



The Dynamics Between Self-Regulated Learning and Learning Outcomes: an Exploratory Approach and Implications

Joep van der Graaf^{1,5} · Lyn Lim² · Yizhou Fan³ · Jonathan Kilgour³ · Johanna Moore³ · Dragan Gašević^{3,4} · Maria Bannert² · Inge Molenaar¹

Received: 2 June 2021 / Accepted: 31 May 2022 / Published online: 5 July 2022
© The Author(s) 2022

Abstract

Self-regulated learning (SRL) has been linked to improved learning and corresponding learning outcomes. However, there is a need for more precise insights into how SRL during learning contributes to specific learning outcomes. We operationalised four learning outcomes that varied on two dimensions: structure/connectedness and level/deepness of knowledge. Specifically, we assessed how surface knowledge measured with a domain knowledge test (independent concepts) and a concept map (connected concepts), and deep knowledge measured with a transfer test (independent concepts) and an essay (connected concepts) were associated with frequencies of SRL activities during learning, assessed by concurrent think aloud, while taking into account students' metacognitive and prior knowledge. Forty-four university students performed a 45-minute problem-solving task integrating information about three topics to write a vision essay on the future of education. Results of the pre-/post-test analysis revealed a learning gain in domain knowledge and concept maps. Low cognitive activities were associated with all knowledge measures, except the concept maps and transfer. Furthermore, specific low cognitive activities showed either a positive or negative association; for example, processing showed a positive association with essay, but a negative association with domain knowledge. High cognitive activities were associated with the essay (connected concepts), but not with the concept map. Both metacognitive activities and knowledge were related to transfer. To conclude, taking the level and structure of knowledge into account helps specify the association between SRL activities during learning and the related learning outcomes.

Keywords Self-regulated learning · Learning outcomes · Metacognition · High cognition · Deep strategies · Low cognition · Surface strategies

The Dynamics between Self-Regulated learning and learning outcomes: an exploratory Approach and Implications

An important competence students need is Self-Regulated Learning (SRL), especially relevant for “learning to learn” (European Union, 2019). Self-regulating learners use cognitive activities (read, practice, elaborate) to study a topic, and use metacognitive activities (planning, monitoring, evaluation) to actively monitor and control their learning and motivate themselves to reach learning goals (Schunk & Greene, 2018). Generally, SRL has been associated with improved learning (Schunk & Greene, 2018), but diverse effects have been found for different outcomes of learning (Bannert, 2006). Often, SRL is associated with improved deep knowledge, which can be measured with transfer tests (Bannert et al., 2009; Bannert, Sonnenberg, et al., 2015) and essays (e.g., Greene & Azevedo 2007, 2009), but its impact on surface knowledge is less clear. It is important to recognise that learning outcomes differ in how they assess knowledge and its representation in mental structures (Frey et al., 2017). Therefore, we systematically investigated the relations between particular SRL activities during learning, as measured with think aloud, and different learning outcomes in an exploratory study. We characterised four learning outcomes using two dimensions of mental representation of knowledge (de Jong & Ferguson-Hessler, 1996).

Self-regulated learning

SRL is depicted as a goal-oriented process, in which learners actively steer their learning (Winne & Hadwin, 1998). Models of SRL consist of the so-called CAMM activities, which are Cognition, Affect, Metacognition, and Motivation (Azevedo et al., 2018). SRL research often investigates cognition and metacognition (e.g., Winne, 2018). Cognition refers to the mental action of representing and processing information (Sternberg, 1981), and includes low-level information processing during learning, such as reading and repeating information, as well as high-level information processing like elaboration and organisation of the information processed (e.g., Bannert et al., 2014). Metacognition is the knowledge one has of their own cognitive processes (Flavell, 1979) and includes activities such as, planning, orientation, monitoring, and evaluation (e.g., Veenman 2013).

A commonly used theory of SRL is the COPES model (Winne, 2018a), which describes four loosely sequenced phases (Winne & Hadwin, 1998). The first phase entails the definition of the task. It includes the metacognitive act of orientation, which is searching the learning environment and activating existing prior knowledge to identify conditions relevant to the task. The second phase focuses on setting goals and creating plans. Via metacognitive planning, the learner constructs goals to work on the task and plans an approach to pursue those goals. In the third phase, the task is executed using cognitive tactics and strategies. A range of cognitive activities is used to perform the task and construct new knowledge and skills. Low-level cognitive activities (reading, repeating, and processing information) serve the goal of processing and understanding provided information and high-level cognitive activities (elaboration and organisation) help to deepen new understanding (King, 2002; Molenaar & Chiu, 2017; Volet et al., 2009). In addition, the self-regulated learner engages in the metacognitive acts of monitoring and control (Winne, 2018a). This means monitoring which information was relevant for learning and monitoring (changes in) progress towards set goals. Control can be exerted by changes in the enactment of tactics and strategies. In

phase four, learners reflect on their general approach and make changes for future learning (Winne & Hadwin, 1998). Thus, SRL is a multifaceted construct in which metacognitive, high, and low cognitive activities play a role to empower learning.

Learning outcomes

SRL has been associated mostly with learning outcomes that can be categorised as deep, such as transfer test and essay scores, whereas the role of SRL on surface knowledge is less clear (Bannert et al., 2009; Greene & Azevedo, 2007). Studies in education research have measured learning outcomes which often vary along two dimensions—the structure and level of knowledge measured (de Jong & Ferguson-Hessler, 1996). Structure is a dimension that ranges from unconnected and independent concepts to interconnected network structures of multiple concepts (Reif & Heller, 1982). In text comprehension research, this dimension has been introduced as *coherence* (McCarthy & McNamara, 2021). Independent concepts are pieces of information, such as a definition of artificial intelligence, that are not organised around fundamental concepts and thus are not connected to other relevant concepts within the domain (Alexander, 1992). Towards the other end of the dimension, we find highly connected knowledge structures. This means that the learner has organised pieces of information into a network of concepts with meaningful relations (Alexander, 1992). An example in artificial intelligence is organising supervised and unsupervised machine learning around the fundamental concept of machine learning. The relation of “supervised” with “machine learning” can be specified as “dependence on human supervision (by means of labelling data)”. The relation of “unsupervised” with “machine learning”, then would be “independence of human supervision”.

The other important dimension that should be considered more in SRL research is the level of knowledge, which ranges from surface to deep knowledge (de Jong & Ferguson-Hessler, 1996). A surface-level representation comprises concrete pieces of knowledge, whereas a deep-level representation encompasses understanding underlying concepts (Glaser, 1991). When presented with a problem about pulleys, the surface-level representation is made up of “a pulley”, whereas a deep-level representation also includes “conservation of angular momentum” (Chi et al., 1981). Deep knowledge is assumed to enable inferring and making analogies, which allows the transfer to new situations (Glaser, 1991).

Structure and level of knowledge can be measured towards each end of their dimensions. Surface knowledge measures that differ in terms of structure knowledge are a domain test and a concept map. Often unconnected, independent concepts are assessed using multiple-choice questions addressing specific concepts and procedures within a domain (Bannert et al., 2009). A well-known method that has the potential to reveal the global organisation of a learner’s knowledge network is a concept map (Lehmann et al., 2020; Thurn et al., 2020). A concept map visualises the interrelations between concepts within a domain, which may resemble how these concepts are organised in the mind (Hilbert & Renkl, 2008). Deep knowledge measures that differ in terms of structure knowledge are a transfer test and an essay. A common measure of deep knowledge is a far transfer test (Bannert et al., 2009). Far transfer makes use of a new situation to which previously constructed knowledge can be applied. Therefore, it requires students to learn with a deep conceptual understanding (Lin & Lehman, 1999). Furthermore, when self-explaining relevant concepts, students also need to

structure their knowledge. Such a mental model that contains deep, connected knowledge can be assessed using an essay test (Greene & Azevedo, 2007, 2009).

To sum up, learning outcomes can be characterised along two dimensions—the structure and the level of knowledge. We have placed our operationalisation of different learning outcomes in Table 1. Although the table might suggest a dichotomy between surface and deep knowledge, and independent and connected concepts, we consider them as dimensions. Thus, our learning outcomes are located on two continuums. For instance, the domain test is operationalised as more surface than deep and more independent than connected.

SRL and learning outcomes

Next, we are interested in the relation between different SRL activities and the aforementioned learning outcomes. As indicated, we focused on low and high cognitive activities, and metacognitive activities.

Low cognitive activities

Low cognition refers to cognitive processes involved in understanding given information (King, 2002). Low cognitive activities often occur in the phase where learners execute the task per se (phase 3 in the COPES model; Winne 2018a), such as when learners acquire and consolidate an initial knowledge base (Frey et al., 2017). They are, therefore, assumed to be relevant for all learning outcomes. Low cognition includes reading, repeating, and processing, and has also been referred to as surface strategy use (Dinsmore, 2017; Molenaar & Chiu, 2017) studied children from Grades 4 to 6 who learned in triads. Triads who showed more low cognition (an aggregate of low cognitive activities) had more new words in their essay than triads with less low cognition. Their findings indicated that low cognition helped build a common foundation of knowledge. In a study with university students, the low cognitive activities, reading and repeating, were more common in students with less knowledge than students with more knowledge, suggesting the relevance of low cognitive activities for acquiring knowledge (Bannert et al., 2014). In this case, reading can be considered a beneficial process to gain knowledge (Frey et al., 2017), while repeating can be unproductive (Bannert et al., 2014). Another activity deals with students' interactions with their products on a low cognitive level, called processing (Molenaar & Chiu, 2017), but the association of this specific low cognitive activity with learning outcomes has not been investigated yet.

