Computers in Human Behavior 90 (2019) 117–130

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Embedded instruction to learn information problem solving: Effects of a whole task approach

Jimmy Frerejean[a,*], Gerdo J. Velthorst[b], Johan L.H. van Strien[a,c], Paul A. Kirschner[a,d], Saskia Brand-Gruwel[a]

[a] Welten Institute, Open University of the Netherlands, Valkenburgerweg 177, 6419 AT, Heerlen, Netherlands
[b] Iselinge University of Professional Teacher Education, Bachlaan 11, 7002 MZ, Doetinchem, Netherlands
[c] Universiteit Utrecht, Heidelberglaan 1, 3584 CS, Utrecht, Netherlands
[d] LET Research Unit, University of Oulu, Pentti Kaiteran katu 1, 90014 Oulu, Finland

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 paper 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 in the short term, 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. In addition, the improvements were no longer visible after five weeks. 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.

1. Introduction

Contemporary educational programs adopting resource-based, inquiry-based, or problem-based approaches confront students with problems, projects, or tasks that contain insufficient information to find the solution. Students 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 quality of information available on the Internet (Hill & Hannafin, 2001). Finding information online for educational purposes is termed information problem solving (IPS) and constitutes a complex cognitive skill that requires students to search for the necessary information in reliable online sources and combine this information to formulate a complete and correct solution. When performed correctly, this process leads to the construction of knowledge. The ease with which students navigate the Internet and uncover information might lead to the belief that students automatically develop IPS skills without any explicit instruction, but research shows this is not the case (Kirschner & De Bruyckere, 2017; Kirschner & van Merriënboer, 2013). Without formal training, students at all levels of education struggle with IPS (Miller & Bartlett, 2012; Rosman, Mayer, & Krampen, 2014; Van Deursen & van Diepen, 2013; Walraven, Brand-Gruwel, & Boshuizen, 2008).

Unfortunately, educational institutions often encounter problems with the implementation of IPS instruction (Badke, 2010). While its importance as an essential 21st century skill is acknowledged, most schools and teachers are poorly equipped to systematically integrate IPS training, often leading to subpar instruction in short library training sessions (Derakhshan & Singh, 2011; Probert, 2009). Research shows that whole-task approaches for teaching complex skills such as IPS show potential (Wopereis, Frerejean, & Brand-Gruwel, 2015, 2016), 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). This study presents such formal training in the form of whole-task IPS instruction embedded in a content knowledge domain and investigates its effects on students' IPS skills.

1.1. Information problem solving

Information problem solving is a complex cognitive skill requiring the application of constituent 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,

* Corresponding author.
E-mail addresses: j.frerejean@maastrichtuniversity.nl (J. Frerejean), gerdo.velthorst@iselinge.nl (G.J. Velthorst), j.l.h.vanstrien@uu.nl (J.L.H. van Strien), paul.kirschner@ou.nl (P.A. Kirschner), saskia.brand-gruwel@ou.nl (S. Brand-Gruwel).

https://doi.org/10.1016/j.chb.2018.08.043
Received 3 November 2017; Received in revised form 15 July 2018; Accepted 24 August 2018
Available online 25 August 2018

