents a limitation, as the instruments' validity and reliability are therefore not fully guaranteed.

As the results of the final study make clear, the constituent skills in the IPS process are strongly interrelated. It is not possible consider IPS as the application of separate skills in a vacuum, instead one should look at it as an integrated and coordinated whole. Each skill in each phase of problem solving depends on what has occurred before. The learner's perception of the problem affects whether a (research) question is formulated, and which search strategies are selected (Athukorala et al., 2016). This in turn is likely to affect the amount of goal driven and systematic behavior (Russell & Grimes, 2007). The scope of the (research) question and the amount of prior knowledge steer the generation of search terms, and the relevance of the search queries determine the composition of the SERP and the returned list of sources. At this point, evaluation skills come into play. In certain tasks and contexts, the ability to judge the quality of sources may be affected by the students' domain-specific knowledge. In the end, the production of a satisfactory solution, and the amount of domain-specific learning that takes place depends on all of the above. Singling out specific aspects of this process without regard for the process as a whole may very well provide a distorted view of the student's search skills. Assessment of IPS skills should be done more holistically and across multiple contexts, taking into account the iterative nature of the process and the relationships between the constituent skills. It is up to future researchers and assessment experts to shine their light on how this may best be achieved.

Domain–specific learning

As this dissertation focuses on the development of IPS skills, it has largely ignored the acquisition of domain-specific knowledge during IPS. As indicated before, teachers often employ IPS tasks with the goal to teach a domain-specific subject. Future research should further investigate under which conditions domain-specific learning actually takes place. Logic dictates that relevant learning can only take place once the desired learning sources are located and retrieved, and therefore an effective IPS process is required to arrive at the *starting point* for learning. The time and amount of processing that is then spent on these sources is likely a key factor determining the amount of learning (i.e., elaboration). As indicated in Chapter 4, group discussion of the uncovered learning materials is one example of how teachers can stimulate critical thinking, processing, and evaluation of information within an IPS task.

In a literature review, Rieh, Collins-Thompson, Hansen, and Lee (2016) describe research in three domains, namely research on students' web searching in learning environments (i.e., *searching to learn*), research on interventions to improve students' search skills (i.e., *learning to search*), and research on design of search systems and tools. They argue to bring these three domains together and conceptualize *searching as a learning process*. In their conceptual model, different search activities can be linked to different levels of learning as described in Bloom's taxonomy (Anderson & Krathwohl, 2001). For example, they link the lower levels such as remembering and understanding to look-up search and simple fact finding, while higher levels such as evaluating or creating are linked to comprehensive search in which students learn to differentiate between different perspectives and search for diverse opinions and perspectives instead of mere correct answers. They advocate creating rich representations of tasks (e.g., goals, complexity, perception), users (e.g., motivation, prior knowledge, epistemic beliefs), queries (e.g., terms, complexity, reformulation patterns), and content (e.g., readability, length, quality) and connect these aspects to learning outcomes to gain a comprehensive holistic understanding of how learning occurs in IPS. This approach closely relates to the whole-task approach to instruction and assessment as described in this dissertation, and studying IPS from the framework of *searching as a learning process* appears an attractive endeavor.

Conclusion

Research on online IPS started in the nineties and now, 27 years later, it is still relevant due to the continuous technological developments that change the world wide web. There is reason to believe that these developments will continue in the foreseeable future, and therefore, research on IPS should continue as well. Current strategies for effective and efficient IPS will likely be outdated in a few years, when search engines have evolved and countless tools and new information sources have emerged. We must therefore continue this line of research to collect empirical evidence on which instructional interventions work and which do not, and expand these findings to other contexts, settings, and environments. We must further our understanding of factors affecting IPS and learning during IPS, so more detailed and elaborate models can be generated to serve as theoretical bases for the development of instruction. Such instruction should include high-level systematic approaches as well as strategies adapted to specific task demands and domains. We must continue to inform our teachers and equip them with the skills to design and provide IPS instruction to our students now and in the future. As a first step, this dissertation shows how well-established instructional principles can be applied to design of embedded, whole-task IPS instruction using modeling examples and scaffolding methods in an online or blended setting.

APPENDICES

Appendix 1

Scoring rubric for information problem solving assessment

Question 1: What is your first step and why? Maximum points: 2

0 points	for statements that reflect that the student starts searching right away
Add 1 point	for statements reflecting orientation activities: activating prior knowledge, planning, thinking, etc.
Add 1 point	for statements concerning task demands: determining information needs, types of sources, formulating a question, etc.

Question 2: Which problem statements would you formulate? Maximum points: 2

0 points	for statements that are irrelevant for the task
1 point	for statements that are relevant, but incomplete or formulated vaguely
2 points	for statements that contain all three relevant concepts (comparable to "What is the influence of X on Y?")

Question 3: Which search query would you type into Google? Maximum points: 4
Award a point for each relevant search term or synonym thereof. If the student shows a systematic search pattern, award an additional point.

Pretest	gender-specific education, influence, school performance
Posttest	intelligence, change, age
Delayed posttest	red wine, health, influence

Question 4: Which three websites would you select? Why? Maximum points: 4

Pretest	sources #3, #4, and #7 yield 2 points, sources #6 and #8 yield 1 point.
Posttest	sources #4, #5, and #6 yield 2 points, sources #3 and #8 yield 1 point.
Delayed posttest	sources #3, #6, and #8 yield 2 points, sources #4 and #5 yield 1 point.

If the sum of these points is 5 or 6, award 2 points for this question.
If the sum of these points is 2, 3, or 4, award 1 point for this question.
If the sum of these points is lower than 2, award no points for this question.
Award an additional point, but no more than 2 points, for all selection criteria that are mentioned in the comment that do not refer to "relevance". For example: reliability, author, publication date, reputation, etc.

Question 5: What do you do when you visit a text-rich website and want to find out if it contains relevant information? Maximum points: 1

1 point	for mentioning a scanning strategy, such as reading headlines only or using the search function (Ctrl + F)

Question 6: Which criteria do you use to determine whether information is useful for your task? Maximum points: 2

1 point	for each of the following criteria: goal of the text, reliability, author reputation, publication date, language/style, compares to other sources

Question 7: How do you deal with contradicting information? Maximum points: 1

1 point	for statements that reflect critical scrutiny, for example searching for more information or investigating reliability, or if the answer reflects that both sides of the story are incorporated in the solution.

Calculating the score

Subscore for step 1: Define the problem	The sum of scores for questions 1 & 2
Subscore for step 2: Search information	The score for question 3
Subscore for step 3: Select information	The score for question 4
Subscore for step 4: Process information	The sum of scores for questions 5, 6, & 7

Total score: The average of these four subscores forms the final score for the test and is expressed as a percentage of the maximum score (4 points)

APPENDIX 2

OVERVIEW OF MULTIMEDIA PRINCIPLES AND HOW THEY WERE APPLIED IN THE MODELING EXAMPLE

Multimedia principle	Description	Application
Split-attention principle	Avoid formats where learners have to split their attention between multiple sources	All information was contained in the video screen and zoomed on relevant elements where possible
Modality principle	Learners learn better from audio narration than on-screen text	The screencast contains a voice-over narration and no on-screen instructions
Redundancy principle	Avoid simultaneous presentation of verbal and visual text	On-screen text and verbal narration do not overlap
Segmenting principle	Media should be segmented or allow learner to process in own pace	The example was split into fragments and could be paused and replayed
Pre-training principle	Learners should be familiar with domain-specific key concepts	An instruction video prior to the modeling example explained all concepts
Coherence principle	Avoid extraneous, non-relevant material	Not possible to remove these elements from a realistic screencast, but zooming was used to focus on relevant information.
Signaling principle	Focus the leaner's attention to essential material.	The mouse cursor was accentuated and was often used to "point" at on-screen elements the expert was talking about. Zooming was used when possible to move distractions (such as advertisements) off-screen.
Contiguity principles	Visually and temporally align words and graphics	In addition to signaling methods, relevant information was always on-screen when it occurred in the narration
Personalization principle	Deliver instruction in a conversational tone	While scripted, the narration resembled an expert who thinks out-loud during the search
Voice principle	Learners learn better from narration in a standard-accented human voice	The speaker was a standard-accented Dutch woman.
Image principle	Adding the speakers image on screen does not necessarily lead to better learning	The image of the speaker was not included

APPENDIX 3

SCORING PROCEDURE FOR ASSESSING QUERY RELEVANCE

For each unique query, determine which concepts are used. Look up the concepts in the table below and add the corresponding points together for a maximum score of nine. Then calculate the average score for each student, expressed as a percentage (0-100).

Example query:
mandatory school uniforms help against bullying (verplichte schoolkleding helpt tegen pesten)
mandatory school uniforms (verplichte schooluniformen) = 3 points
help against (helpt tegen) = 2 points
bullying (pesten) = 2 points.

Total points for this query: 7 / 9 points (77.78%)

Pretest	Concept 1	Concept 2	Concept 3
3 points	Verplicht + uniform + school Schooluniform(en) Uniform(en) + school	Argumenten Discussie Voorargumenten Tegenargumenten Voordelen Nadelen	Onderzoek Gevolgen
2 points	Verplicht + kleding + school Verplichte schoolkleding Kledingvoorschriften + school	Pesten Pestgedrag Gelijkheid Kosten Duur	Preventie Helpt … tegen Oplossingen Voorkomen Tegengaan Minder Tegen
1 point	Kledingvoorschriften Kleding (+ school) Schoolkleding	Politiek Meningen Betoog Debat	Forum

Posttest	Concept 1	Concept 2	Concept 3
3 points	Mediagebruik Beelscherm Blauw licht	Nachtrust Kwaliteit van slaap Slaapgedrag Slaappatroon	Invloed (van) Effect Relatie Gevolgen Zorgen … voor Impact Discussie Onderzoek
2 points	Computer (gebruik) Laptop (gebruik) Smartphone (gebruik) Mobiel 9gebruik) Spelconsole iPad Computeren Internetten Gamen Televisie / tv	Slaap / slapen Slapengaan Slaapproblemen Slecht slapen Slaapritme REM-slaap Melatonine	Positieve effecten Negatieve effecten Positieve invloed Bevorderen Verbeteren Verstoren
1 point	Multimedia Appen Internet Social media Computerspellen Games	Slapeloosheid Concentratie (probleem) Opwindend Slaaptekort Slaapgebrek Hersenactivatie Hersenactiviteit	Voordeel Nadeel Storende factoren Risico’s Gevaren Schade Slecht (voor) Goed (voor) Minder

Delayed posttest	Concept 1	Concept 2	Concept 3
3 points	Verkeersveiligheid Verkeer + veilig(heid)	Navigatiesysteem Navigatieapparatuur GPS	Invloed Gevolgen Effecten Beïnvloeding Discussie
2 points	(Verkeers)ongeval (Verkeers)ongeluk Rijgedrag	TomTom Autonavigatie Navigatiemiddelen Kaart lezen	Voordelen Nadelen Afleiden Risico's Gevaar(lijk) Afweging Meningen over Verbetert Verslechtert Beter door
1 point	Veilig(heid) Veiliger Onveilig Gedrag in het verkeer Auto + ongelukken	Navigatie(gebruik) TomTomgebruik	Negatief Minpunten Goed Slecht

Scoring procedure for assessing systematic approach

Assess to which degree the instructions for systematically searching information are followed:

- Start with a narrowly scoped, relevant query
- Example: [influence navigationsystem on traffic safety] is more specific and focused than [navigation improves safety], though both contain three concepts.
- Subsequently make logical adjustments
- Example: following up [advantages navigationsystems] with [disadvantages navigationsystems] or [advantages GPS navigation] makes more sense than repeatedly switching to non-sequitur queries.
- Use sufficient queries to cover the problem domain
- As a rule of thumb, least three queries should be used to cover an acceptable part of the problem domain, while more than 10 queries might indicate the student is using a trial-and-error approach.