High cognitive activities

In contrast, high cognitive activities, aimed at (re)organising previously acquired knowledge (King, 2002) may contribute to more connection among knowledge concepts. High cogni-

Table 1 Learning Outcomes Operationalised as Surface vs. Deep Knowledge and Independent vs. Connected Concepts

	Surface knowledge	Deep knowledge
Independent concepts	Domain test	Transfer test
Connected concepts	Concept maps	Essay

tion has also been referred to as deep strategy use (Dinsmore, 2017). For instance, knowledge organization can be fostered when students combine the purposes and consequences of machine learning when they are making notes. The study by Molenaar & Chiu (2017) also showed that triads with more high cognition had a higher essay quality, measured as originality, than triads with less high cognition. Previous studies have also demonstrated that the process of organising selected information in a coherent structure aids acquisition of domain-specific knowledge (e.g., Cook & Mayer 1988).

Metacognitive activities

Finally, metacognitive activities can help to deepen understanding (Bannert et al., 2009). Metacognitive activities consist of orientation, planning, regulation of cognitive activities, monitoring execution of planned actions, and evaluation of the outcome of task processing (Efklides, 2008; Veenman & Elshout, 1999). When studying students' learning process, Lin & Lehman (1999) found an association between the metacognitive activities of planning and monitoring, and creating deep knowledge, assessed with a transfer test in college students. They explained that planning activates prior knowledge to integrate new knowledge. Monitoring helps by tracking knowledge development and identifying potential gaps and misconceptions that can be resolved. The association between metacognitive activities and deep knowledge has often been replicated (e.g., Bannert 2006; Bannert et al., 2009).

Prior and metacognitive knowledge

Although the focus of this paper was on associations between the frequency of SRL activities during learning and learning outcomes, we acknowledge the relation between a learner's resources and SRL activities during learning. Learners with sufficient resources for dealing with the demand of self-regulation seem to be able to successfully engage in SRL and cope with the task requirements (Seufert, 2019). Two such resources are prior domain knowledge and metacognitive knowledge (Seufert, 2019), which are also two of the internal conditions in the COPES model (Winne, 2018a). According to the expertise reversal effect (Kalyuga, 2007), learners with more prior domain knowledge perceive tasks as less intrinsically loading. They will be able to compensate for additional extraneously loading aspects imposed by the task. For example, in a study that used think aloud to capture SRL, prior domain knowledge was positively associated with metacognitive activities (planning and monitoring), using diverse strategies, and elaboration, suggesting more effective learning (Moos & Azevedo, 2008). Note that more prior knowledge does not mean more regulation; instead, these results suggest a different and more effective way of regulation. The next step would be to identify how these regulation activities are associated with posttest measures.

Whereas prior domain knowledge constitutes knowledge about a specific topic of study, metacognitive knowledge is knowledge about different cognitive approaches and their usefulness in other contexts (Händel et al., 2014). According to the COPES model, internal conditions, including metacognitive knowledge, are elements the learner perceives could affect work on the task (Winne, 2018a). While metacognitive knowledge has been associated with academic achievement in reading and mathematics (Neuenhaus et al., 2011) and models have formulated how metacognitive knowledge can affect SRL (Efklides, 2008), we are unaware of studies explicitly addressing the association between metacognitive knowl-

edge and SRL *during* learning. Previous research does provide two clues. Firstly, sufficient resources (Seufert, 2019) seem to help students consider relationships among content, which is a central element in SRL (Moos & Azevedo, 2008). Secondly, timed tasks are often used in SRL research (e.g., Bannert et al., 2014; Moos & Azevedo, 2008; Deekens et al., 2018), limiting the number of activities that can be performed and imposing an extraneous load on the learner. Indeed, increased time pressure has been associated with increased extraneous cognitive load (Barrouillet et al., 2007) and less-effective cognitive activities (e.g., Sidi, et al., 2017). Thus, having sufficient resources can help learners deal with time pressure. More specifically, metacognitive knowledge might help learners choose and perform effective cognitive activities, such as high cognition, resulting in better learning outcomes, at the expense of less effective cognitive activities, such as low cognition.

The Present Study and Hypotheses

As described so far, the beneficial effects of SRL activities during learning have been associated with learning outcomes. Different effects have been found depending on the conceptualisation of learning outcomes. However, a systematic analysis of how different SRL activities during learning relate to different learning outcomes has yet to be conducted. There have been few attempts, to our knowledge, to contrast different learning outcomes. Research by Deekens and colleagues (2018) is a notable exception. In two studies, they used think aloud to measure SRL during learning and analysed associations between SRL frequencies and pre- and post-test measures. In both studies, the SRL activities labelled as monitoring (part of metacognition) were positively associated with high cognition, which in turn were positively associated with post-test performance on both a declarative (domain test) and conceptual knowledge (essay) measure. Low cognitive activities negatively associated with essay quality (Deekens et al., 2018). Although surface knowledge of independent concepts

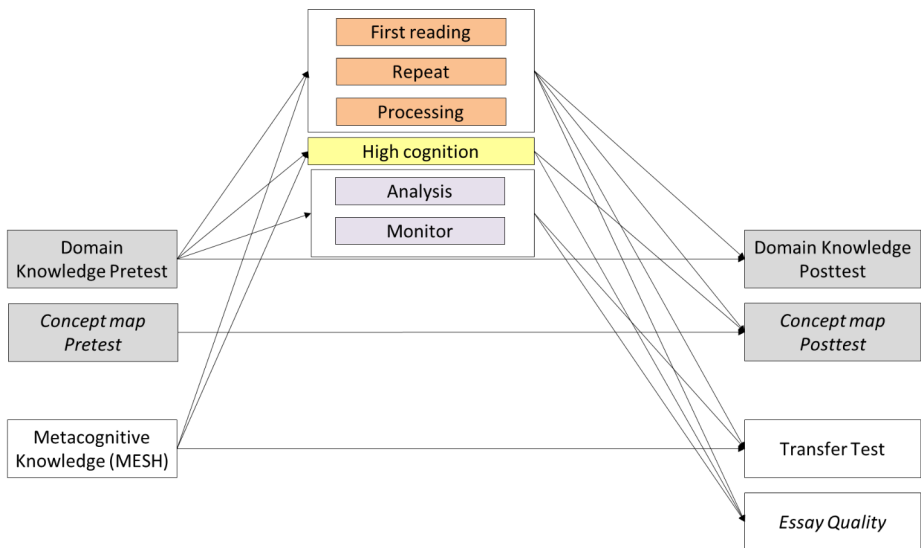


Fig. 1 Our Conceptual Model with Hypothesised Associations

(domain test) and deep knowledge of connected concepts (essay) was assessed, they were not contrasted with surface knowledge of connected concepts and deep knowledge of independent concepts. In addition, it would be useful to determine which low cognitive activities contribute in what way to learning, because some low cognitive activities might be needed when prior knowledge is low (Dinsmore & Alexander, 2016), implying a mediation effect of prior knowledge on SRL to learning outcomes on the posttest. Therefore, the present study investigated the association between SRL activities (metacognitive, high cognitive, and low cognitive) and learning outcomes (deep vs. surface knowledge, and independent vs. connected concepts), see Fig. 1. This study was exploratory, due to the small number of participants.

In a pre-/post-test design, students performed a learning task of 45 min, during which they had to read about three topics and write an essay about the future of education, see Fig. 2. We recorded think aloud during learning and coded it to identify SRL activities. Our analyses identified the associations between SRL (low and high cognition, and metacognition) and learning outcomes (domain test, concept maps, transfer test, and essay). We also controlled for metacognitive knowledge, as assessed at the pretest.

The present study aimed to identify the associations between SRL activities, measured as frequencies of think aloud codes, and learning outcomes, measured as prior knowledge at pretest and learning products at posttest. Our hypotheses concern the association of frequency of SRL activities during learning with particular learning outcomes, see Fig. 1. Hypothesis 1: We hypothesised that low cognitive activities would be associated with all learning outcomes: Reading can contribute to domain knowledge (H1a) (Frey et al., 2017), while repeating might be unproductive for all learning outcomes (H1b) (Bannert et al., 2014). Also, we explored the role of processing because its relation to learning outcomes has not been investigated yet (H1c). Hypothesis 2: We hypothesised high cognitive activities would be associated with measures of connected knowledge (Cook & Mayer, 1988). Hypothesis 3: We hypothesised metacognitive activities, both analysis and monitor, would be associated with measures of deeper knowledge (Lin & Lehman, 1999). In addition to our main hypotheses, we expected prior knowledge to be negatively related to low cognitive activities and positively with high cognitive and metacognitive activities (Moos & Azevedo, 2008). We expected metacognitive knowledge, as assessed on the pretest, to be associated with cognitive activities (Händel et al., 2014) and transfer scores (Bannert et al., 2014; Bannert & Reimann, 2012).

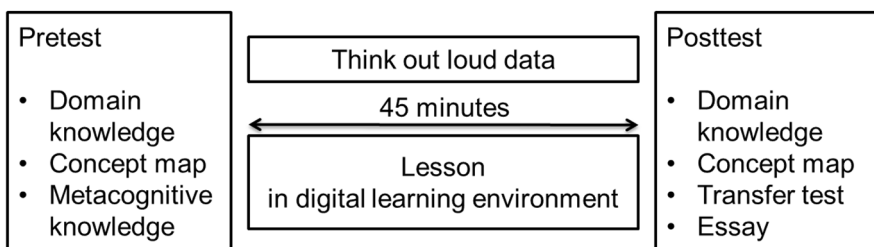


Fig. 2 A Graphical Overview of the Present Study. (Note. The essay was written during the learning session and was, therefore, regarded as a posttest measure.)