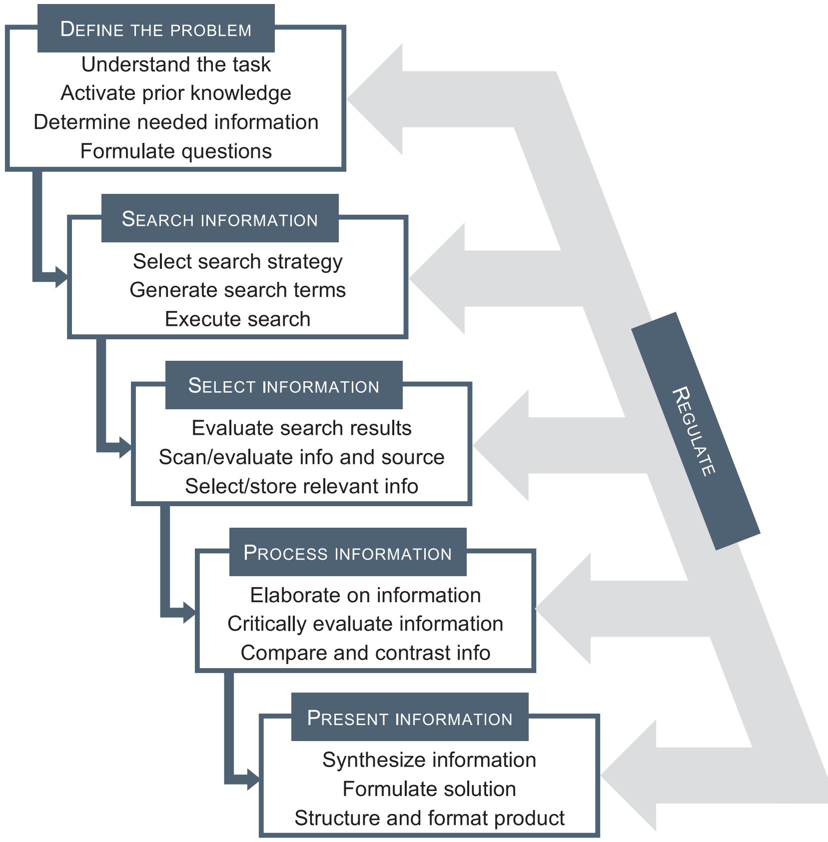

Fig. 1. Five-step systematic approach to information problem solving (based on: Brand-Gruwel et al., 2005; Frerejean et al., 2016).

2013). Fig. 1 presents a five-step approach to solving information problems based on a decomposition of the skill into constituent skills (Brand-Gruwel et al., 2005; Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2016).

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. Experts in IPS do this more explicitly and more often than novices, who hardly spend time on problem definition (Brand-Gruwel et al., 2005; Walraven et al., 2008). The 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. Students often struggle with query formulation and SERP evaluation (Walraven, Brand-Gruwel, & Boshuizen, 2009; Walraven et al., 2008). By judging the source's relevance and trustworthiness, the student determines if the information is useful. Evaluation of sources is often problematic, and novices tend to focus only on relevance of the source, while paying less attention to reliability (Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Walraven et al., 2009). However, research shows that students can be trained to generate more relevant search queries, adopt more evaluation criteria and select higher quality sources (e.g., Kroustallaki, Kokkinaki, Sideridis, & Simos, 2015). 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.

1.2. Instruction for information problem solving

For teaching such complex skills, the four-component instructional design (4C/ID) model (Van Merriënboer & Kirschner, 2018) advocates the design of four components (see Fig. 2): (1) *Learning tasks* form the backbone of the instructional blueprint. Learning tasks are based on authentic real-life situations that are encountered in practice and require the integration and coordination of skills, knowledge and attitudes. Learning tasks should contain sufficient built-in task support and guidance to assist the learner. Examples of task support mechanisms are the *completion strategy* (Van Merriënboer & De Croock, 1995; Van Merriënboer, 1990), 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 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).

1.2.1. Whole-task instruction

The 4C/ID model underlines the prevailing view 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