Weigh these criteria equally when determining the final score on a scale of 0 to 100. If Boolean operators are used incorrectly, deduct up to 10% of the final score.

Indicators of good performance	Indicators of bad performance
Starts with a narrowly-scoped, relevant, query	Starts with a query using one broad and common term
Makes small, logical adjustments to prior query	Queries seem random, trial-and-error, or repeat
Uses sufficient relevant queries	Uses not enough relevant queries or too many queries
Uses boolean operators correctly	Consistently uses Boolean operators incorrectly

Appendix 4

Scoring procedure for assessing source trustworthiness

To determine source trustworthiness, use the descriptions in the matrix below to choose the best-fitting label: *untrustworthy, questionable, trustworthy, very trustworthy*.

	Untrustworthy	Questionable	Trustworthy	Very trustworthy
Author	Students	Non-expert, (commercial) institutions	Expert, knowledgeable institutions	Expert, researcher
Argumentation, sources	Weak, no mention of sources	Questionable, little mention of sources	Adequate, unedited or unreviewed sources	Strong, edited source, references to research
Motive/goal	Giving opinion, writing for oneself or school	Informing, persuading, (subjectively) writing down existing knowledge	Transfer of knowledge, increasing own knowledge	Presenting new knowledge
Layout, format, language	Unstructured, sloppy, spelling mistakes	Adequately structured, readable text	Well-structured, edited copy	Well-structured, clear writing, edited copy
Typical type of source	Blog, personal texts, non-expert Example: Scholieren.com (discussion board where students post assignments to ask for feedback)	Commercial sites, magazines Example: Plazilla.com (blogging platform where everyone can share stories or articles)	National news outlets Example: Tweakers.net (technology website providing news, reviews, community)	Scientific or government sources Example: ProQuest.com (access to dissertations and theses)

Scoring procedure for assessing systematic approach

Assess to which degree the instructions for systematically selecting sources are followed:

- Carefully review the information in the results page (domain name, extension, snippet, etc.) and do not rely only on the top hits. Also, explore more than the first page.
- *Award points for exploring more than only top hits and first-page results.*
- When visiting a page, briefly scan the page by looking at headings and the introductory or concluding paragraph to assess its relevance. Check the author or publisher to indicate source quality.
- *Award points when the student spends more time on highly relevant and trustworthy sources and less time on irrelevant and untrustworthy sources.*
- Use sufficient sources to cover the problem domain.
- *As a rule of thumb, students using less than five sources are unlikely to cover sufficient information. Using more than 15 sources might indicate a superficial processing of the sources in a trial-and-error approach.*
- Keep track of the information you collected and select sources that contain additional relevant information instead of information you already know
- *Award points when the sources collected at the end of the task still contain new information. Award no points if the student reverts to low-quality sources to fill the time.*

	Indicators of good performance	Indicators of average performance	Indicators of bad performance
Number of sources	Average	Average	Much more or less than seems necessary
Variation	Explores more than the top hits and visits subsequent result pages	Explores more than just the top hits in the SERP	Clicks only top hits in SERP
Persistence	Finds high quality sources during the whole task	Finds most high-quality sources at the beginning of the task, less at the end.	Finds only high-quality sources at the beginning of the task, none at the end.
Judgment	Quickly discards low-quality sources and spends most time on high-quality sources	Spends more time on high-quality sources than on low-quality sources	Spends equal time on all sources, or more time on low-quality sources

SUMMARY

Consider a teacher in a classroom of student teachers, discussing whether reading aloud to young children is an effective method to increase their vocabulary. The teacher helps the group formulate questions and instructs them to find answers by gathering information from online sources. The next time they meet, the students present their findings. This is a prime example of an authentic setting where students are required to find their own learning materials. The rise of the Internet has provided quick and easy access to a wealth of online information sources that can serve as learning resources, making students less dependent on the library. But where libraries have gatekeepers to guard against low-quality information, the Internet does not. Anyone can fill the web with anything ranging from completely correct and reliable information to false or fake information with absolutely no reliability. Apart from that, the world wide web is a world where numerous actors constantly compete for the attention of the visitor, such as commercial companies that want to sell products, news agencies that to lure users to their articles, and Internet companies that want clicks on their advertisements. On top of that, trustworthy information and scientific publications are often locked away behind a paywall. Too often it is assumed that students are *digital natives* who know how use digital tools and distinguish trustworthy from untrustworthy information, but in reality, they are unprepared and unaware of who to trust or what to believe.

Finding information online for educational purposes constitutes a complex task that requires knowledge, skills, and attitudes to perform correctly. This is often called information literacy or information problem solving (IPS; Brand-Gruwel, Wopereis, & Vermetten, 2005). Successful IPS starts with a problem definition, where searchers familiarize themselves with the problem and its domain to establish which information they already know. They consequently determine which information is still needed in order to produce a satisfactory answer. Ideally, searchers then formulate one or more specific questions to guide their search process. Such a goal-driven approach using focused questions helps searchers stay on topic and recognize when they have gathered sufficient information. After defining the problem, the searcher decides on the best approach to collect the needed information. In most cases, they will be using a search engine such as Google™ to find online resources. The problem description, the formulated questions, and a searcher's background knowledge can be used to generate specific search terms to use in the search engine. It is important that searchers understand how search engines work, to determine which search strategies or combination of search terms are likely to lead to the best results.

Executing a search then results in a search engine results page (SERP), generally containing 10 links to information sources accompanied by the resource page's title, URL, and a small snippet of text. Careful evaluation of these information elements is needed to judge which seem useful enough to click on. Making good choices on a SERP avoids wasting time on irrelevant or untrustworthy sources of information and makes for an efficient IPS process. After accessing a source, its relevance and trustworthiness needs to be evaluated. *Relevance* relates to the amount of *on-topic* or *sought-for* information, while *trustworthiness* is an indication of the *reliability* of the information, determined by its publication date, author's expertise, reputation of the author's affiliation, quality of argumentation, etc. Searchers iterate between search queries, SERPs, and information sources to extract the information deemed useful for solving the problem. In the end, the searcher processes the information to construct new knowledge and formulate an answer to the question(s). In the educational context, these solutions are often presented as reports, presentations, or essays.

Research exploring the problems students encounter with IPS shows there are major deficiencies in all aspects in people of all age groups (Walraven, Brand-Gruwel, & Boshuizen, 2008). This once more indicates that IPS is a complex cognitive skill, and instruction should focus on teaching these skills and subskills in an integrated and coordinated fashion, working on developing the necessary declarative knowledge, practicing the necessary skills, and forming the right attitudes in order to learn how to deal with new information problems. While its importance as an essential 21st century skill is widely acknowledged, most educational institutions struggle with the application of instructional guidelines and encounter problems with the implementation of IPS instruction (Platform Onderwijs2032, 2016; Thijs, Fisser, & van der Hoeven, 2014). This often leads to subpar instruction in short library training sessions that forego the benefits of whole-task instruction. Embedding IPS instruction within a meaningful context, presenting it simultaneously with domain-specific instruction can lead to deeper learning and improved transfer (Perin, 2011; Wopereis, Brand-Gruwel, & Vermetten, 2008). Research on instructional interventions for IPS often focus on a subset of the constituent skills, or do not let learners apply their IPS skills in an authentic context, for example by restricting the number of potential information sources or making use of prefabricated SERPs (e.g., Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Gerjets, Kammerer, & Werner, 2011). In addition, many of the studies on IPS interventions focus on short-term learning effects and lack measurements of transfer or delayed learning effects.

The aim of the research carried out for this dissertation is to investigate Merrill's instructional design principles in order to formulate practical guidelines for teachers and instructional designers who wish to design instruction for effective and efficient IPS. It attempts to overcome shortcomings of previous research and focuses on the application of IPS skills in ecologically valid and realistic settings, making use of authentic learning tasks that require integration of the skills, knowledge, and attitudes necessary for effective and efficient IPS. In addition, the instructional interventions presented incorporate measurements of transfer or delayed learning effects. More specifically, the following research questions are addressed:

- What are the effects of built-in task support (e.g., completion tasks, emphasis manipulation) on the acquisition of IPS skills?
- What are the effects of a modeling example on the acquisition of IPS skills?
- What are the effects of embedded IPS instruction on the acquisition of IPS skills?
- What are the general characteristics of students' IPS process, and how do student, query, and source characteristics predict the selection of relevant and trustworthy sources?

The research presented in *Chapter 2* investigates the principle of applying new knowledge and skills while receiving *built-in task supp*ort during task performance. More specifically, two approaches to task support are compared in a standalone online IPS training using whole tasks: The completion strategy, a sequence of learning tasks containing a decreasing number of worked-out steps, is compared to emphasis manipulation, an approach where students receive additional support on a single aspect of the task in each learning task. The results of the experiments showed no clear benefit of either of these approaches, preventing any conclusions about the effectiveness of those types of task support. In fact, none of the conditions outperformed the control condition, who received no additional task support apart from being guided through the problem-solving phases. In addition, mental effort measurements during the learning phase revealed no differences in experienced mental effort. However, test scores increased significantly from pretest to posttest for all conditions, showing that a learning sequence consisting of a short instruction video, a modeling example, and four whole tasks is effective to foster IPS skills in the target group. It was hypothesized that the modeling example was responsible for a large part of the learning effect for the setting at hand. To verify this hypothesis, a follow-up study was conducted on the effects of the modeling example for teaching IPS skills.

Chapter 3 presents a study investigating the principle of demonstration. Using the same online training as described in Chapter 2, students receiving a video *modeling example* were compared to students performing a practice task. For these two studies, an online learning environment was developed consisting primarily of web search tasks and video materials for support. The studies reported in Chapters 2 and 3 were implemented as standalone training sessions as part of a university curriculum. Results of the experiment show that viewing a modeling example, presented as a screencast of an expert thinking out loud and interspersed with prompts, leads to a higher posttest performance than performing a practice task. The effect persisted on a delayed posttest one week later. Interestingly, posttest scores of the group receiving the modeling example approximate the posttest scores in the study reported in Chapter 2. Taking into account that all built-in task support was removed, the increase in scores can be explained by the instructional sequence consisting of an instructional video, a modeling example, and authentic learning tasks. The results make clear that a modeling example is an effective instructional method in this context, and illustrate that Merrill's (2002) principle of demonstration holds true for online IPS instruction. IPS instruction in an online setting can therefore benefit from employing video-based modeling examples.

In contrast to the standalone sessions in Chapters 2 and 3, the study reported in *Chapter 4* deals with whole-task instruction embedded in an existing educational program. As such, it deals with the principle of *application* in an ecologically valid setting. In this study, an existing curriculum in a teacher training program was partly redesigned to include embedded whole-task IPS training. As the original program offered mainly face-to-face education, it was decided to include a parallel online environment for practicing IPS tasks. The resulting blended learning setting was evaluated by comparing students receiving the regular curriculum with students receiving the redesigned curriculum including IPS training. Students' performance on authentic tasks was assessed by logging and retrospectively scoring all learner actions, such as selected sources and generated queries. Skill measurements show that student teachers receiving the training search and select information more systematically, but their search queries, sources, and solutions are not of significantly higher quality than those of student teachers who received the regular course without IPS training. The training, thus, succeeded in developing cognitive strategies for solving an information problem, but did not improve all skills relevant to the IPS process. In addition, a delayed posttest showed the learning effects dissipated five weeks later. In conclusion, it appears the embedded, blended instruction using whole tasks to supplement domain-specific instruction shows potential, but did only create short-term learning effects. It was hypothesized that an educational program encompassing more opportunities for practice, using more varied learning tasks, is necessary to obtain the desired long-term results.