Methods

Participants

University students ($n=46$) with an average age of 21 years ($SD=3$ years) participated in the present study. Two participants were removed from the analyses, because there were problems with their think alouds. Thus, analyses were conducted with 44 participants (34 female and ten male). Thirty-nine participants were enrolled in a bachelor's degree program and five in a master's program¹. They were from a wide range of degree programs, of which psychology (11) and communication science (six) were the most common. We informed participants about the present study, and they were given the opportunity to ask questions, after which they gave active consent to collect data. Our research lab's ethical committee approved the present research.

Procedure

Students started with the pretest, which consisted of demographic questions, a domain knowledge test, three concepts maps (one per topic), and a metacognitive knowledge questionnaire. Next was the learning session. Students' task was to write an essay about the future of education using the provided informative texts. They were given 45 min to read the texts and write an essay. We recorded think aloud during their learning session. The final part was the posttest. Students completed three knowledge tasks: the same domain knowledge test and concepts maps, and a transfer test.

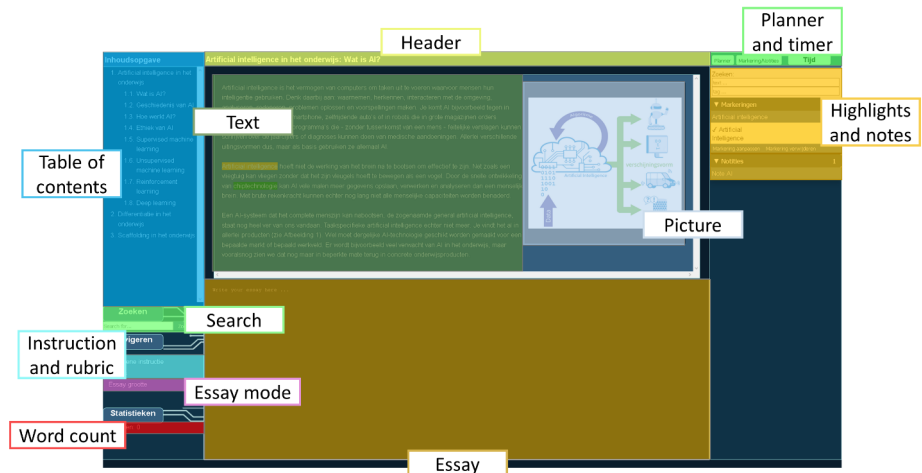


Fig. 3 The Digital Learning Environment

¹ Preliminary analyses showed that master's scored within the same range as bachelor's students on all measures.

Materials

Apparatus

The learning environment was presented via a laptop on a separate monitor (23 inch; 1920×1080 pixels) to the participants. Windows 10 with default settings was installed on the laptop. Students used a keyboard and mouse, which were connected to the laptop. The learning environment was created for this study's purpose. It ran on a local PHP-server and was presented via an internet browser.

Learning environment

The learning environment consisted of three panels, see Fig. 3. The left-hand panel was used to present the menu, a search function, links to the instruction and rubric, a button to change the essay mode (small, medium, or large essay), and a count of the number of words in the essay. The middle panel presented informative texts (of which six pages also had a picture) and the essay. The right-hand panel was used to interact with four tools: a planner, timer, highlighter, and note-taker.

Regarding the texts, there were instructions, a rubric, and informative texts about three different topics. The instructions were on the landing page. Students were instructed to write an essay of 300 to 400 words in 45 min. The essay should incorporate information from the three learning topics: artificial intelligence (AI), differentiation, and scaffolding. Furthermore, the essay should offer a vision of the future of education based on these topics. The instruction explicitly stated that 45 min is a short time and that participants could skip texts and start writing the essay if they wanted. Most of the text was relevant for the essay, but some parts were not. Students had access to the essay rubric, which contained the details of the essay assessment criteria based on the learning instruction and goals. The informational texts addressed: AI—what is AI, how does AI work, and four common forms of AI (Van Wetering et al., 2019); differentiation—what is differentiation, and how to apply it in the classroom (Deunk et al., 2018); and scaffolding—roots of scaffolding, what is scaffolding, and applying scaffolding (Reiser & Tabak, 2014). These three topics were chosen because of their potential relevance for education in 2035. To elicit decisions from the students about what to read, we added irrelevant texts for AI—history of AI (Russell et al., 2010), differentiation—standards for teacher education (Darling-Hammond, 2017), and for scaffolding—cognitive apprenticeship (Collins & Kapur, 2014). All texts were presented in Dutch. The original English texts were translated by the author(s) of the current paper, and the translations were discussed and finalised in collaboration with experts in the respective fields.

To draft the essays, the learning environment had a text field, where students could type. The number of words in the essay was automatically detected and presented. The essay had three modes, which were the read mode (small-sized essay), a hybrid mode for reading and writing (medium-sized essay) as in Fig. 3, which filled half of the page, allowing a hybrid reading-writing mode, and write mode (large essay), which almost completely filled the page. The default mode of the essay area was the read mode and students could change the essay mode via the essay size button.

With respect to the tools, students could use a search function, planner, timer, highlighter, and note-taker. The search function could be used to type in a search term, and matching

results from the informative texts were presented. A student could click on a result to go to that page or cancel the search. The planner showed a timeline of 45 min with six blocks below it. Each block represented an activity: orientation, processing AI, processing differentiation, processing scaffolding, essay writing, and free choice. These blocks could be dragged to the timeline to create a plan. The free choice could be used to type in an activity of the student's choice. The timer displayed the remaining time, but only when students clicked on it. The displayed time would disappear after two seconds. The highlighter and

Table 2 Main and Subcategories for Coding Think Aloud, Including Descriptions and Examples

Coding category	Final code	Description	Example
Metacognition			
Orientation	<i>Analysis</i>	Orientation on the learning-related activities; on prior knowledge; on the task, and feeling about the task.	“Okay, so four things are important in the assignment”
Planning	<i>Analysis</i>	Planning of the learning process by arranging activities and determining strategies. Proceeding to the next topic. Goal-directed search.	“I will explain these topics in my essay”
Monitoring	<i>Monitor</i>	Monitoring and checking the learning process; checking of progress according to instruction or plan. Standalone utterances, such as “Hm” and “Okay”.	“This information does not appear to be relevant” / “I only have a few minutes left to complete the assignment”
Evaluation	<i>Monitor</i>	Evaluation of the learning process; checking of content-wise correctness of learning activities. Saying that one owns work is correct.	“Is AI part of scaffolding?”
Low cognition			
First time reading	<i>First reading</i>	Reading out loud information from the texts and superficial describing of pictorial representations.	“Artificial Intelligence is the ability of ...” (first time)
Rereading	<i>Repeat</i>	Rereading of information in the text or figures.	“Artificial Intelligence is the ability of ...” (second time)
Superficial repetition	<i>Repeat</i>	When the intention is to memorise information, learn it by heart, or use it for the essay, notes, or highlights.	“Artificial Intelligence is the ability of ...” (2 or more times)
Superficial writing down	<i>Repeat</i>	Copying information by typing it in a note or in the essay, or by creating highlights.	“Artificial Intelligence is the ability of ...”
Processing	<i>Processing</i>	Rereading or editing of essay, highlights, and notes. Search in highlights and notes.	“I should start a new paragraph here”
High cognition			
Elaboration	<i>High cognition</i>	Connecting content-related comments and concepts; reasoning and association. Asking questions. Elaboration can be done using prior knowledge, text, figures, highlights, notes, and the essay.	“This means that a similarity between Artificial Intelligence and humans is that both learn”
Organisation	<i>High cognition</i>	Organising content by creating an overview; summarising; adding information generated by oneself; and editing information by rephrasing or integrating the information with prior knowledge.	“To summarise, ... are important for future education”

note-taker functioned similarly. Students started by selecting text. Then they could choose to highlight or take a note. It was optional for both highlights and notes to select any tags to be associated with them. The final step was to save the highlight and, for notes, to insert a note or a note title and save the note.

Coding SRL in think aloud protocols

Students were instructed and trained to think aloud. There was a short training beforehand whereby the experimenter demonstrated how to think aloud, and students could practise. When a participant fell silent, the experimenter reminded them to continue thinking aloud. During the learning session, utterances were recorded. Utterances were categorised as SRL activities using a coding scheme. Our coding scheme was based on previously developed coding schemes (Bannert, 2007; Molenaar et al., 2011). We coded metacognitive, high and low cognitive, motivational (activities indicating positive or negative appraisal), procedural (activities concerning the procedure of learning), and non-codable utterances (such as murmuring), see Table 2. Only metacognitive and cognitive activities were used for analyses, because other categories had a low frequency and were not relevant for our research questions. Two trained raters coded the utterances. Before coding, segments were created based on sound detection. The length of these segments was changed in rare cases, where sound detection went wrong (when a participant spoke very softly or did not speak, but there was another sound, such as moving the keyboard). Coding was done within ELAN software (ELAN, 2020). We then calculated a modified kappa, which takes potential differences in segment length into account (Holle & Rein, 2015). Inter-rater reliability was substantial (Munoz & Bangdiwala, 1997), $\kappa=0.53\text{--}0.65$ ($\kappa_{\max}=0.81\text{--}0.82$). Due to a low frequency and in line with the theoretical framework described in the introduction, we merged four categories, comparable to Engelmann & Bannert (2019). We merged orientation and planning into *analysis*; monitoring and evaluation into *monitor*; rereading, superficial repetition, and superficial writing down into *repeat* (note that superficial repetition and superficial writing down capture behaviour that include literal repetition of information from the text and rereading captures reading the same words again); and elaboration and organisation into *high cognition*.