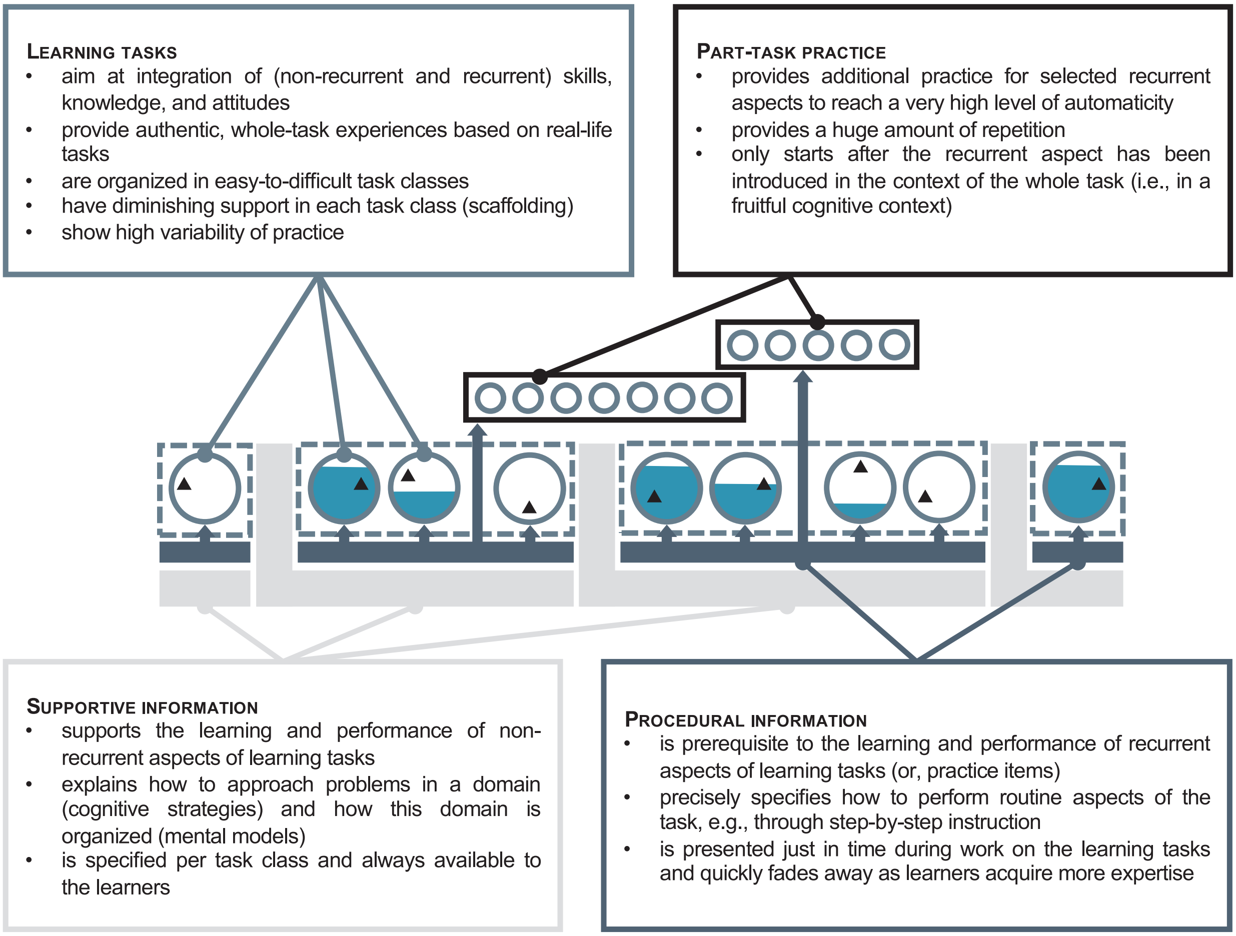

Fig. 2. Four-component instructional design (4C/ID) model (based on Van Merriënboer & Kirschner, 2018).

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 2-h 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 et al., 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.

In contrast, instructional interventions that focus only on improving specific aspects of the IPS skill, such as source evaluation skills (Britt & Aglinskas, 2002; Gerjets, Kammerer, & Werner, 2011), teach constituent skills separately. Such part-task training programs present little opportunity for integration and coordination of the constituent skills, which may hinder performance when these skills are needed in everyday settings or in complex assignments, such as a research project or writing a thesis (Van Merriënboer & Kirschner, 2018). Indeed, when Walraven et al. (2013) integrated part-task 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.

1.3. Embedded instruction

Van Merriënboer and Kirschner (2018) further recommend that training of IPS skills should be intertwined with the teaching of domain-specific skills. They describe IPS as a domain-general skill, which is not bound to a specific domain, yet must be taught in the context of one or more domains. Research on contextualization shows that embedding instruction within a meaningful context and presenting it simultaneously with domain-specific instruction can lead to more motivation, better engagement, deeper learning and improved transfer (Cordova & Lepper, 1996; Perin, 2011). In contrast, mere abstract training of domain-general skills in standalone sessions outside of a domain, is usually ineffective (Anderson, Reder, & Simon, 1996; Tricot & Sweller, 2014).