The study presented in *Chapter 5* further applied this method for assessment to provide detailed insight on students' search skills. It elaborates on the method applied in Chapter 4 and uses the collected log files to perform a deep inspection of students' search processes. Here, in a typical 20-minute search task, students used five queries, eight sources, and spend 65 seconds on a source. Approximately half of the visited sources were not trustworthy or were at best questionable, and half were trustworthy or very trustworthy. In addition, a selected source contained 18% relevant information. In an attempt to uncover predictors of source selection, the conducted analysis showed that query relevance was weakly but positively associated with the amount of relevant information in a source, but it was weakly and negatively related to the trustworthiness of a source. Sources with much relevant information generally reside at the top of the SERP, but this is not the case for trustworthy sources. Furthermore, there was a positive relationship between the time a student spent on a source and its relevance and trustworthiness. This study primarily showed students' search skills show much room for improvement, and that instructional interventions should incorporate well-tested strategies for generating effective search queries. In addition, it suggests the formation of different strategies for selecting relevant sources and trustworthy sources.

Unfortunately, the study described in Chapter 2 does not provide a clear answer to the first research question concerning the differential effects of completion tasks, emphasis manipulation, or a combination of both. Further research is needed to draw solid conclusions about the effectiveness of different types of task support, preferably in longer curricula. Concerning the second research question involving the effects of a modeling example on the acquisition of IPS skills, the study in Chapter 3 shows that a 10-minute screencast of an expert solving an information problem following the IPS-I model (Brand-Gruwel, Wopereis, & Walraven, 2009), designed according to multimedia principles (Mayer, 2014), interspersed with prompts to activate prior knowledge and stimulate comparison of approaches, and followed by practice, is effective for fostering IPS skills. Therefore, teachers and instructional designers are advised to incorporate well-designed examples in their IPS instruction. The third research question on embedded IPS instruction was investigated in Chapter 4, but the study did not show strong learning effects. It did however demonstrate the application of established instructional principles to design task-centered instruction incorporating scaffolding, examples, cognitive feedback, and blended delivery of instructional materials in an ecologically valid setting. Future research will have to further investigate best practices for embedding IPS instruction in existing curricula in a way that domain-specific content as well as self-directed learning skills can be developed simultaneously. The final research question investigated students' natural search process and source evaluation behavior. The findings in Chapter 5 mainly have implications for assessment and modeling of the IPS process, but also show that search and evaluation strategies could be tailored to the task goals.

Current strategies for effective and efficient IPS will likely be outdated in a few years, when search engines have evolved and countless tools and new information sources have emerged. We must therefore continue this line of research to collect empirical evidence on which instructional interventions work and which do not, and expand these findings to other contexts, settings, and environments. Such instruction should include high-level systematic approaches as well as strategies adapted to specific task demands and domains. We must continue to inform our teachers and equip them with the skills to design and provide IPS instruction to our students now and in the future. As a first step, this dissertation shows how instructional principles can be applied to design of embedded, whole-task IPS instruction using modeling examples and scaffolding methods in an online or blended format.

SAMENVATTING

In een lerarenopleiding voert een klas een discussie met hun docent over de effecten van voorlezen op het vergroten van woordenschat bij jonge kinderen. Samen met de docent stellen de studenten concrete vragen op, en vervolgens worden ze geacht zelfstandig op zoek te gaan naar informatiebronnen waarmee ze deze vragen kunnen beantwoorden. De volgende keer dat de klas bij elkaar komt, presenteren de studenten hun bevindingen. Dit schetst een authentieke situatie in het hedendaags onderwijs waarin studenten op zoek moeten naar hun eigen leermaterialen. Door de groei van het internet hebben zij tegenwoordig eenvoudig en snel toegang tot een grote hoeveelheid informatie die als leermateriaal kan dienen, en zijn ze minder afhankelijk van de bibliotheek. Maar waar bibliotheken poortwachters hebben die waken over de kwaliteit van de bronnen, heeft het internet dat niet. Iedereen kan informatie op internet plaatsen, variërend van feitelijk correct en betrouwbaar tot volstrekt onbetrouwbaar nepnieuws. Daarnaast is het world wide web een plaats waar personen en organisaties voortdurend om de aandacht van de bezoeker strijden, zoals commerciële bedrijven die producten willen verkopen, mediabedrijven die bezoekers naar hun artikelen willen lokken, en internetbedrijven die kliks op hun advertenties willen. Daarbij worden betrouwbare bronnen en wetenschappelijke publicaties vaak verscholen achter een betaalmuur. Te vaak neemt men aan dat studenten *digital natives* zijn, die weten hoe ze digitale middelen moeten gebruiken en hoe ze betrouwbare en onbetrouwbare informatie kunnen onderscheiden. In de praktijk zijn ze onvoorbereid, en weten ze niet wie te vertrouwen en wat te geloven.

Het zoeken en vinden van online informatie voor educatieve doeleinden is een complexe taak die kennis, vaardigheden en attituden vereist. Vaak wordt hiervoor de term informatievaardigheden gebruikt, of *information problem solving:* het oplossen van informatieproblemen (IPS; Brand-Gruwel, Wopereis, & Vermetten, 2005). Een succesvol proces start met een probleemdefinitie, waar de zoeker zichzelf vertrouwd maakt met het probleem en het domein en nagaat welke informatie al bekend is en welke informatie nog gezocht moet worden om tot een adequaat antwoord te komen. Daarna formuleert de zoeker idealiter één of meerdere specifieke vragen om het zoekproces te sturen. Een doelgerichte aanpak met gerichte vragen is nuttig om afdwalen te voorkomen en om te bepalen wanneer er voldoende informatie verzameld is. Nadat het probleem gedefinieerd is, kiest de zoeker de beste aanpak om de benodigde informatie te verzamelen. In de meeste gevallen zal hij of zij een zoekmachine zoals Google™ gebruiken. De probleemstelling, de geformuleerde vragen, en de achtergrondkennis worden vervolgens gebruikt om specifieke zoektermen te genereren. Hierbij is het belangrijk dat de zoeker weet hoe zoekmachines werken, zodat de juiste zoekstrategie kan worden gebruikt en de juiste combinatie van zoektermen kan worden gemaakt om tot de gewenste resultaten te komen.

Het uitvoeren van de zoekopdracht levert vervolgens een resultatenpagina (*search engine results page: SERP*) op met doorgaans 10 links naar informatiebronnen, inclusief de paginatitel, de URL, en een korte uitsnede van de tekst. De zoeker dient deze informatie kritisch te evalueren om te bepalen welke link nuttig genoeg lijkt om aan te klikken. Door op dit punt al de juiste keuzes te maken kan de zoeker veel tijd besparen en irrelevante of onbetrouwbare bronnen vermijden. Eenmaal op een informatiebron is het zaak de relevantie en betrouwbaarheid te bepalen. De *relevantie* verwijst naar de hoeveelheid informatie die aansluit bij het onderwerp of waarnaar gezocht wordt. De *betrouwbaarheid* wordt bepaald aan de hand van indicatoren als publicatiedatum, de expertise van de auteur, de reputatie van de organisatie achter de bron, de kwaliteit van de argumentatie, etc. Zoekers pendelen tussen zoekopdrachten, SERPs, en informatiebronnen om die informatie te verzamelen die zij bruikbaar achten om het probleem op te lossen. Uiteindelijk verwerken zij die informatie om nieuwe kennis op te doen en een antwoord te formuleren op hun vragen. In de schoolse setting betekent dat vaak dat ze hun oplossing presenteren in een rapport, presentatie, of essay.

Onderzoek naar de problemen die studenten ervaren met IPS toont aan dat er grote tekortkomingen zijn in alle aspecten en in alle leeftijdscategorieën (Walraven, Brand-Gruwel, & Boshuizen, 2008). Dit toont andermaal aan dat IPS een complexe cognitieve vaardigheid is met verscheidene subvaardigheden die op een geïntegreerde en gecoördineerde wijze dienen te worden onderwezen. Daarbij moet aandacht zijn voor het ontwikkelen van de benodigde declaratieve kennis, het oefenen van de vaardigheden, en het formeren van de gewenste attitudes zodat studenten in staat zijn ook nieuwe informatieproblemen aan te pakken. Het belang van informatievaardigheden als 21^{e}-eeuwse vaardigheid wordt algemeen erkend, maar de meeste onderwijsinstellingen worstelen met het toepassen van onderwijsprincipes en ervaren problemen met de implementatie in hun curriculum (Platform Onderwijs2032, 2016; Thijs, Fisser, & van der Hoeven, 2014). Dit leidt vaak tot suboptimale instructie in korte bibliotheektrainingen die voorbijgaan aan de voordelen van een hele-taak benadering. Het integreren van informatievaardighedenonderwijs in een betekenisvolle context waarbij het gelijktijdig aangeboden wordt met domeinspecifiek onderwijs kan leiden tot dieper leren en een betere transfer (Perin, 2011; Wopereis, Brand-Gruwel, & Vermetten, 2008). Onderzoek naar informatievaardighedenonderwijs richt zich vaak op een onderdeel van de samenstellende vaardigheden, of laat studenten niet werken aan authentieke taken. In dat laatste geval wordt vaak het aantal potentiële informatiebronnen beperkt, of worden gefabriceerde SERPs gebruikt (zie Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Gerjets, Kammerer, & Werner, 2011). Daarbij richt onderzoek zich vaak op leereffecten op de korte termijn en ontbeert het toetsing van transfer of leereffecten op de lange termijn.

Het doel van het onderzoek dat is gedaan in het kader van dit proefschrift is om instructieprincipes te onderzoeken om tot praktische richtlijnen te komen voor onderwijzers en instructieontwerpers die zich bezighouden met het ontwerpen van effectief en efficiënt informatievaardighedenonderwijs. Het onderzoek poogt de tekortkomingen van voorgaand onderzoek te overkomen en richt zich op de toepassing van informatievaardigheden in ecologisch valide situaties met authentieke taken die de integratie van kennis, vaardigheden, en attitudes vereisen. De interventies in dit onderzoek bevatten tevens transfermetingen of metingen van leereffecten op de langere termijn. De volgende onderzoeksvragen werden onderzocht:

- Wat zijn de effecten van ingebouwde taakondersteuning (aanvultaken, nadrukmanipulatie) op het aanleren van informatievaardigheden?
- Wat zijn de effecten van een voorbeelduitwerking op het aanleren van informatievaardigheden?
- Wat zijn de effecten van geïntegreerd informatievaardighedenonderwijs op het aanleren van informatievaardigheden?
- Wat zijn de algemene karakteristieken van een IPS proces, en hoe voorspellen student-, query-, en broneigenschappen de selectie van relevante en betrouw-bare bronnen?