Knowledge tests

Domain test Domain knowledge was assessed at pretest and posttest. A domain test was developed to assess surface knowledge of independent concepts. Students answered 30 questions with four response options in total, ten questions per topic. Questions addressed conceptual knowledge. An example was: “What is unsupervised machine learning?” with the options, A: “You teach the algorithm what the relationships are between data labelled by humans”, B (correct answer): “You ask the algorithm to cluster data itself by finding patterns in a dataset”, C: “You teach the algorithm what the data are, which are needed to perform a task”, and D: “You ask the algorithm to choose its own data and use it to perform a task”. When the correct answer was chosen, one point was scored. Zero points were given for incorrect answers. The maximum score was 30. Reliability was acceptable (Kline, 2000), $\alpha=0.60$, $\lambda_2=0.65$, $\omega_1=0.68$, at pretest and $\alpha=0.59$, $\lambda_2=0.64$, $\omega_1=0.66$, at posttest.

Concept maps Concept maps were assessed at pretest and posttest. Three concept maps were used to assess surface knowledge of connected concepts. There was one concept map per topic. Three experts created a concept map based on the informative texts per topic. Experts were free to choose the concepts they included. One of the authors made a synthesis that resulted in three expert concept maps with 15 concepts per topic. Concepts that were in only one of the three experts' maps were omitted from the final expert concept map. Students' task was to organise these 15 concepts in such a way that the concept map helps to explain the topic and to connect concepts that they thought were related. Thus, students were given a fixed set of 15 concepts per map. Concept maps were scored by comparing the links between concepts and the width of the concept map to the expert concept map, as proposed by (Pirnay-Dummer et al., 2010). This means that two similarity indices were calculated per concept map, which indicated a close similarity to the expert maps when they were close to 1 and a large deviation when they were close to 0. The first similarity index was the number of correct links. A link was counted as correct when two concepts were connected that were also connected in the expert map. The second index was the path length. The length of the longest path in a concept map was compared to the expert map reflecting a students' range of connected knowledge (Pirnay-Dummer et al., 2010). We wanted to create one concept map score for pretest and one for posttest, for parsimony, to be used in subsequent analysis. To verify that two indices per topic for three topics composed a measure of connected surface knowledge, separately for pretest and posttest, we conducted a CFA. We specified a pretest component, a posttest component, and a path between them. Furthermore, we added covariances between the correct link scores of the same topic between pretest and posttest. The model showed that the number of correct links and path length of differentiation on the pretest did not load on the pretest concept map component, which might be explained by the students' low prior knowledge of this topic. We excluded these from the final model, which had a good fit, $\chi^2(32)=35.359$. $p=.313$, CFI=0.979, RMSEA=0.049, 90% CI [0.000–0.127], SRMR=0.078. Thus, the data supported a general concept map score on the pretest of AI and scaffolding, and a general concept map score on the posttest of AI, differentiation, and scaffolding. Therefore, we calculated an average of four indices at the pretest and six indices at the posttest to analyse concept maps.

Metacognitive knowledge Metacognitive knowledge was assessed at the pretest stage using the MESH (Bannert et al., 2015) based on a previous questionnaire (Händel et al., 2013). The participants were asked to read seven learning-related scenarios and to rate five to six statements per scenario. The statements referred to strategies that varied in their degree of effectiveness for the given situation. The participants filled out the questionnaire by ticking one option per statement. The options ranged from "not useful" to "very useful", a six-point Likert scale. Responses were scored by comparing how statements within a scenario were rated. One point was given whenever one of the 43 key comparisons was in line with the expert rating. A comparison was in line when one statement was preferred over the other in the same way as in the expert rating. Thus, a high score means high metacognitive knowledge. One statement was missing in the present study, resulting in 41 key comparisons in total and, thus, a maximum score of 41. Reliability was good (Kline, 2000), $\alpha=0.92$, $\lambda_2=0.92$.

Transfer test Transfer of knowledge was assessed at the posttest. A transfer test was developed to assess deep knowledge of independent concepts. Ten questions addressed the transfer of AI to the medical domain. There were four answer options. An example was: "Which

is an example of how artificial intelligence has been used in hospitals?” with the options, A: “Sophisticated tube transport system to transport patient records within the hospital”, B: “Using robot vacuums for cleaning hospital floors to minimise infection”, C: “Advanced wireless communication system which minimises disruption to equipment”, and D (correct answer): “Using sophisticated algorithms to diagnose diseases”. When the correct answer was chosen, one point was scored. Zero points were given for incorrect answers. Four questions were removed to achieve acceptable reliability. Thus, the maximum score was 6. Reliability was acceptable (Kline, 2000), $\alpha=0.60$, $\lambda_2=0.62$.

Essay The essay was written during the learning session and scored by two independent raters using a coding scheme aligned with the rubric. The essay was scored to assess deep knowledge of connected concepts. The coding scheme described five categories: (1) explanation of three topics (maximum of 9 points; 3 points per topic), (2) connection of the topics to the future of education (maximum of 6 points; 2 points per topic), (3) suggestions of how the topics can be used in the future (maximum of 3 points; 1 point per topic), (4) originality, which was a scaled inverse of a copy score obtained using WCopyfind (maximum of 3 points), and (5) the number of words: 250–450 words resulted in 3 points, 200–249 or

Table 3 Descriptive Statistics (Median with 25th and 75th percentile)

	<i>Mdn</i> (25th, 75th)	<i>M</i> (<i>SD</i>)	<i>Skewness</i> (<i>SE</i>)	<i>Kurtosis</i> (<i>SE</i>)
Pretest				
1. Metacognitive knowledge	26.50 (15.00, 33.00)	24.25 (9.58)	-0.27	-1.57
2. Domain knowledge	17.00 (14.00, 19.00)	16.36 (3.99)	-0.78	0.61
3. Concept maps	0.49 (0.41, 0.56)	0.46 (0.14)	-1.44	2.10
SRL activities				
4. First reading	247.00 (117.50, 282.75)	249.66 (102.12)	1.91	5.39
5. Repeat	55.50 (35.25, 103.75)	71.18 (52.32)	0.94	-0.12
6. Processing	40.00 (17.50, 67.75)	49.45 (44.15)	1.11	0.84
7. High cognition	210.50 (140.00, 248.25)	197.55 (87.53)	-0.01	-0.10
8. Analysis	85.00 (60.50, 108.25)	86.64 (38.06)	0.49	0.04
9. Monitor	162.00 (122.50, 192.50)	161.98 (55.63)	0.29	-0.14
Posttest				
10. Essay	15.00 (12.00, 17.00)	13.95 (4.87)	-1.13	1.08
11. Transfer test	5.00 (3.00, 5.25)	4.30 (1.50)	-0.70	-0.15
12. Domain knowledge	21.00 (18.00, 23.00)	20.57 (3.57)	0.22	-0.89
13. Concept maps	0.54 (0.51, 0.59)	0.52 (0.11)	-1.83	4.24

451–500 words in 2 points, 150–199 or 501–550 in 1 point, and any number below 150 or above 550 in 0 points. Thus, the maximum score was 24 points. Two raters were trained to score the essays. They were instructed in the use of the coding scheme and discussed any discrepancies during training to reach agreement. Inter-rater reliability was almost perfect (Munoz & Bangdiwala, 1997), Fleiss-Cohen $\kappa=0.89$.

Data analysis

Data used for analyses were: number of correct answers on the metacognitive knowledge test (1 in Tables 3 and 4), number of correct answers on the knowledge test on pretest (2) and posttest (12), mean of the similarity and path length index of the concept maps at pretest (3) and posttest (13), frequencies of SRL activities: first reading (4), repeat (5), processing (6), high cognition (7), analysis (8), and monitor (9), number of points on the essay based on a coding scheme (10), and number of correct answers on the transfer test (11).

First, we tested whether learning occurred using Wilcoxon signed-rank tests. Second, as a first investigation of the associations and in preparation of the structural equation modelling (SEM), we calculated correlations using the Spearman rank coefficients, because not all data were normally distributed. Third, we conducted SEM using the lavaan package (Rosseel, 2012) in R (R Core Team, 2020). The Maximum Likelihood (ML) estimator with Huber-White robust statistics was used for the estimation of the model and path coefficients. Not all variables had a normal distribution, see Table 3. Therefore, we first modelled different distributions of the SRL activities (frequencies of think aloud), as Greene et al., (2011) proposed. This analysis showed that regular Ordinary Least Squares (OLS), $BIC=6358.29$, outperformed the other models with our data: Poisson, negative binomial, and their zero-inflated variants, $BIC's > 7700$. We continued with an OLS model, but we requested robust statistics to deal with deviations from non-normality.