Prior research has investigated training programs where IPS instruction was embedded in a domain-specific course. For example, in primary education, Kuiper, Volman, and Terwel (2008) showed that fifth graders benefited 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, outperformed students following the regular curriculum. Similarly, Squibb and Mikkelsen (2016) showed that university students' information literacy skills improved most after receiving IPS instruction integrated in a writing curriculum, when compared to standalone library training or no training at all. These findings indicate that embedding IPS instruction in a domain-specific course is preferred over presenting IPS training separate from the domain-specific course (Farrell & Badke, 2015).

From these results we can expect that IPS training is most effective when it is taught using a whole-task approach, and when those learning tasks are embedded in the context of domain-specific instruction. Research on the effectiveness of IPS training combining these two characteristics is scarce. In one study, Brand-Gruwel and Wopereis

(2006) embedded an IPS course developed according to 4C/ID in a resource-based learning course about dyslexia for ten student teachers. Five other student teachers served as a control group and received no integrated IPS training. They found that the trained student teachers took more task requirements and needed information into account and judged sources more often. Trained students also spent more time processing information, regulated their process more often, and produced better products. In a study by Wopereis et al. (2008), 16 psychology students received embedded IPS instruction in a research methodology course. They found similar increases in the frequencies of constituent skills such as scanning information, judging information and regulating the process.

In summary, these two studies found that an embedded whole-task approach lead to more explicit, more frequent, and longer execution of many IPS skills. However, in both studies, the low number of participants prevents the researchers from drawing any strong conclusions and generalizing the findings. Secondly, their results focus mainly on frequencies and durations of performed constituent skills, and less on learning outcomes and actual increase of IPS performance. Third, the lack of a delayed posttest prevents any claims about the long-term effects of the interventions.

1.4. The present study

The present study attempts to further investigate the effects of embedded whole-task instruction in an ecologically valid setting and overcome the limitations from comparable studies by presenting less constraints to task performance, by measuring IPS performance in a more detailed fashion with an extra measurement weeks after the intervention, and by testing a larger sample. It can be expected that adopting a whole-task approach to IPS instruction in which the skill as a whole is trained, will lead to improvements in each of the constituent skills: problem definition, searching for information, selecting information, processing information, and presenting information (see Fig. 1). This study therefore 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 principles provided by the 4C/ID model, 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.

As this training makes use of whole-tasks that address all constituent skills of IPS, we expect that students receiving embedded training display more expert-like behavior (see section 1.1) in each area 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.

2. Method

2.1. 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 first cohort consisted of 75 student teachers (M_{age} = 19.0, SD = 1.56, 27 male, 48 female) in four classes and received the course with embedded IPS instruction. The following year's cohort of 80 student teachers (M_{age} = 19.2, SD = 1.72, 20 male, 60 female) also comprised four classes and received the regular course. In the remainder of this article, 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.

2.2. Materials

2.2.1. Regular course

The targeted 20-week course with a study load of 112 h 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.

2.2.2. 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, 2016) 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 *prompting triad:* an approach that emphasizes

Table 1
Overview of the IPS training.

Training element	Description
Supportive information	Seven instructional videos (total: 32 min), 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.

Table 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

parts of task execution with anticipative, instructional and reflective prompts (for a description of the prompting 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 1 shows an overview of the design of the online IPS training. Table 2 shows a timeline of the semester in both years and displays the differences and similarities.

2.3. 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.

2.4. 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 4 min to create a mind map of their prior knowledge, without consulting online sources. Students were then given 20 min 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 5 min. 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,*

Table 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.

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.

2.5. Data analysis

2.5.1. 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 5-min mark prompt. Table 3 presents an overview of all variables in the current study.

2.5.2. 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 0.989, indicating high agreement.