Het onderzoek beschreven in *Hoofdstuk 2* onderzoekt het principe van het toepassen van nieuwe kennis en vaardigheden tijdens het werken aan een taak met *ingebouwde taakondersteuning*. Twee aanpakken werden vergeleken in een losstaande informatievaardighedentraining waarbij gebruik gemaakt werd van hele taken. De *aanvulstrategie*, waarbij opeenvolgende leertaken telkens een afnemende hoeveelheid uitgewerkte stappen bevatten, werd vergeleken met *nadrukmanipulatie*, waarbij studenten in elke leertaak aanvullende ondersteuning ontvangen op slechts één aspect van de taak. De resultaten tonen geen duidelijke voorkeur voor een van beide aanpakken, waardoor er geen conclusies kunnen worden getrokken over de effectiviteit van deze vormen van taakondersteuning. Geen van de onderzochte condities presteerde beter dan de controleconditie die geen ingebouwde taakondersteuning ontving buiten de eenvoudige begeleiding door de stappen van het IPS proces. Waar metingen van mentale belasting tijdens het leren geen verschillen lieten zien, toonde de significante stijging van scores van voormeting naar nameting duidelijk aan dat de training, bestaande uit een instructievideo, een voorbeelduitwerking, en vier hele taken, effectief was om informatievaardigheden te ontwikkelen in deze doelgroep. Er werd verondersteld dat de voorbeelduitwerking verantwoordelijk was voor een groot deel van het leereffect, en derhalve werd er een vervolgstudie uitgevoerd naar de effecten van deze voorbeelduitwerking op het aanleren van informatievaardigheden.

Dit onderzoek is beschreven in *Hoofdstuk 3* en richt zich op het principe van demonstratie. Gebruikmakend van dezelfde training als beschreven in Hoofdstuk 2, werden studenten die een video van een *voorbeelduitwerking* ontvingen, vergeleken met studenten die een oefentaak uitvoerden. Voor deze twee studies werd een online leeromgeving ontwikkeld die primair bestond uit online zoektaken en ondersteunende videomaterialen. De onderzoeken in Hoofdstukken 2 en 3 werden als losstaande trainingssessies aangeboden in het kader van een universitaire opleiding. Resultaten van het onderzoek tonen aan dat het zien van een video van een voorbeelduitwerking in de vorm van een screencast waarbij de expert hardop nadenkt, en waarbij prompts beantwoord moeten worden, leidde tot een hogere score op de nameting dan wanneer een oefentaak werd uitgevoerd. Dit effect hield stand tot een week na de nameting. Een interessante bevinding is dat de scores van de studenten die de voorbeelduitwerking kregen, ongeveer gelijk zijn aan de scores uit het onderzoek in Hoofdstuk 2. In acht nemende dat alle ingebouwde taakondersteuning in dit latere onderzoek verwijderd was, kan worden aangenomen dat de stijging in scores verklaard wordt door de interventie bestaande uit een instructievideo, een voorbeelduitwerking, en authentieke leertaken. Deze resultaten maken duidelijk dat het aanbieden van een voorbeelduitwerking een effectieve instructiemethode is in deze context, en dat Merrill's (2002) principe van demonstratie geldt voor online informatievaardighedenonderwijs. Het aanbieden van voorbeelduitwerkingen in video's is daarmee aan te raden bij het aanleren van informatievaardigheden.

In tegenstelling tot de losstaande trainingssessies in Hoofdstukken 2 en 3, richt het onderzoek in *Hoofdstuk 4* zich op een hele-taak benadering geïntegreerd in een bestaand curriculum. Het draait om het principe van het toepassen van kennis en vaardigheden in een ecologisch valide setting. In dit onderzoek werd een bestaand curriculum in een lerarenopleiding deels herontworpen en werd een informatievaardighedentraining ingebouwd bestaande uit hele taken in een parallelle online leeromgeving. Het resulterende *blended* onderwijsaanbod inclusief informatievaardighedentraining werd geëvalueerd en vergeleken met het reguliere onderwijsaanbod. De prestaties van studenten werden gemeten door ze aan authentieke zoektaken te laten werken en gelijktijdig al hun handelingen automatisch vast te leggen in logbestanden. Zo konden hun acties, zoals de geproduceerde zoektermen en de geselecteerde bronnen, achteraf gescoord worden. De metingen toonden aan dat studenten die de informatievaardighedentraining ontvingen systematischer te werk gingen dan de studenten die het reguliere onderwijs ontvingen, maar dat hun zoektermen, bronnen, en oplossingen niet van hogere kwaliteit waren. De training zorgde dus voor een ontwikkeling van cognitieve strategieën voor het aanpakken van informatieproblemen, maar verbeterde niet alle benodigde relevante vaardigheden. Daarbij toonde een latere test aan dat de leereffecten vijf weken later verdwenen waren. Er kan worden geconcludeerd dat dit geïntegreerd en blended informatievaardighedenonderwijs, gebruik makend van een hele-taak benadering potentie toont voor het ontwikkelen van informatievaardigheden, maar enkel leereffecten teweegbracht op de korte termijn. Er werd verondersteld dat een onderwijstraject met meer ruimte voor oefening met variërende taken nodig is voor het behalen van de gewenste leereffecten op de langere termijn.

Het onderzoek in *Hoofdstuk 5* paste deze manier van meten opnieuw toe om gedetailleerd inzicht te verkrijgen in de zoekvaardigheden van studenten. Het gaat dieper in op de methode die in Hoofdstuk 4 werd toegepast en gebruikt een diepgaande inspectie van de logbestanden om het zoekproces in kaart te brengen. In een typische zoektaak van 20 minuten gebruikten studenten vijf zoekopdrachten, acht bronnen, en verbleven 65 seconden op een bron. Ongeveer de helft van de bezochte bronnen waren niet betrouwbaar of twijfelachtig, en de andere helft was betrouwbaar of zeer betrouwbaar. Een geselecteerde bron bevatte gemiddeld 18% relevante informatie. In verdere analyse bleek dat de relevantie van de zoektermen zwak doch positief geassocieerd was met de hoeveelheid relevante informatie in een bron, maar ook zwak doch negatief geassocieerd met de betrouwbaarheid van de bron. Bronnen met veel relevante informatie waren doorgaans te vinden bovenaan de SERP, maar dat patroon gold niet voor betrouwbare bronnen. Voorts was er een positieve samenhang tussen de hoeveelheid tijd die een student doorbrengt op een bron, de hoeveelheid relevante informatie in die bron, en de betrouwbaarheid van die bron. Dit onderzoek toont voornamelijk aan dat er nog veel ruimte voor verbetering is bij de zoekvaardigheden van studenten, en dat er in de instructie aandacht moet worden besteed aan het aanleren van goed onderbouwde strategieën voor het formuleren van geschikte zoekopdrachten. Daarnaast wordt gesuggereerd dat er verschillende strategieën nuttig kunnen zijn voor het selecteren van respectievelijk relevante en betrouwbare bronnen.

Het onderzoek in Hoofdstuk 2 geeft helaas geen eenduidig antwoord op de eerste onderzoeksvraag betreffende de verschillende effecten van aanvultaken, nadruk-manipulatie, of een combinatie van deze methoden. Verder onderzoek is nodig voordat conclusies kunnen worden getrokken over de effectiviteit van deze vormen van taakondersteuning, bij voorkeur in langlopende curricula. Betreffende de tweede onderzoeksvraag over de effecten van een voorbeelduitwerking, toont het onderzoek in Hoofdstuk 3 aan dat een 10 minuten durende screencast waarin een expert een informatieprobleem oplost volgens het IPS-I model (Brand-Gruwel, Wopereis, & Walraven, 2009), ontworpen volgens multimedia principes (Mayer, 2014), onderbroken door prompts om voorkennis te activeren en vergelijking van aanpakken te stimuleren, en gevolgd door oefening, effectief is voor het aanleren van informatievaardigheden. Onderwijzers en instructieontwerpers wordt daarom aangeraden gebruik te maken van goed ontworpen voorbeelden in hun instructie. De derde onderzoeksvraag over geïntegreerd informatievaardighedenonderwijs werd onderzocht in Hoofdstuk 4, maar uit dat onderzoek bleken geen sterke leereffecten. Het onderzoek geeft wel een voorbeeld van de toepassing van heersende instructieprincipes voor het ontwerp van taakgebaseerd onderwijs gebruik makend van scaffolding, voorbeelden, cognitieve feedback, en een blended aanbod van instructiematerialen in een ecologisch valide setting. Toekomstig onderzoek zal verder moeten uitwijzen welke best practices gelden bij het integreren van informatievaardighedenonderwijs in bestaande curricula, zodat domeinspecifieke instructie en instructie voor vaardigheden voor zelfgestuurd leren gelijktijdig kunnen worden ontwikkeld. De laatste onderzoeksvraag draaide om het natuurlijke zoekproces van studenten en hun evaluatiegedrag. De bevindingen hebben voornamelijk implicaties voor onderzoek naar het meten en modeleren van het IPS proces, maar tonen ook aan dat zoek- en evaluatiestrategieën zouden kunnen worden afgestemd op het doel van de betreffende taak.

De huidige strategieën voor het efficiënt en effectief oplossen van informatieproblemen zijn waarschijnlijk over een aantal jaren alweer achterhaald, wanneer zoekmachines doorontwikkeld zijn en talloze nieuwe tools en informatie-bronnen zijn ontstaan. We moeten daarom deze onderzoekslijn doorzetten en bewijs verzamelen over welke interventies werken en welke niet, en deze resultaten doortrekken naar andere contexten en omgevingen. Dergelijke instructie zal zowel systematische aanpakken op hoger niveau moeten bevatten, als strategieën op lager niveau, toegespitst op de specifieke taak. We moeten doorgaan met het informeren van onze onderwijzers, en hen uitrusten met de vaardigheden om informatievaardighedenonderwijs te ontwerpen en aan te bieden aan onze studenten nu en in de toekomst. Dit proefschrift neemt een eerste stap en toont aan hoe gevestigde instructieprincipes kunnen worden toegepast om geïntegreerd informatievaardighedenonderwijs te ontwerpen gebruik makend van een hele-taak benadering, voorbeelduitwerkingen, en taakondersteuning, aangeboden in online of blended vorm.

REFERENCES

Agosti, M., Crivellari, F., & Di Nunzio, G. M. (2012). Web log analysis: A review of a decade of studies about information acquisition, inspection and interpretation of user interaction. *Data Mining and Knowledge Discovery, 24*, 663-696. doi:10.1007/s10618-011-0228-8

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*, 147-179. doi:10.1016/S0364-0213(02)00061-7

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*, 87-113. doi:10.1080/00461520.2013.775712

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior, 28*, 515-526. doi:10.1016/j.chb.2011.10.024

Argelagós, E., & Pifarré, M. (2016). Key information-problem solving skills to learn in secondary education: A qualitative, multi-case study. *Journal of Education and Learning, 5*(4), 1-14. doi:10.5539/jel.v5n4p1

Athukorala, K., Głowacka, D., Jacucci, G., Oulasvirta, A., & Vreeken, J. (2016). Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *Journal of the Association for Information Science and Technology, 67*(11), 2635-2651. doi:10.1002/asi.23617

Atkinson, R., Derry, S., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*, 181-214. doi:10.3102/00346543070002181

Atkinson, R., Renkl, A., & Merrill, M. (2003). Transitioning from studying examples to solving problems: Effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology, 95*, 774-783. doi:10.1037/0022-0663.95.4.774

Badke, W. (2010). Why information literacy is invisible. *Communications in Information Literacy, 4*, 129-141. doi:10.1109/ICELMACH.2010.5608206

Bandura, A. (1971). *Social Learning Theory*. New York: General Learning Press.

Bandura, A. (1977). *Social Learning Theory*. Englewood Cliffs: Prentice Hall.