The type of model we built was a so-called structural model without latent variables. The way variables entered the model is as follows: we ordered variables in the model to reflect our pre-/post-test design, where pretest measures were used as independent variables, SRL activities as mediators, and posttest measures as dependent variables (see Fig. 1). Note that essay quality was regarded as a posttest measure, because it was the product of the learning process (and most of the time, it was also the final action performed during the learning session). We first tested a hypothesis model with paths specified based on our hypotheses. Then, we respecified the model to improve model fit. Model fit is acceptable when CFI and TLI are close to 0.95, RMSEA close to 0.06, and SRMR close to 0.08 (Hu & Bentler, 1999). Respecification was done by trimming down the model, removing paths with p -values above 0.10. Then, paths were added based on modification indices and whether they were in line with correlations, in line with the exploratory nature of the present study. We stopped when the Chi-square test indicated a good model fit ($p > .05$).

In the final model, mediation effects (indirect effects), direct effects, and total effects were tested. Mediation effects were tested by multiplying the paths involved in the mediation. An example was the indirect effect of metacognitive knowledge on pretest via first reading to essay quality. The presence of direct effects in the final model depended on the respecification step. During respecification, empirical tests were used to trim the model. This meant that in two cases in the present analyses, mediation was tested, where there was no direct effect. Note that a significant direct effect is not a prerequisite for significant

mediation (Hayes, 2009). The total effect was the sum of the direct and indirect effect and thus, was only calculated when both direct and indirect effects were present.

Results

Before addressing the research questions, we tested whether learning occurred. Results showed that learning occurred, domain test scores were higher on posttest than pretest, $p < .001$, $r = .51$, (a large effect), and concept map scores as well, $p = .003$, $r = .31$, a medium effect (Cohen, 1992). The descriptive statistics are depicted in Table 3, and correlations are depicted in Table 4. The median score on the domain knowledge at the pretest was 17 correct and at posttest 21 correct. The increase in concept maps was from 0.49 to 0.54. The frequencies of the SRL activities showed that first reading was most frequent (274 times), and processing was least frequent (40 times).

The correlations showed associations between the SRL activities. Low cognitive activities, first reading, repeat, and processing, and high cognition were correlated. Metacognitive activities, analysis and monitor, were correlated to a lesser extent with the other SRL activities. Furthermore, learning outcomes were correlated with prior knowledge, except for the essay score. The essay was correlated with low and high cognitive activities. The transfer test showed two correlations with metacognitive activities: analysis, $p = .071$, and monitor, $p = .054$. Domain knowledge correlated negatively with repeat, $p = .071$. Concept maps did not correlate with SRL activities.

In line with our pretest-posttest design, we constructed a model with pretest measures predicting SRL activities (as mediators) and both pretest measures and SRL activities as predictors of posttest measures. Based on our hypotheses, we constructed a model, which had a poor fit. We respecified the model based on the correlations and modification indices. After respecification, the final model had an acceptable fit (Hu & Bentler, 1999), see Fig. 4. Only the SRMR showed a lesser fit, which indicated more variables might be associated. This could be due to associations within SRL activities and/or variables being dropped in the SEM model due to their low statistical detectability in the case of multiple independent variables. Therefore, fit was considered as acceptable. All specified paths can be found in Fig. 4 and Appendix A. No additional paths were specified.

We found *first reading* to be negatively associated with essay quality, and not to be associated with other learning outcomes, in contrast to hypothesis H1a, see Table 5. We found a negative association of *repeat* with domain knowledge, in line with hypothesis H1b. *Processing* was negatively associated with domain knowledge and positively with essay quality. We did not have explicit expectations about *processing* and explored its associations (H1c). Concept maps and transfer were not associated with low cognitive activities. We found that *high cognition* was positively associated with essay quality, but there was no association with concepts maps. This partially supported our hypothesis that *high cognition* would be associated with measures of connected concepts (H2). We found *analysis* to be positively associated with transfer, but no associations of metacognitive activities and essay quality were found. This partially supported our hypothesis that *metacognition* would be associated with measures of deep knowledge (H3).

Regarding the effect of prior knowledge, we found, as expected, associations of domain knowledge with low cognition, namely *processing*, and of metacognitive knowledge with

Table 4 Spearman's Rank Correlations of Learning Outcomes and SRL Activities

	1	2	3	4	5	6	7	8	9	10	11	12	13
Pretest													
1. Metacognitive knowledge	1												
2. Domain knowledge	0.23	1											
3. Concept maps	0.23	0.43**	1										
SRL activities													
4. First reading	-0.26 [†]	-0.26 [†]	1										
5. Repeat	-0.15	-0.13	-0.14	0.34*	1								
6. Processing	0.18	0.17	-0.01	-0.67***	-0.18	1							
7. High cognition	0.24	0.18	0.17	-0.51***	-0.69***	0.37*	1						
8. Analysis	0.21	-0.10	0.04	-0.28 [†]	0.21	0.28 [†]	-0.16	1					
9. Monitor	0.20	0.26 [†]	0.20	-0.21	0.08	0.24	-0.01	0.25	1				
Posttest													
10. Essay	0.20	0.17	0.12	-0.61***	-0.52***	0.51***	0.61***	0.07	0.04	1			
11. Transfer test	0.51***	0.25	0.45**	-0.12	0.03	0.06	-0.06	0.27 [†]	0.29 [†]	0.02	1		
12. Domain knowledge	0.20	0.32*	0.32*	-0.11	-0.27 [†]	-0.14	0.12	-0.13	0.18	0.08	0.35*	1	
13. Concept maps	0.12	0.05	0.45**	-0.14	-0.01	-0.05	0.12	0.09	0.12	0.02	0.09	0.12	1

Note. [†] $p < .100$, * $p < .050$, ** $p < .010$, *** $p < .001$

Table 5 Hypotheses and Whether Findings Support Them

Hypothesis	Supported
H1a First reading positively associated with learning outcomes	No
H1b Repeat negatively associated with learning outcomes	Partial support
H1c No specific hypothesis for processing	
H2 High cognitive activities positively associated with connected concepts	Partial support
H3 Metacognitive activities positively associated with deep knowledge	Partial support

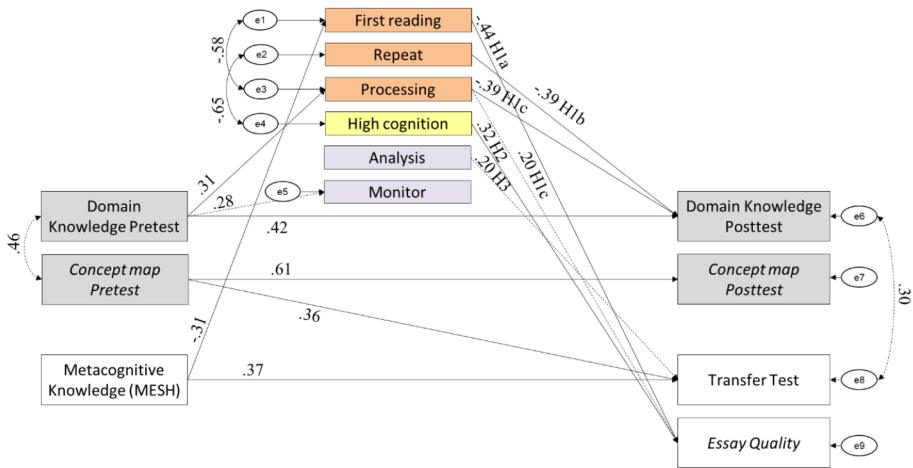


Fig. 4 Model of Metacognitive Knowledge, Knowledge Measures at Pretest and Posttest, and SRL Activities. (Note. Model fit was acceptable: $\chi^2(61)=74.26, p=.119, CFI=0.92, TLI=0.90, RMSEA=0.069$ with 90% CI [$<0.001 - 0.118$], SRMR=0.148. Surface knowledge has a grey background and deep knowledge a white background. Connected concepts are in italics and independent concepts are not. Orange represents low cognitive activities and purple metacognitive activities. Dotted lines represent paths with a p -value larger than 0.05. See Appendix A for all path coefficients.)

first reading. In addition to the hypothesised associations, we found autoregressive effects of domain knowledge and concept maps. Furthermore, domain knowledge and concept maps were positively related at pretest, and domain knowledge and transfer test were positively related at posttest as revealed by correlated errors.

Finally, mediation effects were found. Domain knowledge (pretest) was associated via *processing* with domain knowledge (posttest), $p=.015$. This means that higher scores on domain knowledge (pretest) were associated with higher scores on *processing*, and higher scores on *processing* were associated with lower scores on domain knowledge (posttest). Metacognitive knowledge (pretest) was associated via *first reading* with essay quality, $p=.048$. This means that higher scores on metacognitive knowledge (pretest) were associated with lower scores on *first reading*, and higher scores on *first reading* were associated with lower scores on essay quality (posttest).

Discussion

The present study investigated the association between the frequency of SRL activities during learning (low and high cognitive, and metacognitive activities) and learning outcomes (deep vs. surface knowledge, and independent vs. connected concepts). We want to stress the exploratory nature of our study and therefore, results should be interpreted with caution. Our hypotheses were partially confirmed. Low cognitive activities were related to learning outcomes, but not all. High cognitive activities were associated with one of two connected concepts measures: essay quality. Metacognitive activities were associated with one of two measures of deep knowledge, namely, with transfer. Note that we used frequencies of SRL activities in our analyses, which limits drawing conclusions about sequential or temporal aspects, but it does allow conclusions about the overall occurrence of specific learning activities in relation to learning outcomes.