2.5.3. 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.

2.5.4. 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 1.

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 0.873 for *query relevance* and 0.956 for *systematic approach.* One researcher scored the remaining queries and students. To investigate Hypothesis 2 that trained students would display better search processes, MANCOVA's 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.

2.5.5. 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., 2016). *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 et al., 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 0–100 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 2.

Approximately 10% of all sources were scored by two raters, obtaining an intra-class correlation of 0.935 for trustworthiness and 0.989 for coverage. The interrater agreement for the *systematic approach* score was determined by double scoring 15 students and amounted to 0.755. 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.

2.5.6. Assessing processing of information

For assessing the processing of information, we investigated the *time spent on a source*. 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, we conducted a multiple linear regression analysis 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.

2.5.7. 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.

3. Results

Table 4 presents an overview of the means and standard deviations of the variables collected in this study.

3.1. 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 h *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.

3.2. Problem definition

To assess the skill *problem definition*, students' answers to the 5-min 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, we reject the hypothesis that trained students display more activities concerning problem definition (H1).

3.3. 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: $F(1, 133) = 0.792$, $p = .375$. 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*: $F(1, 119) = 41.499$, $p < .001$, $\eta^2_{\text{partial}} = .259$. 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: $F(2, 132) = 0.764$, $p = .468$. On the posttest, a significant difference was found between trained and untrained students: $F(2, 117) = 16.177$, $p < .001$, $\eta^2_{\text{partial}} = .217$. Subsequent univariate analyses showed no significant difference on *query relevance*: $F(1, 118) = 0.077$, $p = .782$, but a significant difference on *systematic approach* $F(1, 118) = 12.856$, $p < .001$, $\eta^2_{\text{partial}} = .098$. 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.

3.4. 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

Table 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)

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) = 0.488, 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.

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) = 0.893, 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.

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.

3.5. 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 5 provides an overview of the results of the multiple regression. On the pretest, model 1 explained 16.5% of variance (R^2_{adj} = 0.164, F (2, 1057) = 104.597, p < .001). Including *training* as a predictor did not improve the model (ΔR^2 = 0.000, F_{change} (1, 1056) = 0.159, 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} = 0.09, F (2, 1190) = 59.676, p < .001). Including *training* as a predictor marginally improved the model (ΔR^2 = 0.003, 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 5, which indicates that trained students spent approximately 5.5 s less on a source than untrained students. On the delayed posttest, model 1 explained 43.5% of variance (R^2_{adj} = 0.434, F (2, 964) = 371.082, p < .001). Including *training* as a predictor did not improve the model (ΔR^2 = 0.001, 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.

3.6. 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) = 0.456, p = .501, posttest F (1, 126) = 0.091, 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.

Table 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

4. Discussion

This study investigated effects of a curriculum containing embedded whole-task 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 et al., 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. Furthermore, a delayed posttest provided information on IPS performance five weeks after the training session.

Results show that activities pertaining to *defining the problem* were scarce, which is common for novices (Brand-Gruwel et al., 2005). This finding reflects results by Brand-Gruwel and Wopereis (2006) and Wopereis et al. (2008) who also found the time invested on problem definition activities was low. 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. In comparison, Argelagós and Pifarré (2012) show that an embedded IPS curriculum in secondary education was effective to increase the frequency of problem definition activities by using worksheets and driving questions. 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 (Brand-Gruwel et al., 2005). 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. This is more in line with expert behavior (Monchaux et al., 2015). The results therefore indicate that the training succeeded in developing a systematic approach that students could apply in the test setting. Unfortunately, this improvement disappeared after five weeks.

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. Similar improvements in efficiency of searching and selecting information were found in the study on embedded IPS instruction in secondary education by Argelagós and Pifarré (2012). In that study, students also selected better sources, while in the current study, the trained students did not select sources that were 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 reached 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. The improvements were only visible on the posttest and disappeared on the delayed posttest.

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). In addition, the training was only able to generate short-term learning effects, as trained and untrained students performed equally on all aspects on the delayed posttest five weeks after the training. In comparison, an evaluation of embedded instruction in secondary education by Argelagós and Pifarré (2012) showed stronger improvements on more of the constituent skills. The lack of strong learning effects in this study 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, we can be confident that both groups are 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.

4.1. 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. As such, we fully agree with Rieh, Collins-Thompson, Hansen, and Lee (2016) that 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.

Conflict of interest

The authors declare that they have no conflict of interest.

Appendices

Appendix A. (translated from Dutch)

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
mandatory school uniforms = 3 points
help against = 2 points
bullying = 2 points.
Total points for this query: 7/9 points (77.78%).

Pretest	Concept 1	Concept 2	Concept 3
3 points	Mandatory + uniform + school School uniform(s) Uniform(s) + school	Arguments Discussion Arguments for Arguments against Benefits Cons	Research Consequences

2 points	Mandatory + clothing + school Mandatory school clothing Dress code + school	Bullying Bullying behavior Equality Costs Expensive	Prevention Help … against Solutions Prevents Counteract Less Against
1 point	Dress code Clothing (+school) School clothing	Politics Opinions Argument Debate	Forum

Posttest	Concept 1	Concept 2	Concept 3
3 points	Media use Screen Blue light	Night's rest Sleep quality Sleeping behavior Sleeping pattern	Influence (of) Effect Relation Consequences Leads to Impact Discussion Research
2 points	Computer (use) Laptop (use) Smartphone (use) Mobile (use) Gameconsole iPad Using a computer Browsing the internet Gaming Television/tv	Sleep/sleeping Going to sleep Sleeping problems Poor sleep Sleeping rhythm REM-sleep Melatonin	Positive effects Negative effects Positive influence Promote Improve Disrupt
1 point	Multimedia Texting Internet Social media Computergames Games	Sleeplessness Concentration (problems) Exciting Sleep deprivation Lack of sleep Brain activation Brain activity	Advantage Disadvantage Disturbing factor Risks Dangers Damage Bad (for) Good (for) Less

Delayed posttest	Concept 1	Concept 2	Concept 3
3 points	Traffic safety Traffic + safe	Navigation system Navigation equipment GPS	Influence Consequences Effects Affect Discussion
2 points	(Traffic)accident (Traffic)mishap Driving behavior	TomTom Car navigation Navigationtools Reading map	Advantages Disadvantages Distract Risks Danger Consideration Opinions on Improves Worsens Better
1 point	Safe/safety Safer Unsafe Behavior in traffic Car + accidents	Navigation (use) TomTom use	Negative Minus Good Bad

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 navigation system 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 0–100. 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 B. (translated from Dutch)

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

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. https://doi.org/10.1007/s10618-011-0228-8.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher, 25*(4), 5–11. https://doi.org/10.3102/0013189x025004005.

Argelagós, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior, 28*, 515–526. https://doi.org/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. https://doi.org/10.5539/jel.v5n4p1.

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

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. https://doi.org/10.1080/11356405.2015.1092265.

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*. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.compedu.2009.06.004.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*, 485–522. https://doi.org/10.1207/S1532690XCI2004_2.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*, 715–730. https://doi.org/10.1037/0022-0663.88.4.715.

Derakhshan, M., & Singh, D. (2011). Integration of information literacy into the curriculum: A meta-synthesis. *Library Review, 60*, 218–229. https://doi.org/10.1108/00242531111117272.

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. https://doi.org/10.1037/0033-295X.100.3.363.

Farrell, R., & Badke, W. (2015). Situating information literacy in the disciplines. *Reference Services Review, 43*, 319–340. https://doi.org/10.1108/rsr-11-2014-0052.

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. https://doi.org/10.1016/j.chb.2016.03.048.

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. https://doi.org/10.1016/j.learninstruc.2010.02.005.

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. https://doi.org/10.1016/0001-6918(89)90007-3.

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

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

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. https://doi.org/10.1016/j.chb.2012.10.012.