Bannert, M., & Reimann, P. (2011). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*, 193-211. doi:10.1007/s11251-011-9167-4

Baum, B. E., & Gray, J. J. (1992). Expert modeling, self-observation using videotape, and acquisition of basic therapy skills. *Professional Psychology: Research and Practice, 23*, 220-225. doi:10.1037/0735-7028.23.3.220

Becerril, L., & Badia, A. (2015). Information problem-solving skills and the shared knowledge construction process: a comparison of two learning tasks with differing levels of cognitive complexity. *Culture and Education, 27*, 766-801. doi:10.1080/11356405.2015.1092265

Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology, 39*, 775-786. doi:10.1111/j.1467-8535.2007.00793.x

Bjerrum, A. S., Hilberg, O., van Gog, T., Charles, P., & Eika, B. (2013). Effects of modelling examples in complex procedural skills training: A randomised study. *Medical Education, 47*, 888-898. doi:10.1111/medu.12199

Braaksma, M. A. H., Rijlaarsdam, G., van den Bergh, H., & van Hout-Wolters, B. H. A. M. (2004). Observational learning and its effects on the orchestration of writing processes. *Cognition and Instruction, 22*(1), 1-36. doi:10.1207/s1532690Xci2201_1

Braaksma, M. A. H., Rijlaarsdam, G., van den Bergh, H., & van Hout-Wolters, B. H. A. M. (2006). What observational learning entails: A case study. *L1 Educational Studies in Language and Literature, 6*(1), 31-62.

Braaksma, M. A. H., van den Bergh, H., Rijlaarsdam, G., & Couzijn, M. (2001). Effective learning activities in observation taks when learning to write and read argumentative texts. *European Journal of Psychology of Education, 16*, 33-48. doi:10.1007/BF03172993

Brand-Gruwel, S., Kammerer, Y., van Meeuwen, L., & van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning, 33*. doi:10.1111/jcal.12162

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 by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior, 21*, 487-508. doi:10.1016/j.chb.2004.10.005
Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education, 53*, 1207-1217. doi:10.1016/j.compedu.2009.06.004
Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly, 44*(1), 6-28. doi:10.1598/RRQ.44.1.1
Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*, 485-522. doi:10.1207/S1532690XCI2004_2
Bulger, M. E., Mayer, R. E., & Metzger, M. J. (2014). Knowledge and processes that predict proficiency in digital literacy. *Reading and Writing, 27*, 1567-1583. doi:10.1007/s11145-014-9507-2
Chevalier, A., Dommes, A., & Marquié, J.-C. (2015). Strategy and accuracy during information search on the Web: Effects of age and complexity of the search questions. *Computers in Human Behavior, 53*, 305-315. doi:10.1016/j.chb.2015.07.017
Chi, M. T. H., Bassok, M., & Lewis, M. W. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 182*, 145-182.
Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477. doi:10.1207/s15516709cog1803_3
Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152. doi:10.1207/s15516709cog0502_2
Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator, 15*(3), 6-11.
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. doi:10.1016/j.chb.2007.01.021
Derakhshan, M., & Singh, D. (2011). Integration of information literacy into the curriculum: a meta-synthesis. *Library Review, 60*, 218-229. doi:10.1108/00242531111117272
DeStefano, D., & LeFevre, J. A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior, 23*, 1616-1641. doi:10.1016/j.chb.2005.08.012
Dirkx, K. J. H., Kester, L., & Kirschner, P. A. (2014). The testing effect for learning principles and procedures from texts. *The Journal of Educational Research, 107*, 357-364. doi:10.1080/00220671.2013.823370
Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83-87. doi:10.1111/1467-8721.01235
Eisenberg, M. B., & Berkowitz, R. E. (1990). *Information problem-solving: the Big Six Skills approach to library & information skills instruction*. Norwood, New Jersey: Ablex Publishing Corporation.
Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406. doi:10.1037/0033-295X.100.3.363
Fabiani, M., Buckley, J., Gratton, G., Coles, M. G. H. H., Donchin, E., & Logie, R. (1989). The training of complex task performance. *Acta Psychologica, 71*, 259-299. doi:10.1016/0001-6918(89)90012-7
Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., & Tauber, E. R. (2003). How do users evaluate the credibility of Web sites? *Proceedings of the 2003 conference on Designing for user experiences*, 1-15. doi:10.1145/997078.997097
Ford, N., Miller, D., & Moss, N. (2005). Web search strategies and human individual differences: Cognitive and demographic factors, internet attitudes, and approaches. *Journal of the American Society for Information Science and Technology, 56*, 741-756. doi:10.1002/asi.20168
Francom, G. M., & Gardner, J. (2014). What is task-centered learning? *TechTrends, 58*(5), 27-35. doi:10.1007/s11528-014-0784-z
Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2016). Completion strategy or emphasis manipulation? Task support for teaching information problem solving. *Computers in Human Behavior, 62*, 90-104. doi:10.1016/j.chb.2016.03.048

Frerejean, J., Velthorst, G., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2017). Embedded instruction to learn information problem solving: Effects of a whole task approach. *Manuscript submitted for publication.*

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness fading in mathematics and science instruction: A systematic review. *Educational Psychology Review, 26*(1), 9-25. doi:10.1007/s10648-014-9249-3

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. doi:10.1016/j.chb.2007.01.029

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

Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: supporting learners with processing prompts and an interactive comparison tool. *Educational Technology Research and Development, 56*, 73-92. doi:10.1007/s11423-007-9068-z

Gopher, D. (2007). Emphasis change as a training protocol for high-demand tasks. In A. F. Kramer, D. A. Wiegmann, & A. Kirlik (Eds.), *Attention: From Theory to Practice* (pp. 209-224). Oxford: Oxford University Press.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71*, 147-177. doi:10.1016/0001-6918(89)90007-3

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013). Learning to be creative. The effects of observational learning on students' design products and processes. *Learning and Instruction, 28*, 35-47. doi:10.1016/j.learninstruc.2013.05.001

Hart, S. G., & Staveland, L. E. (1998). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). Amsterdam: North Holland.

Hilbert, T. S., Renkl, A., Schworm, S., Kessler, S., & Reiss, K. (2008). Learning to teach with worked-out examples: a computer-based learning environment for teachers. *Journal of Computer Assisted Learning, 24*, 316-332. doi:10.1111/j.1365-2729.2007.00266.x

Hill, J., & Hannafin, M. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*, 37-52. doi:10.1007/bf02504914

Hoffman, B., & Schraw, G. (2010). Conceptions of efficiency: Applications in learning and problem solving. *Educational Psychologist, 45*, 1-14. doi:10.1080/00461520903213618

Holmes Finch, W., Bolin, J. E., & Kelley, K. (2014). *Multilevel Modeling Using R*. New York: CRC Press.

Hölscher, C., & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks, 33*, 337-346. doi:10.1016/S1389-1286(00)00031-1

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014a). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80-91. doi:10.1016/j.chb.2014.09.013

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2014b). Effects of creating video-based modeling examples on learning and transfer. *Learning and Instruction, 33*, 108-119. doi:10.1016/j.learninstruc.2014.04.005

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2015). Learning from video modeling examples: does gender matter? *Instructional Science, 44*, 69-86. doi:10.1007/s11251-015-9360-y

Hsu, C.-Y., Tsai, M.-J., Hou, H.-T., & Tsai, C.-C. (2014). Epistemic beliefs, online search strategies, and behavioral patterns while exploring socioscientific issues. *Journal of Science Education and Technology, 23*, 471-480. doi:10.1007/s10956-013-9477-1

Johnson, A. M., Reisslein, J., & Reisslein, M. (2014). Representation sequencing in computer-based engineering education. *Computers and Education, 72*, 249-261. doi:10.1016/j.compedu.2013.11.010

Kalyuga, S. (2008). When less is more in cognitive diagnosis: A rapid online method for diagnosing learner task-specific expertise. *Journal of Educational Psychology, 100*, 603-612. doi:10.1037/0022-0663.100.3.603

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*, 23-31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., & Hanham, J. (2011). Instructing in generalized knowledge structures to develop flexible problem solving skills. *Computers in Human Behavior, 27*, 63-68. doi:10.1016/j.chb.2010.05.024

Kalyuga, S., Rikers, R., & Paas, F. (2012). Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills. *Educational Psychology Review, 24*, 313-337. doi:10.1007/s10648-012-9195-x

Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology, 96*, 558-568. doi:10.1037/0022-0663.96.3.558

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2012). The role of Internet-specific epistemic beliefs in laypersons' source evaluations and decisions during Web search on a medical issue. *Computers in Human Behavior, 29*, 1193-1203. doi:10.1016/j.chb.2012.10.012

Kammerer, Y., & Gerjets, P. (2012). How search engine users evaluate and select web search results: The impact of the search engine interface on credibility assessments. In D. Lewandowsky (Ed.), *Web Search Engine Research (Library and Information Science)* (Vol. 4, pp. 251-279). Bingley, England: Emerald Group Publishing Ltd.

Keil, F. C., & Kominsky, J. F. (2013). Missing links in middle school: Developing use of disciplinary relatedness in evaluating internet search results. *PloS one, 8*(6), 1-5. doi:10.1371/journal.pone.0067777

Kennedy, G. E., Judd, T. S., Churchward, A., Gray, K., & Krause, K. L. (2008). First year students' experiences with technology: Are they really digital natives? *Australasian Journal of Educational Technology, 24*, 108-122. doi:10.1.1.85.9526

Kim, Y. M., & Rieh, S. Y. (2005). Dual-task performance as a measure of mental effort in searching a library system and the Web. *Proceedings of the American Society for Information Science and Technology, 42*(1), 1-14. doi:10.1002/meet.14504201155

Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135-142. doi:10.1016/j.tate.2017.06.001

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86. doi:10.1207/s15326985ep4102_1

Kirschner, P. A., & van Merriënboer, J. J. G. (2009). Ten steps to complex learning. In T. L. Good (Ed.), *21st century education: a reference handbook* (pp. 244-253). Thousand Oaks, CA: Sage.

Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do learners really know best? Urban legends in education. *Educational Psychologist, 48*, 169-183. doi:10.1080/00461520.2013.804395

Kong, S. C. (2014). Developing information literacy and critical thinking skills through domain knowledge learning in digital classrooms: An experience of practicing flipped classroom strategy. *Computers & Education, 78*, 160-173. doi:10.1016/j.compedu.2014.05.009

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121-1134. doi:10.1037/0022-3514.77.6.1121

Kuhlthau, C. C. (1988). Developing a model of the library search process: Cognitive and affective aspects. *Reference Quarterly, 28*, 232-242.