Low cognitive activities

Low cognitive activities were assumed to be associated with all learning measures, with positive associations of reading and negative associations of repeat, while we explored the associations of processing. Reading was not positively associated with learning outcomes, rather there was a negative association. Repeat was negatively associated with domain knowledge. Processing showed a negative association with domain knowledge and a positive association with essay quality. Thus, we found partial support for our first hypothesis.

Regarding domain knowledge, the effect of repeat is in line with previous findings (Banert et al., 2014; Moos & Azevedo, 2008). The explanation is that when learners have difficulties understanding, they tend to repeat the information. Hence, repeat actions can be an indicator of problems with creating an initial knowledge base. In addition, repeat might also indicate that learners successfully identified low comprehension via metacognitive monitoring (Kim, 2017), but failed to remedy it, because they did not have an effective cognitive approach available to replace the repeating approach. With respect to processing, the negative association with domain knowledge is harder to explain. When learners interacted with their essay on a low level (rereading or copying information without elaboration or translation of information), we coded it as processing. This means it is impossible to engage in knowledge construction when engaging in processing, which can explain a negative relation between processing and domain knowledge at posttest. In the present study, we used a timed task commonly used in SRL research (e.g., Moos & Azevedo 2008; Deekens et al., 2018).

In line with this reasoning, we found that essay quality was positively associated with processing. Processing reflects interacting, and thus also creating one's products, such as the essay. It has been shown that time spent on the essay positively relates to essay quality (Guo et al., 2018). This might explain our result: As interacting with the essay likely leads to adding more elements in the essay and/or revising the essay, the quality of the essay improves as well. With reference to first reading, we found a negative association with essay quality. This might be explained in a similar way: When engaging in reading, it is not possible to engage in writing, leading to a negative association between reading and essay quality. Furthermore, increased time pressure has been associated with less effective cognitive activities (e.g., Sidi, et al., 2017). In our study, some reading appeared to be needed to write an essay, but prolonged reading seemed to be a less effective cognitive activity.

In sum, the current study found that low cognition appeared to be relevant for two out of four learning outcomes. However, we expected that low cognitive activities were associated with all learning outcomes, because they may lead to an emerging understanding of the topics (Frey et al., 2017). Our results did not show an association between low cognition and concept maps or transfer. For both learning outcomes, low cognition still may have been relevant. For instance, first reading had a high frequency, suggesting that most students did read. Therefore, reading still may have been important, but we could not assess the effect due to the overall relatively high frequency of reading. Another explanation is that low cognition was not needed to score better on concept maps and transfer. There was a medium-sized learning gain on the concept maps, and we found a large-sized effect on domain knowledge. Thus, the learning gain in concept maps could have been larger, and if this would have been the case, the influence of SRL activities could have been larger. This reasoning can also be applied to domain knowledge. Perhaps learning gains can be larger in other learning settings, such as a whole course instead of a single session. The transfer test assessed far transfer. Therefore, information from the text could not directly be translated to the transfer test (Lin & Lehman, 1999). Our results showed that far transfer was more strongly associated with controlled processing of the information via the metacognitive act of analysis. Thus, although a knowledge base might be needed, which can be created via reading, we found the association of transfer with metacognitive activities to be more prominent. This explanation aligns with the complex interplay of SRL activities and the information acted upon (Winne, 2018b).

High cognitive activities

High cognitive activities were positively associated with essay quality (connected concepts/deep knowledge). This result is in line with a previous study that revealed the influence of high cognition on essay quality in individual university students (Deekens et al., 2018) and groups of primary school students (Molenaar & Chiu, 2017). High cognition serves to (re-)organise and (re-)structure information (King, 2002). These activities are relevant to creating an essay. A knowledge base should be provided in an essay, but inferences should also be made. In our case, students were asked to write their vision for the future of education.

Thus, high cognition appeared to be relevant for one out of two measures of structured knowledge. We did not find an association of high cognitive activities with concept maps. This might have been due to the small learning gain in concept maps. Alternatively, an explanation might be that the participants were requested to create the concept maps themselves based on the texts they read, but without these texts being available during concept mapping, and without additional instruction on how to use the text to create concept maps. Thus, the participants might not have been adequately equipped to create concept maps. Studies that have revealed beneficial effects of creating concept maps on learning also include instruction on creating concept maps (Schroeder et al., 2018).

Metacognitive activities

Metacognitive activities were positively associated with transfer (independent concepts/deep knowledge), a positive effect of analysis. In addition, we found a positive effect of metacognitive knowledge on transfer, and an indirect effect of metacognitive knowledge via

first reading on essay quality. Taken together, these results confirm our hypothesis that metacognitive activities help in constructing deep knowledge. The association of metacognitive activities with transfer (Bannert, 2006; Bannert et al., 2009; Lin & Lehman, 1999) and essay quality (Greene & Azevedo, 2009) has been found before. This indicates that metacognitive activities help in creating deep knowledge. It is possible that students use orientation and planning to identify knowledge gaps or confusion and plan activities to resolve them (Winne, 2020), which would help translate incoming knowledge into existing knowledge, creating deep knowledge.

Metacognitive knowledge

Metacognitive knowledge was indirectly and positively related to essay quality via first reading. Higher scores on metacognitive knowledge at pretest were associated with lower frequencies of first reading, and lower frequencies of first reading were associated with higher essay scores. Metacognitive knowledge has been associated with academic achievement (Neuenhaus et al., 2011), which agrees with our finding that metacognitive knowledge is positively associated with essay quality. Our results add that the effect of metacognitive knowledge on learning outcomes is mediated by the learning process, more specifically by first reading in our study. Thus, by affecting frequencies of learning behaviour, metacognitive knowledge seems to foster construction of deep knowledge, as reflected in the essay. This finding aligns well with another finding, namely the direct association of metacognitive knowledge with our other measure of deep knowledge: the transfer test.

Prior knowledge

Domain knowledge on the pretest was positively associated with processing (low cognition) and monitor (metacognition). There were no other effects of prior knowledge on SRL. The finding of prior domain knowledge in relation to monitoring aligns with a previous study in which students with higher prior domain knowledge planned and monitored more, and took notes, summarised, and memorised less (Moos & Azevedo, 2008). The effect of monitoring can be explained by monitoring of comprehension, such as identifying that the information has been studied before, or by being able to monitor the relevance of the content (Moos & Azevedo, 2008). In a different study, students with higher prior knowledge moved their eye gaze more frequently from the text content area to the note-taking area than students with lower prior knowledge (Taub & Azevedo, 2019), which might be related to our processing measure. Thus, prior knowledge might aid in using text to create products, such as notes or essays.

Regarding the effects of prior knowledge on posttest scores, we found autoregression as expected between pretest and posttest scores of domain knowledge and concept maps. We also found an association between pretest concept maps and transfer. This suggests that having an organised knowledge structure of the relevant topics helps transfer this knowledge to a different context. Transferring knowledge to a contextually dissimilar problem was found to be fostered by a reason-justification treatment, which appeared to help students highlight underlying structures and principles (Lin & Lehman, 1999). This result also corresponds to the debate about the extent to which deep knowledge is organised (connected) and to which organised (connected) knowledge is deep (de Jong & Ferguson-Hessler, 1996). This

study adds that organised and deep knowledge measured at the same point in time are not associated (see posttest), and that organised knowledge before a learning session about the respective topics is associated with a measure of deep knowledge after the learning session. It can, therefore, be speculated that connectedness is a prerequisite for deep knowledge. However, more research is needed to disentangle the types of knowledge and study its interplay during learning, especially considering there was no effect of concepts maps at pretest (surface knowledge of independent concepts) to essay quality (deep knowledge of connected concepts).

Finally, we found two mediation effects. Metacognitive knowledge was indirectly related to essay quality via first reading. This result was explained by the use of metacognitive knowledge to create deep knowledge, as reflected in the essay. The second mediation effect was of domain knowledge at pretest via processing to domain knowledge at posttest. In other words, learners with higher domain knowledge at pretest tended to show more processing and learners with a high amount of processing tended to have lower domain knowledge scores at posttest. Learners with high domain knowledge scores at pretest had a knowledge base to be able to translate their knowledge to other contexts (Frey et al., 2017), in our case, into an essay by means of processing. In contrast, learners with low or average domain knowledge may have engaged less in processing, because they were unable to apply their knowledge when writing the essay. In turn, engagement with processing means not being engaged with other activities, which may have been more beneficial for acquiring domain knowledge.

Underlying mechanisms of SRL in relation to learning outcomes

Overall, our results suggested a trade-off between different SRL activities. A student could perform only one activity at one time, thus, excluding other activities. How students perceived and interpreted the learning goal may have affected which SRL activities were enacted and, in turn, what has been learnt. We found that processing was positively associated with essay quality, but negatively with domain knowledge. First reading was negatively associated with essay quality. These results can be explained by students' investigation of the costs and gains of possible choices, as in the expectancy-value theory (Eccles & Wigfield, 2002), which has been proposed to play an important role in SRL (Winne, 1995). The costs and gains analysis in the present study concerned different learning goals (gains): understanding the fundamentals of AI, scaffolding, and differentiation versus creating a vision about the future of education. The learner could control their learning via their SRL activities (costs): reading for domain knowledge versus processing to create an essay. Furthermore, this process of controlling learning appeared to be, at least partly, informed by metacognitive knowledge, because we found an indirect effect of metacognitive knowledge via first reading on essay quality. The main goal of the present learning assignment was to write an essay. The mediation effect shows that learners with higher metacognitive knowledge were more likely to read less, indicating a beneficial control strategy (Winne, 1996), because they tended to have higher essay scores, which might be due to having more time to write the essay. This effect of metacognitive knowledge, together with the effect of prior knowledge, on essay scores support the theoretical notion of learners' requiring sufficient resources to successfully engage in SRL (Seufert, 2019). The role of prior knowledge has already been

supported by previous research (e.g., Moos & Azevedo 2008). We add that metacognitive knowledge is also a learner characteristic related to successful SRL.