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

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

Kroustallaki, D., Kokkinaki, T., Sideridis, G. D., & Simos, P. G. (2015). Exploring students' affect and achievement goals in the context of an intervention to improve web searching skills. *Computers in Human Behavior, 49*, 156–170. https://doi.org/10.1016/j.chb.2015.02.060.

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. https://doi.org/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 & Development, 48*, 45–66. https://doi.org/10.1007/BF02313485.

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 & Development, 57*, 61–77. https://doi.org/10.1007/s11423-007-9085-y.

Lucassen, T., & Schraagen, J. M. (2013). The influence of source cues and topic familiarity on credibility evaluation. *Computers in Human Behavior, 29*, 1387–1392. https://doi.org/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. https://doi.org/10.1023/a:1014617530297.

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.

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. https://doi.org/10.1016/j.ipm.2015.05.004.

Perin, D. (2011). Facilitating student learning through contextualization: A review of evidence. *Community College Review, 39*, 268–295. https://doi.org/10.1177/0091552111416227.

Probert, E. (2009). Information literacy skills: Teacher understandings and practice. *Computers & Education, 53*, 24–33. https://doi.org/10.1016/j.compedu.2008.12.018.

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*, 19–34. https://doi.org/10.1177/0165551515615841.

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. https://doi.org/10.1080/02602938.2014.950554.

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. https://doi.org/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. https://doi.org/10.1080/0144929X.2015.1066446.

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. https://doi.org/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. https://doi.org/10.1007/s10984-007-9031-2.

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

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. https://doi.org/10.5860/crl.77.2.164.

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. *The Journal of Continuing Education in Nursing, 44*, 278–284. https://doi.org/10.3928/00220124-20130501-78.

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review, 26*(2), 265–283. https://doi.org/10.1007/s10648-013-9243-1.

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. https://doi.org/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. https://doi.org/10.1016/j.intcom.2009.06.005.

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. https://doi.org/10.1037/1076-898x.11.4.237.

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. https://doi.org/10.2190/4NK5-17L7-TWQV-1EHL.

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

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. https://doi.org/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. https://doi.org/10.1002/acp.1250.

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. https://doi.org/10.1207/S15326985EP3801_2.

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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.compedu.2008.08.003.

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

Wirth, W., Sommer, K., von Pape, T., & Karnowski, V. (2016). Success in online searches: Differences between evaluation and finding tasks. *Journal of the Association for Information Science and Technology, 67*, 2897–2908. https://doi.org/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. https://doi.org/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 (Vol. Eds.), *Communications in Computer and Information Science: Vol. 552. Information Literacy: Moving Toward Sustainability. Vol. 552. Communications in Computer and Information Science: Vol. 552. Information Literacy: Moving Toward Sustainability* (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 (Vol. Eds.), *Information Literacy: Key to an Inclusive Society. ECIL 2016. Communications in Computer and Information Science. Vol. 676. Information Literacy: Key to an Inclusive Society. ECIL 2016. Communications in Computer and Information Science* (pp. 678–687). Cham: Springer International Publishing.

Jimmy Frerejean obtained his PhD at the Open University of the Netherlands, where his research focused on instructional design for information problem solving. He currently works as a researcher on instructional design for medical education at the School of Health Professions Education at Maastricht University.

Gerdo J. Velthorst is a learning scientist and teacher trainer at Iselinge University of Professional Teacher Education. He is lecturer in teacher's research practice courses and leading officer regarding institution-wide curriculum development and renewal.

Johan L. H. van Strien works as an educational consultant at Utrecht University, following a position as assistant professor at the Open University of the Netherlands. His academic interests include multiple-text comprehension, information problem solving, epistemic cognition, assessment, academic skills, and blended learning.

Paul A. Kirschner is Distinguished University Professor and professor of Educational Psychology at the Open University of the Netherlands and Professor of Education at the University of Oulu, Finland. His research deals primarily with instructional design and how to make teaching and learning more effective, efficient and enjoyable.

Saskia Brand-Gruwel is Professor of Learning Sciences at the Open University of the Netherlands. She is dean of the Faculty Psychology and Educational Sciences. Her research is about learning from the internet and instructional design for higher-order skills.