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. doi:10.1016/j.chb.2007.01.022

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*, 45-66. doi:10.1007/BF02313485

Lazonder, A. W. (2000). Exploring novice users' training needs in searching information on the WWW. *Journal of Computer Assisted Learning, 16*, 326-335. doi:10.1046/j.1365-2729.2000.00145.x

Lazonder, A. W., Biemans, H. J. A., & Wopereis, I. G. J. H. (2000). Differences between novice and experienced users in searching information on the World Wide Web. *Journal of the American Society for Information Science and Technology, 51*, 576-581. doi:10.1002/(sici)1097-4571(2000)51:6<576::aid-asi9>3.0.co;2-7

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106-131. doi:10.1177/1529100612451018

Lim, J., Reiser, R. A., & Olina, Z. (2009). The effects of part-task and whole-task instructional approaches on acquisition and transfer of a complex cognitive skill. *Educational Technology Research and Development, 57*, 61-77. doi:10.1007/s11423-007-9085-y
Lin, S., & Xie, I. (2013). Behavioral changes in transmuting multisession successive searches over the web. *Journal of the American Society for Information Science and Technology, 64*, 1259-1283. doi:10.1002/asi.22839
Linek, S., Gerjets, P., & Scheiter, K. (2010). The speaker/gender effect: does the speaker's gender matter when presenting auditory text in multimedia messages? *Instructional Science, 38*, 503-521. doi:10.1007/s11251-009-9115-8
List, A., Grossnickle, E. M., & Alexander, P. A. (2016). Undergraduate students' justifications for source selection in a digital academic context. *Journal of Educational Computing Research, 54*, 22-61. doi:10.1177/0735633115606659
Loo, J. L., Eifler, D., Smith, E., Pendse, L., He, J., Sholinbeck, M., . . . Dupuis, E. A. (2016). Flipped instruction for information literacy: Five instructional cases of academic librarians. *The Journal of Academic Librarianship, 42*, 273-280. doi:10.1016/j.acalib.2016.03.001
Lucassen, T., Muilwijk, R., Noordzij, M. L., & Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology, 64*(2), 254-264. doi:10.1002/asi.22743
Lucassen, T., & Schraagen, J. M. (2013). The influence of source cues and topic familiarity on credibility evaluation. *Computers in Human Behavior, 29*, 1387-1392. doi:10.1016/j.chb.2013.01.036
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. doi:10.1023/a:1014617530297
Margaryan, A., Littlejohn, A., & Vojt, G. (2011). Are digital natives a myth or reality? University students' use of digital technologies. *Computers & Education, 56*(2), 429-440. doi:10.1016/j.compedu.2010.09.004
Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers and Education, 76*, 143-157. doi:10.1016/j.compedu.2014.03.016
Mayer, R. E. (2014). *The Cambridge Handbook of Multimedia Learning* (2nd ed.). Cambridge: Cambridge University Press.
Melo, M., & Miranda, G. (2015). Learning electrical circuits: The effects of the 4C-ID instructional approach in the acquisition and transfer of knowledge. *Journal of Information Technology Education: Research, 14*, 313-337.
Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *Libraries and the Academy, 4*, 331-344. doi:10.1353/pla.2004.0055
Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development, 50*(3), 43-59. doi:10.1007/BF02505024
Metzger, M. J. (2007). Making sense of credibility on the web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology, 58*, 2078-2091. doi:10.1002/asi.20672
Miller, C., & Bartlett, J. (2012). 'Digital fluency': towards young people's critical use of the internet. *Journal of Information Literacy, 6*(2), 35-55.
Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management, 51*, 557-569. doi:10.1016/j.ipm.2015.05.004
Moore, P. (1995). Information problem solving: A wider view of library skills. *Contemporary Educational Psychology, 20*, 1-31. doi:10.1006/ceps.1995.1001
Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*, 118-125. doi:10.1016/j.cedpsych.2012.12.004
Paas, F., Tuovinen, J. E., Tabbers, H., & van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-71. doi:10.1207/s15326985ep3801_8
Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429-434. doi:10.1037/0022-0663.84.4.429

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 351-371. doi:10.1007/BF02213420
Pariser, E. (2011). *The filter bubble: What the internet is hiding from you*. New York: Penguin Press.
Pedersen, S., & Lui, M. (2003). The transfer of problem-solving skills from a problem-based learning environment: the effect of modeling an expert's cognitive processes. *Journal of Research on Technology in Education, 35*, 303-320.
Perin, D. (2011). Facilitating student learning through contextualization: A review of evidence. *Community College Review, 39*, 268-295. doi:10.1177/0091552111416227
Platform Onderwijs2032. (2016). *Eindadvies Platform Onderwijs 2032*. Den Haag: Platform Onderwijs 2032.
Prensky, M. (2001). Digital natives, digital immigrants, part II: Do they really think differently? *On the Horizon, 9*(6), 1-9. doi:10.1108/10748120110424843
Probert, E. (2009). Information literacy skills: Teacher understandings and practice. *Computers & Education, 53*, 24-33. doi:10.1016/j.compedu.2008.12.018
Reigeluth, C. M. (1999). What is instructional design theory and how is it changing? In C. M. Reigeluth (Ed.), *Instructional-design theories and models: A new paradigm of instructional theory*. (Vol. 2, pp. 5-29). Mahwah, NJ: Lawrence Erlbaum Associates.
Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*, 1-29. doi:10.1207/s15516709cog2101_1
Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education, 14*, 477-488. doi:10.1007/BF03172974
Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction, 12*, 529-556. doi:10.1016/s0959-4752(01)00030-5
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1-37. doi:10.1111/cogs.12086
Renkl, A., & Atkinson, R. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: a cognitive load perspective. *Educational Psychologist, 38*, 15-22. doi:10.1207/s15326985ep3801_3
Renkl, A., Atkinson, R., & Große, C. (2004). How fading worked solution steps works – A cognitive load perspective. *Instructional Science, 32*, 59-82. doi:10.1023/B:TRUC.0000021815.74806.f6
Renkl, A., & Atkinson, R. K. (2002). Learning from examples: Fostering self-explanations in computer-based learning environments. *Interactive Learning Environments, 10*, 105-119. doi:10.1076/ilee.10.2.105.7441
Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review, 21*, 67-78. doi:10.1007/s10648-008-9093-4
Rieh, S. Y., Collins-Thompson, K., Hansen, P., & Lee, H. J. (2016). Towards searching as a learning process: A review of current perspectives and future directions. *Journal of Information Science, 42*(1), 19-34. doi:10.1177/0165551515615841
Rieh, S. Y., Kim, Y.-M. M., & Markey, K. (2012). Amount of invested mental effort (AIME) in online searching. *Information Processing and Management, 48*, 1136-1150. doi:10.1016/j.ipm.2012.05.001
Rosman, T., Mayer, A.-K., & Krampen, G. (2014). Combining self-assessments and achievement tests in information literacy assessment: empirical results and recommendations for practice. *Assessment & Evaluation in Higher Education, 40*(5), 1-15. doi:10.1080/02602938.2014.950554
Rosman, T., Mayer, A.-K., & Krampen, G. (2015). Measuring psychology students' information-seeking skills in a situational judgment test format. *European Journal of Psychological Assessment, 31*, 1-10. doi:10.1027/1015-5759/a000239
Rosman, T., Mayer, A.-K., & Krampen, G. (2016a). A longitudinal study on information-seeking knowledge in psychology undergraduates: Exploring the role of information literacy instruction and working memory capacity. *Computers & Education, 96*, 94-108. doi:10.1016/j.compedu.2016.02.011
Rosman, T., Mayer, A.-K., & Krampen, G. (2016b). On the pitfalls of bibliographic database searching: comparing successful and less successful users. *Behaviour & Information Technology, 35*, 106-117. doi:10.1080/0144929X.2015.1066446
Rouet, J.-F. (2009). Managing cognitive load during document-based learning. *Learning and Instruction, 19*, 445-450. doi:10.1016/j.learninstruc.2009.02.007

Rouet, J.-F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction, 21*, 205-219. doi:10.1016/j.learninstruc.2010.02.007

Russell, D., & Grimes, C. (2007). *Assigned tasks are not the same as self-chosen Web search tasks*. Paper presented at the 40th Annual Hawaii International Conference on System Sciences (HICSS '07), Washington, DC.

Saito, H., & Miwa, K. (2007). Construction of a learning environment supporting learners' reflection: A case of information seeking on the Web. *Computers & Education, 49*, 214-229. doi:10.1016/j.compedu.2005.07.001

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

Sarfo, F. K., & Elen, J. (2007). Developing technical expertise in secondary technical schools: The effect of 4C/ID learning environments. *Learning Environments Research, 10*, 207-221. doi:10.1007/s10984-007-9031-2

Sarsfield, E. (2014). Differences between novices' and experts' solving ill-structured problems. *Public Health Nursing, 31*, 444-453. doi:10.1111/phn.12100

Scheiter, K., Gerjets, P., Vollmann, B., & Catrambone, R. (2009). The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction, 19*, 387-401. doi:10.1016/j.learninstruc.2009.02.004

Smith, C. L. (2015). Domain-independent search expertise: A description of procedural knowledge gained during guided instruction. *Journal of the Association for Information Science and Technology, 66*(7), 1388-1405. doi:10.1002/asi.23272

Smith, C. L. (2017). Domain-independent search expertise: Gaining knowledge in query formulation through guided practice. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23776

Smith, E. E. (2012). The digital native debate in higher education: A comparative analysis of recent literature. *Canadian Journal of Learning and Technology, 38*(3), 1-18.

Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27*, 46-52. doi:10.1016/j.chb.2010.05.011

Squibb, S. D., & Mikkelsen, S. (2016). Assessing the value of course-embedded information literacy on student learning and achievement. *College & Research Libraries, 77*, 164-183. doi:10.5860/crl.77.2.164

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. doi:10.1016/j.chb.2007.01.023

Stark, R., & Krause, U. M. (2009). Effects of reflection prompts on learning outcomes and learning behaviour in statistics education. *Learning Environments Research, 12*, 209-223. doi:10.1007/s10984-009-9063-x

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction, 12*, 39-60. doi:10.1016/s0959-4752(01)00015-9

Susilo, A. P., van Merriënboer, J., van Dalen, J., Claramita, M., & Scherpbier, A. (2013). From lecture to learning tasks: use of the 4C/ID model in a communication skills course in a continuing professional education context. *Journal of continuing education in nursing, 44*, 278-284. doi:10.3928/00220124-20130501-78

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*, 257-285. doi:10.1207/s15516709cog1202_4

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction, 16*, 165-169. doi:10.1016/j.learninstruc.2006.02.005

Tabachnick, B. G., & Fidell, L. S. (2007). Principal components and factor analysis. In *Using Multivariate Statistics* (5th ed., pp. 607-675). Boston: Pearson Education.

Tapscott, D. (1999). Educating the net generation. *Educational Leadership, 56*(5), 6-11.

Thijs, A., Fisser, P., & van der Hoeven, M. (2014). *21e eeuwse vaardigheden in het curriculum van het funderend onderwijs*. Enschede: SLO.

Timmers, C., & Veldkamp, B. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education, 56*, 923-930. doi:10.1016/j.compedu.2010.11.007

Timmers, C. F., Walraven, A., & Veldkamp, B. P. (2015). The effect of regulation feedback in a computer-based formative assessment on information problem solving. *Computers & Education, 87*, 1-9. doi:10.1016/j.compedu.2015.03.012

Van den Boom, G., Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior, 20*, 551-567. doi:10.1016/j.chb.2003.10.001

Van Deursen, A. J. A. M., & van Diepen, S. (2013). Information and strategic Internet skills of secondary students: A performance test. *Computers & Education, 63*, 218-226. doi:10.1016/j.compedu.2012.12.007

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

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-oriented worked examples: Improving transfer performance through enhanced understanding. *Instructional Science, 32*, 83-98. doi:10.1023/B:TRUC.0000021810.70784.b0

Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2008). Effects of studying sequences of process-oriented and product-oriented worked examples on troubleshooting transfer efficiency. *Learning and Instruction, 18*, 211-222. doi:10.1016/j.learninstruc.2007.03.003

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. doi:10.1037/1076-898x.11.4.237

Van Gog, T., & Rummel, N. (2010). Example-based learning: Integrating cognitive and social-cognitive research perspectives. *Educational Psychology Review, 22*, 155-174. doi:10.1007/s10648-010-9134-7

Van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323-327. doi:10.1016/j.compedu.2013.12.004

Van Laar, E., van Deursen, A. J. A. M., van Dijk, J. A. G. M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577-588. doi:10.1016/j.chb.2017.03.010

Van Meeuwen, L. W. (2008). *Tracking the Brain*. Technische Universiteit Eindhoven, Eindhoven. Retrieved from http://repository.tue.nl/634895

Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of Educational Computing Research, 6*, 265-285. doi:10.2190/4NK5-17L7-TWQV-1EHL

Van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education, 64*, 153-160. doi:10.1016/j.compedu.2012.11.025

Van Merriënboer, J. J. G., & Ayres, P. (2005). Research on cognitive load theory and its design implications for e-learning. *Educational Technology Research and Development, 53*, 5-13. doi:10.1007/BF02504793

Van Merriënboer, J. J. G., Clark, R. E., & de Croock, M. B. M. (2002). Blueprints for complex learning: The 4C/ID-model. *Educational Technology Research and Development, 50*(2), 39-61. doi:10.1007/BF02504993

Van Merriënboer, J. J. G., & de Croock, M. B. M. (1995). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research, 8*, 365-394. doi:10.2190/MJDX-9PP4-KFMT-09PM

Van Merriënboer, J. J. G., Kester, L., & Paas, F. (2006). Teaching complex rather than simple tasks: Balancing intrinsic and germane load to enhance transfer of learning. *Applied Cognitive Psychology, 20*, 343-352. doi:10.1002/acp.1250

Van Merriënboer, J. J. G., & Kirschner, P. A. (2013). *Ten steps to complex learning: A systematic approach to four-component instructional design* (2nd ed.). New York: Routledge.

Van Merriënboer, J. J. G., & Kirschner, P. A. (2018). *Ten steps to complex learning: A systematic approach to four-component instructional design* (3rd ed.). New York: Routledge.

Van Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner's mind: Instructional design for complex learning. *Educational Psychologist, 38*, 5-13. doi:10.1207/S15326985EP3801_2

Van Strien, J. L. H., Brand-Gruwel, S., & Boshuizen, H. P. A. (2014). Dealing with conflicting information from multiple nonlinear texts: Effects of prior attitudes. *Computers in Human Behavior, 32*, 101-111. doi:10.1016/j.chb.2013.11.021

Vandewaetere, M., Manhaeve, D., Aertgeerts, B., Clarebout, G., van Merriënboer, J. J. G., & Roex, A. (2015). 4C/ID in medical education: How to design an educational program based on whole-task learning: AMEE Guide No. 93. *Medical Teacher, 37*, 4-20. doi:10.3109/0142159X.2014.928407

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. doi:10.1016/j.chb.2007.01.030

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education, 52*, 234-246. doi:10.1016/j.compedu.2008.08.003

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2010). Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*, 716-728. doi:10.1016/j.chb.2010.01.008

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2013). Fostering students' evaluation behaviour while searching the internet. *Instructional Science, 41*, 125-146. doi:10.1007/s11251-012-9221-x

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*, 249-270.

Wirth, W., Sommer, K., von Pape, T., & Karnowski, V. (2015). Success in online searches: Differences between evaluation and finding tasks. *Journal of the Association for Information Science and Technology*. doi:10.1002/asi.23389

Wopereis, I., Brand-Gruwel, S., & Vermetten, Y. (2008). The effect of embedded instruction on solving information problems. *Computers in Human Behavior, 24*, 738-752. doi:10.1016/j.chb.2007.01.024

Wopereis, I., Frerejean, J., & Brand-Gruwel, S. (2015). Information problem solving instruction in higher education: A case study on instructional design. In S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi, & L. Roy (Eds.), *Communications in Computer and Information Science: Vol. 552. Information Literacy: Moving Toward Sustainability* (Vol. 552, pp. 293-302). Basel, Switzerland: Springer.

Wopereis, I., Frerejean, J., & Brand-Gruwel, S. (2016). Teacher perspectives on whole-task information literacy instruction. In S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi, L. Roy, & T. Çakmak (Eds.), *Information Literacy: Key to an Inclusive Society. ECIL 2016. Communications in Computer and Information Science* (Vol. 676, pp. 678-687). Cham: Springer International Publishing.

Wu, W.-C., & Kelly, D. (2014). Online search stopping behaviors: An investigation of query abandonment and task stopping. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1-10. doi:10.1002/meet.2014.14505101030

Xie, I. (2011). Information searching and search models. In M. J. Bates & M. Niles Maack (Eds.), *Encyclopedia of Library and Information Sciences* (3rd ed., pp. 37-41). New York: Taylor & Francis.

Yechiam, E., Erev, I., & Gopher, D. (2001). On the potential value and limitations of emphasis change and other exploration-enhancing training methods. *Journal of Experimental Psychology: Applied, 7*, 277-285. doi:10.1037/1076-898x.7.4.277

Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the transition from touch-typing training to everyday use. *Human Factors, 45*, 671-684. doi:10.1518/hfes.45.4.671.27085

Zhou, M. (2013). A systematic understanding of successful web searches in information-based tasks. *Educational Technology & Society, 16*, 321-331.

Nawoord

Dit zijn ze dan, de laatste woorden van dit proefschrift. Met blijdschap en opluchting kom ik aan het einde van een lange periode waarin er in mijn leven meer is veranderd dan er gelijk is gebleven. Dit proefschrift schrijven was verreweg de moeilijkste taak die ik heb volbracht, en zeker de laatste maanden heb ik mezelf tot het uiterste moeten drijven om letterlijk de laatste punt te kunnen zetten. Maar het is nu dan zo ver!

Ik begon als net afgestudeerd arbeids– en organisatiepsycholoog aan een promotietraject dat draaide om het ontwerpen en onderzoeken van onderwijs. Niet direct een onderwerp dat aansloot op mijn studie, dus ik kwam binnen zonder achtergrond in onderwijskunde of onderwijswetenschappen, en ik had nog maar heel eventjes aan onderzoek geproefd. Familie en vrienden uit mijn directe omgeving hadden nauwelijks affiniteit met de academische wereld, en maar weinigen wisten wat promoveren inhield. Nu terugkijkend realiseer ik me dat ik dat toen evenmin wist, en als naïeve beginner met een minimum aan voorkennis begon aan een traject waarop zich meer uitdagingen en valkuilen zouden openbaren dan ik ooit had verwacht. Het was niet altijd makkelijk, maar ik ben ontzettend blij en trots dat het proefschrift na al dat harde werken eindelijk af is.

De perfectionist in mij kon soms maar lastig omgaan met alle druk en stress die een promotietraject met zich meebrengt. Ik twijfelde vaak over de gemaakte keuzes, als de onderzoeksresultaten anders uitvielen dan gehoopt, of er weer een periode zonder noemenswaardige progressie voorbijging. Mijn motivatie schommelde, en ik wisselde productieve periodes af met periodes waarin ik uitvluchten zocht in allerlei afleidingen. Ik liep tegen mijn grenzen aan en ging er soms ongewild en onbewust overheen. Zo kwam het dat ik de laatste zomer moest zwoegen om de laatste hoofdstukken op het scherm te krijgen. Ik schreef verbeten door, wetende dat ik niet over alles tevreden zou zijn, maar dat was oké. Want dat is een van de vele lessen die ik heb geleerd. Soms kost perfectie simpelweg te veel. Ondanks die onvolkomenheden en de scherpe randjes ben ik trots op het werk dat er ligt. Want is het niet het doel van een promotietraject dat je *leert* onderzoek te doen? Leren gaat met vallen en opstaan. Ik ben blij dat ik beide heb meegemaakt, en dankbaar voor de enorme ontwikkeling die ik heb doorgemaakt.

Ik leerde veel, over cognitieve belasting, multimedia principes, het 4C/ID model, en informatievaardigheden, en begon langzaam in te zien hoe ik een goede blauwdruk voor onderwijs kon ontwerpen. Het daadwerkelijk implementeren en aanbieden van online onderwijs vormde de volgende uitdaging. Ik had echter wat ervaring opgedaan met programmeren en koos ervoor om op eigen houtje een leeromgeving te bouwen waarin studenten netjes in condities werden ingedeeld en vervolgens de voor hen bestemde instructievideo's, taken, en vragenlijsten ontvingen. Dat was ontzettend veel werk, en ik leerde hoeveel keuzes je als ontwerper moet maken. Keuzes op grote schaal, over welke vaardigheden getraind moeten worden, maar ook keuzes op kleine schaal, over het kiezen van een animatie of statische tekst in een instructievideo. Later programmeerde ik een analysetool, die de data van elke participant op een leesbare en gestructureerde wijze weergaf. Dat hielp enorm bij de interpretatie en data-analyse, en ik kijk met veel voldoening daarop terug, want de ervaring die ik hiermee heb opgedaan zal mij in de toekomst ongetwijfeld van pas komen. Het ontwerpen en ontwikkelen van goed doordacht onderwijs kost enorm veel tijd, en ik heb geleerd dat mensen zich dat niet altijd realiseren. Ik vond mijn passie in het ontwerpen, ontwikkelen, en aanbieden van effectief en efficiënt onderwijs, en ben vastberaden hier in de toekomst mee door te gaan.

Omdat ik gedurende mijn promotietraject van meer mensen geleerd heb dan ik hier kan noemen, wil ik de laatste zinnen in dit proefschrift richten aan Paul, Saskia, Johan, Wendy, Olga, Iwan, en Gerdo.

Paul, jouw rol in dit project is gaandeweg verschoven. Was je aanvankelijk nog direct betrokken bij het onderzoek, later vroeg een veranderende organisatie binnen de OU om het verschuiven van prioriteiten. De afstand werd groter en we kozen om jou enigszins te ontlasten en je meer in de schrijffase te betrekken. Ik wil je bedanken voor je scherpe feedback, die ondanks dat 'ie soms verpakt zat in prikkelende bewoordingen, toch sterk heeft bijgedragen aan de kwaliteit van dit werk.

Saskia, jouw rol verschoof gedurende dit traject van dagelijks begeleider naar co-promotor naar promotor. Ook jij kwam in de draaikolk van de veranderende OU, en dat leidde ertoe dat de hulp van Johan werd ingeroepen om dit team te ondersteunen. Ik heb onze samenwerking altijd als zeer prettig ervaren. Ik kon altijd even binnenlopen met een vraag en rekenen op jouw vrolijke humeur. Had ik vragen of zorgen, dan werd er altijd snel gerelativeerd en vertrouwen opgebouwd. Misschien leidde tijdgebrek of afleiding soms wel tot iets te makkelijk relativeren en had je me iets strakker aan het lijntje kunnen houden, maar ach, dat is voortschrijdend inzicht, zullen we maar zeggen.

Johan, ik denk niet dat een begeleider zijn promovendus ooit Dr. Oetlul genoemd heeft en op dezelfde avond grappen maakt over zijn dialect. Dat geeft wel mooi aan hoe onze omgang was: los en gemakkelijk, maar daarnaast hebben we wel zeer nuttige discussies gehad over de inhoud van mijn proefschrift. Toen Paul en Saskia wat meer afstand namen heb jij de taak op je genomen om mijn eerste schrijfproducten diepgaand te verwerken en treffende feedback te formuleren. Ik heb daar veel van geleerd en daarvoor ook mijn grote dank.

Wendy, jij was aan het begin kortstondig betrokken bij mijn traject, maar ook jouw rol veranderde en je verdween een tijd uit beeld. Toen je terugkwam werden we samen verantwoordelijk voor de cursus Ontwerpen van Onderwijs, die we in recordtijd moesten reviseren. Maar ondanks die druk konden we snel op elkaar vertrouwen en zonder enige problemen samenwerken. De hilariteit tijdens het opnemen van de video's was voor mij oprecht een van de hoogtepunten van mijn tijd bij de OU. Ik ben blij dat we elkaar nu ook op de UM weer zien, en hoop dat we onze lunchafspraken in stand houden.

Olga, ik heb je leren kennen als een enthousiaste, betrokken, en nieuwsgierige vrouw, en vond onze gesprekken tijdens de autoritjes van en naar Eijsden altijd erg plezierig. We hebben weinig daadwerkelijk samen gewerkt, maar we konden wel altijd fijn praten over onderwijs, de OU, en over Eijsden. Bedankt dat ik zo vaak mocht meerijden. Ik hoop dat we elkaar nog geregeld tegenkomen.

Iwan, die ietwat eigenzinnige maar consciëntieuze man waarmee ik samen de cursus Informatievaardigheden voor Sociale Wetenschappers draaide. De diepte en nauwkeurigheid waarmee jij iemands werk las en van feedback voorzag was voor mij een eyeopener. Van jou leerde ik hoeveel denkwerk er kwam kijken bij het ontwerpen van onderwijs. We voerden vele nuttige en leerzame gesprekken over informatievaardigheden, onderwijs, maar soms zaten we gewoon ordinair te roddelen over het reilen en zeilen op de OU. Ik hoop dat we elkaar blijven zien, in het kader van 4C/ID bijeenkomsten of daarbuiten, en ik hoop vooral ooit nog een keer bij jouw promotie aanwezig te mogen zijn.

Tot slot, Gerdo. Wat kan ik tegen jou zeggen dat recht doet aan de dank die ik jou verschuldigd ben? Ik leerde je kennen toen ik mijn derde studie in Doetinchem kwam uitvoeren, en we hebben sindsdien veelvuldig contact gehouden. Je wilde graag op de hoogte blijven en meedenken met het onderzoek, en zo gebeurde het dat je eigenlijk de helft van mijn promotietraject met mij hebt meegelopen. Uren brachten we door op Skype, en we verzetten enorme bergen denkwerk om onze resultaten te verwerken en te begrijpen. Wat hebben we samen ontzettend veel geleerd. Ik voel me dan ook enigszins schuldig dat ik degene ben die mag promoveren, terwijl jij eigenlijk ook wel recht hebt op een stukje daarvan. Het lijkt me dan ook niet meer dan logisch dat jij naast me mag staan op het podium. Als je ooit besluit om te gaan promoveren, weet dan dat ik je daarin zal bijstaan zoals jij ook mij hebt bijgestaan.

ICO Dissertation Series

In the ICO Dissertation Series dissertations are published of graduate students from faculties and institutes on educational research within the ICO Partner Universities: Eindhoven University of Technology, Leiden University, Maastricht University, Open University of the Netherlands, University of Amsterdam, University of Twente, Utrecht University, VU University Amsterdam, and Wageningen University, and formerly University of Groningen (until 2006), Radboud University Nijmegen (until 2004), and Tilburg University (until 2002). The University of Groningen, University of Antwerp, University of Ghent, and the Erasmus University Rotterdam have been 'ICO 'Network partner' in 2010 and 2011. From 2012 onwards, these ICO Network partners are full ICO partners, and from that period their dissertations will be added to this dissertation series.

303. Strien, J.L.H. van (19-12-2014). *Who to Trust and What to Believe? Effects of Prior Attitudes and Epistemic Beliefs on Processing and Justification of Conflicting Information From Multiple Sources*. Heerlen: Open University of the Netherlands.
304. Wijaya, A. (21-01-2015). *Context-based mathematics tasks in Indonesia: Towards better practice and achievement*. Utrecht: Utrecht University.
305. Goossens, N.A.M.C. (22-01-2015). *Distributed Practice and Retrieval Practice in Primary School Vocabulary Learning*. Rotterdam: Erasmus University.
306. Jupri, A. (28-01-2015). *The use of applets to improve Indonesian student performance in algebra*. Utrecht: Utrecht University.
307. Griethuijsen, R.A.L.F. van (11-03-2015). *Relationships between students' interest in science, views of science and science teaching in upper primary and lower secondary education*. Eindhoven: Eindhoven University of Technology.
308. De Smet, C. (11-05-2015). *Using a learning management system in secondary education: Design and implementation characteristics of learning paths*. Ghent: Ghent University.
309. Aesaert, K. (19-05-2015). *Identification and assessment of digital competences in primary education*. Ghent: Ghent University.

310. Ardies, J. (22-05-2015). *Students’ attitudes towards technology. A cross-sectional and longitudinal study in secondary education*. Antwerp: University of Antwerp

311. Donker, A.S. (11-06-2015). *Towards effective learning strategies*. Groningen: University of Groningen

312. Veldhuis, M. (24-06-2015). *Improving classroom assessment in primary mathematics education*. Utrecht: Utrecht University.

313. Leeuwen, A. van (30-06-2015). *Teacher Regulation of CSCL: Exploring the complexity of teacher regulation and the supporting role of learning analytics*. Utrecht: Utrecht University.

314. Boschman, F.B. (28-08-2015). *Collaborative design of ICT-rich early literacy learning material: Design talk in teacher teams*. Enschede: Twente University.

315. Dijk, M.L. van (04-09-2015). *Physical Activity, Cognitive Performance and Academic Achievement in Adolescents*. Heerlen: Open University of the Netherlands.

316. Jaarsma, T. (04-09-2015). *Expertise Development Under the Microscope: Visual Problem Solving in Clinical Pathology*. Heerlen: Open University of the Netherlands.

317. Isac, M.M. (01-10-2015). *Effective civic and citizenship education. A cross-cultural perspective*. Groningen: University of Groningen.

318. Spelt, E.J.H. (26-10-2015). *Teaching and learning of interdisciplinary thinking in higher education in engineering*. Wageningen: Wageningen University.

319. Ritzema, E.S. (05-11-2015). *Professional development in data use: The effects of primary school teacher training on teaching practices and students' mathematical proficiency*. Groningen: University of Groningen.

320. Gijselaers H.J.M. (06-11-2015). *Biological Lifestyle Factors in Adult Distance Education: Predicting Cognitive and Learning Performance*. Heerlen: Open University of the Netherlands.

321. Oude Groote Beverborg, A. (12-11-2015). *Fostering sustained teacher learning: Co-creating purposeful and empowering workplaces*. Enschede: Twente University.

322. Dijkstra, E.M. (13-11-2015). *Teaching High-Ability Pupils in Early Primary School*. Heerlen: Open University of the Netherlands.

323. Want, A.C. van der (17-11-2015). *Teachers' Interpersonal Role Identity*. Eindhoven: Eindhoven University of Technology.

324. Wal, M.M van der (4-12-2015). *The Role of Computer Models in Social Learning for Participatory Natural Resource Management*. Heerlen: Open University of the Netherlands.

325. Herten, M. van (11-12-2015). *Learning communities, informal learning and the humanities. An empirical study of book discussion groups*. Heerlen: Open University of the Netherlands.

326. Lucieer, S.M. (06-01-2016). *Selecting Students for Medical Education: Exploring Novel Approaches*. Rotterdam: Erasmus University Rotterdam.

327. Ouwehand, K.H.R. (15-01-2016). *Effects of Observing and Producing Deictic Gestures on Memory and Learning in Different Age Groups*. Rotterdam: Erasmus University Rotterdam.

328. Bogert, N.J. van der (21-01-2016). *On teachers' visual perception and interpretation of classroom events using eye tracking and collaborative tagging methodologies*. Eindhoven: Eindhoven University of Technology.

329. Wolff, C. (16-02-2016). *Revisiting 'withitness': Differences in teachers' representations, perceptions, and interpretations of classroom management*. Heerlen: Open University of the Netherlands.

330. Kok, E.M. (01-04-2016). *Developing visual expertise; from shades of grey to diagnostic reasoning in radiology*. Maastricht: Maastricht University.

331. De Beer, H.T. (11-05-2016). *Exploring Instantaneous Speed in Grade Five: A Design Research*. Eindhoven: Eindhoven University of Technology.

332. Ebbeler, J. (12-05-2016). *Implementing data use in schools: effects on the professional development of educators and the role of school leaders in data teams*. Enschede: University of Twente.

333. Draaijer, S. (10-06-2016). *Supporting Teachers in Higher Education in Designing Test Items*. Amsterdam: Vrije Universiteit Amsterdam.

334. Bos, L.T. (15-06-2016). *Moving Beyond Words. Supporting Text Processing Using a Situation Model approach*. Amsterdam: Vrije Universiteit Amsterdam.

335. Vrugte, J. ter (16-06-2016). *Serious support for serious gaming*. Enschede: University of Twente.

336. Kock, Z.D.Q.P. (23-06-2016). *Toward physics education in agreement with the nature of science: Grade 9 electricity as a case*. Eindhoven: Eindhoven University of Technology.

337. Trinh Ba, T. (28-6-2016). *Development of a course on integrating ICT into inquiry-based science education*. Amsterdam: Vrije Universiteit Amsterdam.

338. Gerken, M. (29-06-2016). *How do employees learn at work? Understanding informal learning from others in different workplaces*. Maastricht: Maastricht University.

339. Louws, M.L. (06-07-2016). *Professional learning: what teachers want to learn*. Leiden: Leiden University.

340. Geel, M.J.M. van, & Keuning T. (08-07-2016). *Implementation and Effects of a Schoolwide Data-Based Decision Making Intervention: a Large-Scale Study*. Enschede: University of Twente.

341. Bouwer, I.R., & Koster, M.P. (02-09-2016). *Bringing writing research into the classroom: The effectiveness of Tekster, a newly developed writing program for elementary students*. Utrecht: Utrecht University.

342. Reijners, P.B.G. (02-09-2016). *Retrieval as a Cognitive and Metacognitive Study Technique to Learn from Expository Text*. Heerlen: Open University of the Netherlands.

343. Hubers, M.D. (08-09-2016). *Capacity building by data team members to sustain schools' data use.* Enschede: University of Twente.

344. Hsiao, Y.P. (23-09-2016). *Peer Support to Facilitate Knowledge Sharing on Complex Tasks*. Heerlen: Open University of the Netherlands.

345. Scheer, E.A. van der (23-09-2016). *Data-based decision making put to the test*. Enschede: University of Twente.

346. Bohle Carbonell, K. (28-9-2016). *May I ask you….? The influence of Individual, Dyadic, and Network Factors on the Emergence of Information in Exchange Teams*. Maastricht: Maastricht University.

347. Claessens, L.C.A. (30-09-2016). *Be on my side, I'll be on your side: Teachers' perceptions of teacher–student relationships*. Utrecht: Utrecht university.

348. Jansen in de Wal, J. (18-11-2016). S*econdary school teachers' motivation for professional learning*. Heerlen: Open University of the Netherlands.

349. Kock, W.D. de (24-11-2016). *The effectiveness of hints during computer supported word problem solving*. Groningen: University of Groningen.

350. Oonk, C. (07-12-2016). *Learning and Teaching in the Regional Learning Environment: Enabling Students and Teachers to Cross Boundaries in Multi-Stakeholder Practices'*. Wageningen: Wageningen University.

351. Beckers, J. (09-12-2016). *With a little help from my e-portfolio; supporting students' self directed learning in senior vocational education*. Maastricht: Maastricht University.

352. Osagie, E.R. (14-12-2016). *Learning and Corporate Social Responsibility. A study on the role of the learning organization, individual competencies, goal orientation and the learning climate in the CSR adaptation process*. Wageningen: Wageningen University.

Open Universiteit
www.ou.nl