Limitations and Suggestions

The current study had a relatively small sample size. Given our sample size, the number of variables in the model is high. The present analysis should, therefore, be regarded as exploratory. We consider our results to be meaningful due to three reasons. First, we have a relatively high number of data samples for the SRL codes. Second, we cautiously built the model. Third, we had specific hypotheses about the relations, which were mostly confirmed. Future studies should further investigate the proposed association between SRL activities and learning outcomes. Nevertheless, our exploration was grounded in theory, we carefully tested our hypotheses, and we did find evidence to support our hypotheses. Reliability could have been better in a similar vein, although it was still acceptable. Nevertheless, these factors call for caution when interpreting the results.

Another limitation is that we did not take into account the information that was the object of SRL activities. It might have been that students incorrectly comprehended information or wrongly assumed that they understood it. To code such qualities of think aloud data, coding schemes should be extended and adapted to specific contexts, because the information provided affects the potential interactions with it. This brings about another limitation, namely the context of the present study. Although we assume that SRL activities can be relevant for learning regardless of the domain, it has also been shown that SRL interacts with the task's context (Winne, 2018a). One important part of SRL is that the learner takes the task characteristics into account during learning, for example, when planning their activities to be performed during the task. In our scenario, learners were limited in their time and had to choose what activities to perform. Such a decision can affect learning and learning outcomes (Winne, 1995), as we found a trade-off between reading to build a knowledge base and processing to write the essay. In addition, temporal aspects of SRL can be taken into account (Järvelä & Bannert, 2021; Molenaar & Järvelä, 2014). It might be that SRL frequencies vary over time, as has been found previously: More successful students display an equal number of cognitive acts as less successful students in the middle part of a learning session, but less cognitive acts in the first and final part of a learning session (Paans et al., 2019). Such variation might, then, be associated with different learning outcomes. Finally, using a pretest might have affected learners' engagement during the study. However, all participants completed the pretest, and the learners' task was to write an essay given the information provided. Therefore, we expected little to no effect of cueing the students.

Conclusions

Our findings show that frequencies of SRL during learning were associated with learning outcomes, that these associations depend on the particular SRL activity and learning outcome, and that a learner's resources were associated with SRL and learning outcomes. Low cognition might have fostered the construction of knowledge assessed in most learning outcomes. In contrast, high cognition might have been more helpful in creating more organised knowledge structures, and metacognition might have helped construct deeper knowledge. Therefore, it is important to consider the connectedness and deepness of knowledge when

designing and evaluating a learning task, for example, developing tools that foster planning to create deeper knowledge. Furthermore, it is important for learners to adapt their SRL to meet the current goals, because a particular SRL activity is more beneficial for constructing one type of knowledge than another.

Appendix A

To	From	Standardised path coefficient	<i>z</i>	<i>p</i>
Essay	First reading	-0.44	3.46	<0.001
Essay	Processing	0.20	1.64	0.101
Essay	High cognition	0.32	2.21	0.027
Transfer	Analysis	0.20	1.63	0.103
Transfer	Metacognitive knowledge	0.37	3.11	0.001
Transfer	Concept maps (pretest)	0.36	3.43	<0.001
Domain knowledge (posttest)	Repeat	-0.39	2.70	0.007
Domain knowledge (posttest)	Domain knowledge (pretest)	0.42	4.04	<0.001
Domain knowledge (posttest)	Processing	-0.39	3.69	<0.001
Concept maps (posttest)	Concept maps (pretest)	0.61	3.24	0.001
First reading	Metacognitive knowledge	-0.31	2.51	0.012
Processing	Domain knowledge (pretest)	0.31	2.63	0.008
Monitor	Domain knowledge (pretest)	0.28	1.85	0.064
Concept maps (pretest)	Domain knowledge (pretest)	0.46	1.94	0.053
First reading	Processing	-0.58	2.60	0.009
High cognition	Repeat	-0.65	4.26	<0.001
Transfer	Domain knowledge (posttest)	0.30	1.93	0.054
Essay	Metacognitive knowledge via first reading ¹	0.14	1.98	0.048
Essay	Domain knowledge (pretest) via processing ¹	0.06	1.61	0.107
Domain knowledge (posttest)	Domain knowledge (pretest) via processing	-0.12	2.44	0.015
Domain knowledge (posttest)	Total effect of domain knowledge (pretest)	0.31	3.22	0.001

Note.¹Total effects were not calculated due to absence of direct effects

Acknowledgements This work is a result of the FLoRA research project funded by DFG (Germany), NWO (The Netherlands), and ESRC (United Kingdom) as part of the Open Research Area (ORA) call 5 BA20144/10-1, NWO464.18.104, ES/S015701/1.

Declarations

The authors declare that they have no conflict of interest. The authors have received approval from the ethical committee of our university to conduct the present study. The authors have requested and received participants' approval to collect, process, and publish the data anonymously.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence,

and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexander, P. A. (1992). Domain Knowledge: Evolving Themes and Emerging Concerns. *Educational Psychologist*, 27(1), 33–51. https://doi.org/10.1207/s15326985ep2701_4
- Azevedo, R., Taub, M., & Mudrick, N. V. (2018). Understanding and Reasoning About Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation With Advanced Learning Technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (Second edition). Routledge, Taylor & Francis Group
- Bannert, M. (2006). Effects of Reflection Prompts When Learning with Hypermedia. *Journal of Educational Computing Research*, 35(4), 359–375. <https://doi.org/10.2190/94V6-R58H-3367-G388>
- Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia. Erfassung, Beschreibung und Vermittlung wirksamer metakognitiver Lernstrategien und Regulationsaktivitäten*. Waxmann
- Bannert, M., Hildebrand, M., & Mengelkamp, C. (2009). Effects of a metacognitive support device in learning environments. *Computers in Human Behavior*, 25(4), 829–835. <https://doi.org/10.1016/j.chb.2008.07.002>
- Bannert, M., Pieger, E., & Sonnenberg, C. (2015). *MESH - Fragebogen zur Erfassung metakognitiver Strategien beim Lernen mit Hypermedien [Metacognitive Strategy Inventory for Learning with Hypermedia]*. München: TUM School of Education, Technische Universität
- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. <https://doi.org/10.1007/s11251-011-9167-4>
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>
- Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, 52, 293–306. <https://doi.org/10.1016/j.chb.2015.05.038>
- Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(3), 570–585. <https://doi.org/10.1037/0278-7393.33.3.570>
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Collins, A., & Kapur, M. (2014). Cognitive Apprenticeship. *The Cambridge Handbook of the Learning Sciences, Second Edition* (pp. 44–62). Cambridge University Press. <https://www.scholars.northwestern.edu/en/publications/scaffolding>
- Cook, L., & Mayer, R. (1988). Teaching Readers About the Structure of Scientific Text. *Journal of Educational Psychology*, 80(4), 448–456
- Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, 40(3), 291–309. <https://doi.org/10.1080/02619768.2017.1315399>
- de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2
- Deekens, V. M., Greene, J. A., & Lobczowski, N. G. (2018). Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology*, 88, 63–79. <https://doi.org/10.1111/bjep.12174>
- Deunk, M. I., Smale-Jacobse, A. E., de Boer, H., Doolaard, S., & Bosker, R. J. (2018). Effective differentiation Practices: A systematic review and meta-analysis of studies on the cognitive effects of differentiation practices in primary education. *Educational Research Review*, 24, 31–54. <https://doi.org/10.1016/j.edurev.2018.02.002>
- Dinsmore, D. L. (2017). Toward a Dynamic, Multidimensional Research Framework for Strategic Processing. *Educational Psychological Review*, 29, 235–268. <https://doi.org/10.1007/s10648-017-9407-5>

- Dinsmore, D. L., & Alexander, P. A. (2016). A Multidimensional Investigation of Deep-level and Surface-level Processing. *The Journal of Experimental Education*, 84, 213–244. <https://doi.org/10.1080/00220973.2014.979126>
- Eccles, J. S., & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Review of Psychology*, 53(1), 109–132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Efklides, A. (2008). Metacognition: Defining Its Facets and Levels of Functioning in Relation to Self-Regulation and Co-regulation. *Europea Psychologist*, 13, 277–287. <https://doi.org/10.1027/1016-9040.13.4.277>
- ELAN [Computer software] (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- Engelmann, K., & Bannert, M. (2019). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*, 101205. <https://doi.org/10.1016/j.learninstruc.2019.05.002>
- European Union, P. O. (2019, July 8). of the E. *Key competences for lifelong learning*. [Website]. Publications Office European Union. <http://op.europa.eu/nl/publication-detail/-/publication/297a33c8-af3-11e9-9d01-01aa75ed71a1>
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34, 906–911. <https://doi.org/10.1037/0003-066X.34.10.906>
- Frey, N., Fisher, D., & Hattie, J. (2017). Surface, Deep, and Transfer? Considering the Role of Content Literacy Instructional Strategies. *Journal of Adolescent & Adult Literacy*, 60(5), 567–575. <https://doi.org/10.1002/jaal.576>
- Glaser, R. (1991). The maturing of the relationship between the science of learning and cognition and educational practice. *Learning and Instruction*, 1(2), 129–144. [https://doi.org/10.1016/0959-4752\(91\)90023-2](https://doi.org/10.1016/0959-4752(91)90023-2)
- Greene, J. A., & Azevedo, R. (2007). Adolescents' Use of Self-Regulatory Processes and Their Relation to Qualitative Mental Model Shifts While Using Hypermedia. *Journal of Educational Computing Research*, 36(2), 125–148. <https://doi.org/10.2190/G7M1-2734-3JRR-8033>
- Greene, J. A., & Azevedo, R. (2009). A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, 34(1), 18–29. <https://doi.org/10.1016/j.cedpsych.2008.05.006>
- Greene, J. A., Costa, L., & Dellinger, K. (2011). Analysis of self-regulated learning processing using statistical models for count data. *Metacognition and Learning*, 6, 275–301. <https://doi.org/10.1007/s11409-011-9078-4>
- Guo, H., Deane, P. D., van Rijn, P. W., Zhang, M., & Bennett, R. E. (2018). Modeling Basic Writing Processes From Keystroke Logs. *Journal of Educational Measurement*, 55(2), 194–216. <https://doi.org/10.1111/jedm.12172>
- Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal of Educational Research Online*, 5(2), 28
- Händel, M., Lockl, K., Heydrich, J., Weinert, S., & Artelt, C. (2014). Assessment of metacognitive knowledge in students with special educational needs. *Metacognition and Learning*, 9(3), 333–352. <https://doi.org/10.1007/s11409-014-9119-x>
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical Mediation Analysis in the New Millennium. *Communication Monographs*, 76(4), 408–420. <https://doi.org/10.1080/03637750903310360>
- Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: What characterizes good and poor mappers? *Instructional Science*, 36(1), 53–73. <https://doi.org/10.1007/s11251-007-9022-9>
- Holle, H., & Rein, R. (2015). EasyDIAG: A tool for easy determination of interrater agreement. *Behavior Research Methods*, 47, 837–847. <https://doi.org/10.3758/s13428-014-0506-7>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Järvelä, S., & Bannert, M. (2021). Temporal and adaptive processes of regulated learning—What can multimodal data tell? *Learning and Instruction*, 72, 101268. <https://doi.org/10.1016/j.learninstruc.2019.101268>
- King, A. (2002). Structuring Peer Interaction to Promote High-Level Cognitive Processing. *Theory Into Practice*, 41(1), 33–39. https://doi.org/10.1207/s15430421tip4101_6
- Kim, Y. G. (2017). Why the Simple View of Reading Is Not Simplistic: Unpacking Component Skills of Reading Using a Direct and Indirect Effect Model of Reading (DIER). *Scientific Studies of Reading*, 21, 310–333. <https://doi.org/10.1080/10888438.2017.1291643>
- Kline, P. (2000). *The Handbook of Psychological Testing*. Psychology Press

- Lehmann, T., Pirnay-Dummer, P., & Schmidt-Borcherding, F. (2020). Fostering integrated mental models of different professional knowledge domains: Instructional approaches and model-based analyses. *Educational Technology Research and Development*, 68(3), 905–927. <https://doi.org/10.1007/s11423-019-09704-0>
- Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in a computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching*, 36(7), 837–858. [https://doi.org/10.1002/\(SICI\)1098-2736\(199909\)36:7<837::AID-TEA6>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1098-2736(199909)36:7<837::AID-TEA6>3.0.CO;2-U)
- McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist*, 56, 1–19. <https://doi.org/10.1080/00461520.2021.1872379>
- Molenaar, I., & Chiu, M. M. (2017). Effects of Sequences of Cognitions on Group Performance Over Time. *Small Group Research*, 48(2), 131–164. <https://doi.org/10.1177/1046496416689710>
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85. <https://doi.org/10.1007/s11409-014-9114-2>
- Molenaar, I., van Boxtel, C. A. M., & Slegers, P. J. C. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science*, 39(6), 785–803. <https://doi.org/10.1007/s11251-010-9154-1>
- Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology*, 33(2), 270–298. <https://doi.org/10.1016/j.cedpsych.2007.03.001>
- Munoz, S. R., & Bangdiwala, S. I. (1997). Interpretation of Kappa and B statistics measures of agreement. *Journal of Applied Statistics*, 24, 105–112. <https://doi.org/10.1080/02664769723918>
- Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2011). Fifth graders metacognitive knowledge: general or domain-specific? *European Journal of Psychology of Education*, 26, 163–178. <https://doi.org/10.1007/s10212-010-0040-7>
- Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2019). Temporal variation in children's self-regulated hypermedia learning. *Computers in Human Behavior*, 96, 246–258. <https://doi.org/10.1016/j.chb.2018.04.002>
- Pirnay-Dummer, P., Ifenthaler, D., & Spector, J. M. (2010). Highly integrated model assessment technology and tools. *Educational Technology Research and Development*, 58(1), 3–18. <https://doi.org/10.1007/s11423-009-9119-8>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Reif, F., & Heller, J. I. (1982). Knowledge structure and problem solving in physics. *Educational Psychologist*, 17(2), 102–127. <https://doi.org/10.1080/00461528209529248>
- Reiser, B. J., & Tabak, I. (2014). Scaffolding. *The Cambridge Handbook of the Learning Sciences, Second Edition* (pp. 44–62). Cambridge University Press. <https://www.scholars.northwestern.edu/en/publications/scaffolding>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), <https://doi.org/10.18637/jss.v048.i02>
- Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall
- Schroeder, N. L., Nesbit, J. C., Anguiano, C. J., & Adesope, O. O. (2018). Studying and Constructing Concept Maps: A Meta-Analysis. *Educational Psychology Review*, 30(2), 431–455. <https://doi.org/10.1007/s10648-017-9403-9>
- Schunk, D. H., & Greene, J. A. (Eds.). (2018). *Handbook of Self-Regulation of Learning and Performance* (Second edition). Routledge, Taylor & Francis Group
- Seufert, T. (2019). Training for coherence formation when learning from text and picture and the interplay with learners' prior knowledge. *Frontiers in psychology*, 10, 1–11. <https://doi.org/10.3389/fpsyg.2019.00193>
- Sternberg, R. J. (1981). Testing and cognitive psychology. *American Psychologist*, 36, 1181–1189. <https://doi.org/psycnet.apa.org/doi/10.1037/0003-066X.36.10.1181>
- Taub, M., & Azevedo, R. (2019). How Does Prior Knowledge Influence Eye Fixations and Sequences of Learning and Metacognitive SRL Processes during Learning with an Intelligent Tutoring System? *International Journal of Artificial Intelligence in Education*, 29(1), 1–28. <https://doi.org/10.1007/s40593-018-0165-4>
- Thurn, C. M., Hänger, B., & Kokkonen, T. (2020). Concept Mapping in Magnetism and Electrostatics: Core Concepts and Development over Time. *Education Sciences*, 10(5), 129. <https://doi.org/10.3390/educsci10050129>
- Van Wetering, M., Booij, E., & Van Bruggen, W. (2019). *Education in an artificially intelligent world Kennisnet Technology Compass 2019–2020*. Kennisnet. <https://www.kennisnet.nl/app/uploads/kennisnet/publicatie/Kennisnet-Technology-Compass-2019-2020.pdf>

- Veenman, M. V. J. (2013). Assessing Metacognitive Skills in Computerized Learning Environments. In R. Azevedo, & V. Alevan (Eds.), *International Handbook of Metacognition and Learning Technologies*. New York: Springer
- Veenman, M., & Elshout, J. J. (1999). Changes in the relation between cognitive and metacognitive skills during the acquisition of expertise. *European Journal of Psychology of Education*, *14*, 509–523. <https://doi.org/10.1007/BF03172976>
- Volet, S., Summers, M., & Thurman, J. (2009). High-level co-regulation in collaborative learning: How does it emerge and how is it sustained? *Learning and Instruction*, *19*, 128–143. <https://doi.org/10.1016/j.learninstruc.2008.03.001>
- Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, *30*(4), 173–187. https://doi.org/10.1207/s15326985ep3004_2
- Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, *8*(4), 327–353. [https://doi.org/10.1016/S1041-6080\(96\)90022-9](https://doi.org/10.1016/S1041-6080(96)90022-9)
- Winne, P. H. (2018a). Cognition and Metacognition Within Self-Regulated Learning. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (Second edition). Routledge, Taylor & Francis Group
- Winne, P. H. (2018b). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology*, *88*(1), 9–20. <https://doi.org/10.1111/bjep.12173>
- Winne, P. H. (2020). Open Learner Models Working in Symbiosis With Self-Regulating Learners: A Research Agenda. *International Journal of Artificial Intelligence in Education*. <https://doi.org/10.1007/s40593-020-00212-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Joep van der Graaf^{1,5} · Lyn Lim² · Yizhou Fan³ · Jonathan Kilgour³ ·
Johanna Moore³ · Dragan Gašević^{3,4} · Maria Bannert² · Inge Molenaar¹

✉ Joep van der Graaf
joep.vandergraaf@ru.nl

¹ Behavioural Science Institute, Radboud University, Thomas van Aquinostraat 4, Room 3.036, P.O. box 9104, 6525 Nijmegen, GD, The Netherlands

² Technical University of Munich, Munich, Germany

³ School of Informatics, University of Edinburgh, Edinburgh, UK

⁴ Centre for Learning Analytics, Faculty of Information Technology, Monash University, Melbourne, Australia

⁵ Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia