Evaluating the Impact of Instructional Support Using Data Mining and Process Mining: A Micro-Level Analysis of the Effectiveness of Metacognitive Prompts

Christoph Sonnenberg
University of Wuerzburg, Instructional Media, Germany
christoph.sonnenberg@uni-wuerzburg.de

Maria Bannert
Technical University of Munich, Teaching and Learning with Digital Media, Germany
maria.bannert@tum.de

In computer-supported learning environments, the deployment of self-regulatory skills represents an essential prerequisite for successful learning. Metacognitive prompts are a promising type of instructional support to activate students' strategic learning activities. However, despite positive effects in previous studies, there are still a large number of students who do not benefit from provided support. Therefore, it may be necessary to consider explicitly the conditions under which a prompt is beneficial for a student, i.e., so-called adaptive scaffolding. The current study aims to (i) classify the effectiveness of prompts on regulatory behavior, (ii) investigate the correspondence of the classification with learning outcome, and (iii) discover the conditions under which prompts induce regulatory activities (i.e., the proper temporal positioning of prompts). The think-aloud data of an experiment in which metacognitive prompts supported the experimental group (n = 35) was used to distinguish between effective and non-effective prompts. Students' activities preceding the prompt presentation were analyzed using data mining and process mining techniques. The results indicate that approximately half of the presented prompts induced metacognitive learning activities as expected. Moreover, the number of induced monitoring activities correlates positively with transfer performance. Finally, the occurrence of orientation and monitoring activities, which are not well-embedded in the course of learning, increases the effectiveness of a presented prompt. In general, our findings demonstrate the benefits of investigating metacognitive support using process data, which can provide implications for the design of effective instructional support.

Keywords: self-regulated learning, instructional support, micro-level analysis, metacognitive prompting, think-aloud data, process mining

1. Introduction

The research in self-regulated learning (SRL) indicates that many learners have difficulties in spontaneously deploying regulatory activities, which results in lower learning performance (e.g., Azevedo, 2009; Bannert & Mengelkamp, 2013; Greene, Dellinger, Tüysüzoglu, & Costa, 2013;

Winne & Hadwin, 2008; Zimmerman, 2008). Therefore, our main objective is to provide effective instructional support for hypermedia learning. In this context, metacognitive prompting is a promising approach that affects the learning process by inducing regulatory activities as well as learning performance (e.g., Bannert & Mengelkamp, 2013; Bannert & Reimann, 2012). The purpose of these prompts is to foster SRL activities such as orientation, planning, monitoring, and evaluation strategies by asking students to monitor and to control their learning process (Bannert, 2009; Veenman, 1993). However, post-hoc analyses of students' prompt use revealed that approximately half of the sample demonstrated poor compliance with the provided support (Bannert & Mengelkamp, 2013). Thus the students did not benefit from the metacognitive prompting as intended. To improve the provided instructional support, it is our aim to investigate the conditions that influence the effectiveness of metacognitive prompts by taking into account fine-grained process data.

Referring to an event-based view of SRL that describes regulatory activities as dynamically unfolding over time during a learning task (e.g., Azevedo, 2009; Winne, 2014), the current research is increasingly interested in analyzing sequences of learning activities that are measured online during learning (e.g., measured by log files or think-aloud data). Moreover, the development and application of new methods that can take into account the sequential and temporal order of learning events (Martin & Sherin, 2013; Molenaar & Järvelä, 2014), as well as the dynamic relationship between SRL processes (Ben-Eliyahu & Bernacki, 2015), accompany the increasing importance of analyzing process data. In particular, the techniques in the field of educational data mining (EDM) have the potential to support the discovery of event patterns in SRL (Winne & Baker, 2013), for example, by modeling events that are crucial for an understanding of learning through the use of process mining (Reimann & Yacef, 2013; Trčka, Pechenizkiy, & van der Aalst, 2010). As a consequence, these recent developments provide new opportunities for the evaluation of instructional support on the micro level. More precisely, an in-depth process analysis contributes to the investigation of scaffolding effects (e.g., metacognitive prompting) on learning activities, and the evaluation results can inform researchers about how to develop their supporting strategies (e.g., Jeong et al., 2008; Johnson, Azevedo, & D'Mello, 2011; Molenaar & Chiu, 2014; Sonnenberg & Bannert, 2015). In general, we argue that the analysis of fine-grained process data is necessary for the development and the design of effective instructional support (e.g., Bannert & Mengelkamp, 2013; Sonnenberg & Bannert, 2015).

Hence, the aim of the current contribution is to demonstrate the potential of evaluating metacognitive prompting by analyzing process data (i.e., concurrent think-aloud protocols) by using data mining and process mining techniques. Because metacognitive prompts do not always optimally support students (e.g., Bannert & Mengelkamp, 2013; Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015), we investigate the conditions of effectiveness to derive implications for an improved prompt design. According to the concept of adaptive hypermedia systems and adaptive scaffolding (Bouchet, Harley, Trevors, & Azevedo, 2013; Brusilovsky, 2001, 2007; Molenaar & Roda, 2008), the system and the provided support should be tailored to the student's requirements, for example, to prior knowledge and to performed SRL activities (Azevedo, Cromley, & Seibert, 2004; Azevedo, Cromley, Winters, Moos, & Greene, 2005). Moreover, Molenaar and Roda (2008) note that the evaluation of scaffolds needs to be contingent on the learner's current activities and his or her goals. Therefore, it is crucial to examine the conditions that contribute to the effectiveness of a scaffold when its presentation successfully supports a student.

The present paper is organized as follows. First, we outline the effects of scaffolding hypermedia learning through metacognitive prompts. We focus on the challenges regarding prompt design and the concept of adaptive scaffolding, including the results of the related research. Second, we introduce the analysis of fine-grained process data for the evaluation of instructional support. Third, we investigate the effects of prompting using the coded think-aloud data of an experimental study. To that end, we first classify the prompts by considering the increase of metacognitive utterances following the prompt presentation. Then we explore the conditions of effectiveness applying data mining and process mining techniques to the learning activities that precede each prompt. Finally, the significance of the findings for the design of the prompts and the development of a micro-level theory of SRL processes are discussed.

2. Scaffolding Hypermedia Learning Through Instructional Support

Especially in open-ended learning settings, the use of self-regulatory skills represents a predictor of learning success (e.g., Azevedo, 2005; Lin, Hmelo, Kinzer, & Secules, 1999). However, students often do not perform regulatory activities spontaneously, which in general results in lower learning outcomes (e.g., Azevedo, 2009; Bannert & Mengelkamp, 2013). Consequently, instructional support aims to counteract this deficiency by promoting the activation of strategic learning processes. For example, the research on metacognitive prompting has provided evidence for its beneficial effects on learning process and outcome (e.g., Azevedo, Cromley, Moos, Greene, & Winters, 2011; Bannert, 2009; Bannert et al., 2015). In a series of experiments comparing students supported by different types of metacognitive prompts to a control group without support, we found medium effects on metacognitive processes and transfer performance (Bannert & Mengelkamp, 2013). This magnitude is in line with meta-analyses on metacognitive instruction (d = 0.59; Hattie, 2009). The research on metacognitive prompting is presented in more detail below.

2.1. Effects of Metacognitive Prompting

Metacognitive prompts have the purpose of inducing regulatory activities, for example, orientation, planning, monitoring, and evaluation strategies (Bannert, 2007, 2009; Veenman, 1993). According to theories of SRL (e.g., Winne & Hadwin, 2008; Zimmerman, 2008), the recurring deployment of these activities during a learning task is crucial for the successful regulation of one's learning process. For example, Zimmerman (2008) describes SRL as a cyclical model that comprises three phases, namely *forethought* (i.e., task analysis, goal setting, and planning), *performance* (i.e., strategy use, monitoring, and control), and *reflection* (i.e., self-evaluation). The current study refers to a framework for successful hypermedia learning (Bannert, 2007), which comprises the learning activities orientation, goal specification, planning, information search and relevance judgment, information processing, monitoring, and evaluation of goal attainment. Although these activities may imply a typical sequence, their performance can vary dynamically, considering the challenges of a given learning task. Additionally, the framework also considers the motivational aspects such as achievement motivation, action control, and self-efficacy, which are supposed to influence the strategy use in the process of learning. Especially students' motivational states with respect to the perception of their current task (e.g., the value or level of difficulty), their competencies (e.g., the ability to successfully use a specific strategy), and the learning situation (e.g., the authenticity of the situation) might affect their learning behavior.

For example, Zimmerman (2008) highlights the importance of motivational variables such as self-efficacy, task interest, value, and goal orientation, which influence the student's proactive pursuit towards his or her learning goals.

According to the expected effect mechanism, the presentation of metacognitive prompts fosters the activation of one's repertoire of metacognitive skills (e.g., Bannert & Mengelkamp, 2013). Metacognitive prompting thereby attempts to remedy the phenomenon of *production deficit* (e.g., Winne, 1996; Wirth, 2009). That is, students possess knowledge about regulatory capacities, but they do not use such skills spontaneously. Consequently, metacognitive prompts can foster the performance of learning activities that are described in the framework above and even the sequential order of their implementation during a learning task.

Several studies on metacognitive prompting have confirmed its positive effect on learning activities and learning outcome (e.g., Azevedo et al., 2011; Ge, 2013; Johnson et al., 2011; Kramarski & Gutman, 2006). However, the research also indicates that its effectiveness may depend on learner characteristics or the features of the prompt design (e.g., Bannert, 2009). For instance, the use of prompts requires additional cognitive capacities, which makes it necessary to have sufficient prior domain knowledge or to offer training in advance on how to use the prompts during learning (Bannert & Reimann, 2012; Veenman, 1993). Moreover, Thillmann, Künsting, Wirth, and Leutner (2009) found that strategy instruction must be embedded in the ongoing course of learning. Similarly, findings from cognitive support strengthen the necessity to integrate strategy instruction into ongoing cognitive learning activities (e.g., Wittwer & Renkl, 1998). However, a study on metacognitive feedback for help-seeking skills using an intelligent tutoring system showed that real-time feedback may be not beneficial to learning gains for all students (Aleven, Roll, McLaren, & Koedinger, 2016; Roll, Aleven, McLaren, & Koedinger, 2011). Possible explanations that were discussed by the authors are a cognitive overload that is caused by the real-time interventions or a lack of awareness of the value of the provided feedback (i.e., a motivational issue). As a result, more research is needed that investigates the conditions under which metacognitive support is beneficial.

In our previous experiments, we investigated the effects of different types of metacognitive prompts during hypermedia learning (Bannert & Mengelkamp, 2013; Bannert et al., 2015). The presentation of prompts took place during a 40-minute learning session, and the prompts stimulated or even suggested appropriate regulatory activities to university students. Overall, the results showed positive effects on learning processes, measured by coded think-aloud protocols ($0.35 < d < 1.17$) as well as on navigation behavior ($0.42 < d < 0.59$), and finally on transfer performance ($0.42 < d < 0.59$). With regard to the learning process, the students that were supported by prompts showed significantly more metacognitive learning activities and significantly better navigation behavior in comparison to the students in a control group who received no support. In addition, a detailed analysis that used process mining revealed differences among the process models of the students who learned with prompts and the students who learned without support (Sonnenberg & Bannert, 2015). These findings indicate that metacognitive prompting affects not only the frequency of regulatory activities but also the deployment of the sequences of learning activities.

In our most recent work (Bannert et al., 2015), students were asked to provide the reasons for their navigational decisions several times during learning (i.e., prompts to reflect on one's behavior). We found beneficial effects on navigation behavior (frequency of relevant pages visited: $p = .004$, $d = 0.65$; time spent on relevant pages: $p = .009$, $d = 0.58$) and on transfer performance ($p = .035$, $d = 0.44$) compared with those of a control group. The current study extends this work by analyzing think-aloud data to gain a deeper insight into the effectiveness of the presented prompts and to address the challenges of prompt design better.

2.2. Challenges of Prompt Design

Despite the reported beneficial effects in the previous section, an optimal prompt design in open-ended learning environments remains challenging (Azevedo & Hadwin, 2005). In our studies, a post-hoc analysis based on videos and verbal protocols showed that in general only half of the sample used the provided prompts in the intended manner (Bannert & Mengelkamp, 2013). For example, students started to read the contents of the learning environment immediately instead of first planning their learning steps as requested by prompts, or they even ignored the prompts by not considering the requests at all. Moreover, some students reported that they felt restricted in their course of learning by the prompts. Consequently, the important question arises as to why many students do not comply with a provided support device (Clarebout & Elen, 2006; Clarebout, Elen, Collazo, Lust, & Jiang, 2013). The recent research has noted that the temporal positioning of a prompt within the learning process can significantly influence compliance and thereby the effectivness of prompting (Azevedo et al., 2011; Sitzmann, Bell, Kraiger, & Kanar, 2009; Thillmann et al., 2009). As a consequence, more research that concerns the temporal characteristics of SRL is needed to inform the timing of scaffolds (Molenaar & Järvelä, 2014; Sonnenberg & Bannert, 2015).

In general, there are two possible approaches to improve the timing of a scaffold during hypermedia learning. First, researchers can try to involve students in designing their scaffolds. Because students should be the experts in their course of learning, it may be best to provide them with the opportunity to adapt the support to their needs (e.g., Bannert et al., 2015). Second, one can attempt to develop an adaptive hypermedia system that can diagnose the demands of a student and that can present a scaffold when it is needed based on specific embedded rules (e.g., Bouchet et al., 2013). Both are promising approaches for optimizing the positioning of metacognitive prompts during hypermedia learning, but they also present certain challenges, which are described below.

In a recent study (Bannert et al., 2015), we investigated the first approach, that is, the effects of so-called *self-directed metacognitive prompts*. The idea behind these prompts is to involve students in the configuration or even the creation of their prompts before learning, for example, by determining the time when a prompt should appear in their course of learning (e.g., after 5 minutes, after 12 minutes, and so on). Contrary to our expectations, despite the beneficial effects compared to a control group, there was no significant improvement of prompt use. Although the students had been familiarized with the hypermedia learning environment by performing some training tasks, it may be possible that they were overcharged with adapting the scaffolds to their needs. It is possible that they could not correctly determine when they should best be supported by a prompt.

Considering the second approach, the learning environment needs to diagnose the requirements of the learner simultaneously during learning, and it must determine when to position a scaffold, provided, for example, by a pedagogical agent (Bouchet et al., 2013; see the next section on adaptive support). This approach has the advantage of taking into account the current learning progress for presenting adaptive support, but both diagnosis and intervention can be difficult to develop and implement into a learning environment (Azevedo & Hadwin, 2005). Therefore, more research is needed that analyses the data of students' learning activities and how scaffolds affect these activities to derive the conditions under which a scaffold can optimally support a learner. Moreover, these analyses can contribute to the development of more specific models that describe how to support self-regulated behavior and to develop rules for the implementation of adaptive scaffolds into learning environments. For example, which learning activities should trigger a support device or the appropriate timing to present a scaffold. As described in the following, related studies have already addressed the impact of adaptive support on the enhancement of computer-supported learning.

2.3. Effects of Adaptive Support

In general, the research that has investigated adaptive scaffolding has confirmed the beneficial effects of individualized support on the regulation of learning processes and on learning outcome (e.g., Azevedo et al., 2004, 2005, 2011; Lehmann, Hähnlein, & Ifenthaler, 2014; Schwonke, Hauser, Nückles, & Renkl, 2006; Yeh, Chen, Hung, & Hwang, 2010). For example, Azevedo et al. (2005) compared the effects of adaptive scaffolding, fixed scaffolding, and no scaffolding during hypermedia learning. The results showed that adaptive scaffolding, which was provided by a human tutor, facilitated the shift in students' mental models significantly compared to the other two groups, and, further, it improved the deployment of regulatory strategies. Moreover, Lehmann et al. (2014) investigated the effectiveness of so-called preflective and reflective prompts. They found benefits in the preflective prompts, which stimulate to reflect on future events, but only for novice learners. Hence, they concluded that the adaptation of prompting in online research is crucial for future research.

Moreover, computerized settings such as adaptive hypermedia systems (Brusilovsky, 2001, 2007) provide new possibilities for the realization of adaptive support (e.g., Molenaar & Roda, 2008; Walker, Rummel, & Koedinger, 2011). Students' interactions with the system can be analyzed automatically and in real-time, thus informing the intervention type and presentation time of support devices. The current research investigating SRL and metacognitive behaviors uses some intelligent tutoring systems and learning environments such as BioWorld (Lajoie et al., 2013), the Geometry Cognitive Tutor (Aleven, 2013), and Crystal Island (Lester, Mott, Robison, Rowe, & Shores, 2013). These systems incorporate approaches to assess and scaffold SRL-behavior dynamically. A specific example of an existing adaptive hypermedia learning environment is MetaTutor, a system that provides scaffolds and feedback through several pedagogical agents (e.g., Azevedo et al., 2012; Bouchet et al., 2013). Its adaptive presentation is based on a set of system-generated rules that refer to the students' interaction with the system. For example, when a learner begins with a new sub-goal, a pedagogical agent prompts the learner to activate any prior knowledge that may be relevant to the sub-goal before starting to work on the content. Furthermore, a pedagogical agent asks the learners if they have adequately completed the current subgoal after they have spent more than 20 minutes working on it. In summary, MetaTutor uses the students' interactions with the hypermedia system as conditions for triggering the specific actions of pedagogical agents. Based on the analyses of learning activities with MetaTutor and the impact of pedagogical agents on learning processes and learning outcome, the system rules are constantly refined to optimize the provided support. Another example for an adaptive scaffolding environment is the AtgentSchool system (Molenaar & Roda, 2008), which was designed to provide dynamic and adaptive support to school children. In general, AtgentSchool is an attention management system that can diagnose a learner's focus using his or her activities within the system and provide appropriate interventions. The system works adaptively and dynamically by calibrating the scaffolds to the student's progress and his or her characteristics.

Despite the beneficial impact of adaptive scaffolds and the possible implementation of adaptive rules in hypermedia systems, there are still many issues for future research. For example, Yeh et al. (2010) noted the missing theoretical understanding of prompt formats that are tailored to students' different levels of expertise. Moreover, the questions of how to diagnose the proper time to scaffold, how to calibrate the support for the appropriate phase of SRL, and how to gradually reduce support as students progress in self-regulating their learning (e.g., Azevedo & Hadwin, 2005) still need to be addressed in future studies. In the current contribution, we concentrate on the investigation of the appropriate temporal positioning of scaffolds by exploring the conditions of effective scaffolds using process data.

3. Evaluation of Instructional Support Using Process Data

The recent research in SRL that emphasizes the investigation of learning as patterns of events (e.g., Azevedo, 2014; Bannert, Reimann, & Sonnenberg, 2014; Winne, 2014) has begun to concentrate on the microanalysis of process data, which is comprised of different traces of students' behavior during learning (e.g., log files, think-aloud data, or eye movements). The development and application of innovative analysis techniques that are appropriate for this type of data accompany the increasing interest in fine-grained process data. For example, the techniques that address the temporal characteristics and the dynamic relations of regulation activities (Ben-Eliyahu & Bernacki, 2015; Molenaar & Järvelä, 2014) or that support the discovery of event patterns in SRL activities (Winne & Baker, 2013). These approaches provide new potential for the evaluation of the effectiveness of scaffolds on the micro level. Furthermore, evaluation results can contribute to the advancement of a supporting strategy (e.g., Jeong et al., 2008; Johnson et al., 2011; Molenaar & Chiu, 2014; Sonnenberg & Bannert, 2015).

The analysis of process data using EDM techniques can stimulate the development and design of effective instructional support that is based on discovered process patterns. For example, Bouchet et al. (2013) investigated how adaptive versions of their learning environment respond differently to students, clustered by their SRL behavior. The authors drew implications for the implementation of increasingly adaptive, individualized support based on learners' profiles. Moreover, Kinnebrew, Segedy, and Biswas (2014) were able to track students' cognitive skills as well as their use of metacognitive skills in learning environments using log files and sequence mining methods. Additionally, the evaluation of scaffolds using process data can support the development of SRL micro-level models that comprise assumptions on the conditions of effective scaffolding (e.g., positioning, students' level of expertise, type of scaffold).

In our approach, we apply process mining techniques (PM; Reimann & Yacef, 2013; Trčka et al., 2010) to model the sequences of events that are crucial for understanding learning processes as well as the impact of the provided scaffolds on these processes. Those learning activities are measured by concurrent think-aloud protocols (Ericsson & Simon, 1993). In general, PM enables the discovery of process models from event sequences that are stored in an event log, the testing of models through conformance checking with additional data, and the extension of existing models (Trčka et al., 2010). We recommend PM as a promising method in SRL research (Bannert et al., 2014; Sonnenberg & Bannert, 2015) because it allows researchers to describe and to test learning models that incorporate a process-oriented view, and that can represent the workflow of activities (Van der Aalst, Weijters, & Maruster, 2004). Theses process models build on the concept of Petri nets, which represent an executable system of places and transitions (Bannert et al., 2014). Using PM algorithms for discovery, a model can be generated based on event data. The functionality of a discovery algorithm is described in more detail in the section titled 'Analysis techniques'. In the context of computer-supported learning research in particular, PM techniques are increasingly used to study learning from an event-based perspective (e.g., Malmberg, Järvelä, Järvenoja, & Panadero, 2015; Reimann, Markauskaite, & Bannert, 2014; Reimann & Yacef, 2013; Schoor & Bannert, 2012).

In previous analyses that have used PM, we compared the process models of students with high versus low learning performance and demonstrated that PM techniques can reveal differences in the sequential patterns of regulatory activities (Bannert et al., 2014). Furthermore, we investigated the effects of metacognitive prompts on the sequential structure of SRL activities by comparing the process models of students supported by prompts and of students in a control group without support (Sonnenberg & Bannert, 2015). Compared to the traditional frequency-based analysis of learning events, which is not able to take into account the sequential order of activities, additional findings on prompting effects were revealed. Now, we aim to obtain more detailed information on the learning process that precedes the presentation of a prompt to derive the conditions for its effectiveness. For example, discovered patterns could indicate that the performance of certain events or sequences of events is beneficial, or detrimental, for the effectiveness of a subsequent prompt presentation. Again, it is expected that PM can contribute to deeper insights into the sequence of learning activities.

4. Research Questions

Metacognitive prompts can stimulate the activation of one's repertoire of strategic learning activities. Despite the beneficial effects of prompting on learning process and learning outcome, students often do not comply with the provided support optimally (Bannert & Mengelkamp, 2013). Poor compliance is most probably caused by a lack of tailoring of instructional support, that is, the conditions under which a scaffold is needed are often not considered. For instance, the presentation of a prompt may not be necessary, and it can even be disruptive at certain times in the learning process. Therefore, we argue that it is necessary to consider students' current learning process to provide adequate instructional support. The conditions of the learner and his or her learning progress (e.g., learner characteristics, learning material, or current learning activity) need to be analyzed by process analysis to inform the decisions of when and how often a prompt should be presented, that is, the amount of scaffolding, the timing, and the fading out of support. In the present analysis, we aim to classify the effectiveness of metacognitive prompts by considering their impact on subsequent learning activities. Additionally, we focus on the conditions of effectiveness by analyzing the learning activities that precede each prompt. We address the following research questions in detail:

1. Is it possible to distinguish between metacognitive prompts with high and low effectiveness in terms of the activation of regulatory learning activities using process data?

2. Does the effectiveness of prompts correspond with learning outcome; that is, does a student who activates a higher number of metacognitive learning activities following a prompt presentation show a higher learning performance?

3. What are the conditions under which metacognitive prompts effectively induce regulatory activities?

With respect to the effect mechanism of metacognitive prompts, we expect that the presentation of an effective prompt enhances the deployment of metacognitive learning activities. Moreover, the number of induced metacognitive activities should be associated with the learning outcome. Finally, it should be possible to discover the conditions, that is, the learning activities that precede the prompt presentation that influence its effectiveness.

5. METHOD

The present analysis relates to the data of an experimental study that was reported in Bannert et al. (2015), which examined the general effects of metacognitive prompts on navigation parameters and learning performance. In the following analysis, we proceed by concentrating on a microanalysis using think-aloud data. Both contributions refer to the same participants, but they investigate different research questions and mainly consider different data sources.

5.1. SAMPLE AND RESEARCH DESIGN

The participants were N = 70 undergraduate students from a German university (mean age = 20.07, SD = 1.88, 82.9% female). The students were randomly assigned to the experimental group (n = 35, mean age = 20.29, SD = 1.89, 88.6% female), or to the control group (n = 35, mean age = 19.86, SD = 1.87, 77.1% female). All of the participants majored either in media communication or human-computer interaction, and their recruitment was accomplished through an online system that was administered by our institute. Each student received an incentive of 40 Euros (approximately $47 USD) for their participation.

All of the students participated in a hypermedia learning session. The students who were assigned to the experimental group received support through metacognitive prompts, whereas the control group received no support during learning. Due to the research questions of the current study, we will only focus on the data of the experimental group. More detailed information about the procedure and the findings of the group comparison is reported in Bannert et al. (2015).

5.2. LEARNING ENVIRONMENT AND PERFORMANCE MEASUREMENT

Both the learning material and metacognitive prompts were presented in a hypermedia learning environment. The learning content comprised a chapter on learning theories (i.e., classical conditioning, operant conditioning, and observational learning). Altogether, this chapter included 50 nodes with approximately 13,000 words, 20 pictures and tables, and 300 hyperlinks. The pages that were relevant to the learning goals were limited to 10 nodes with approximately 2,300 words, five pictures and tables, and 60 hyperlinks. Thus, each node comprised approximately 230 words, and there was a figure on approximately every second page. All of the remaining nodes comprised overviews, summaries, and pages with content that were not relevant to the learning task.

It was possible to navigate within the learning environment using one of the following elements: (i) a hierarchical navigation menu, (ii) a next-page and previous-page button, (iii) the backward- and forward-button of the browser, and (iv) hyperlinks that were embedded in the content. Support through metacognitive prompts appeared as a pop-up window on the screen

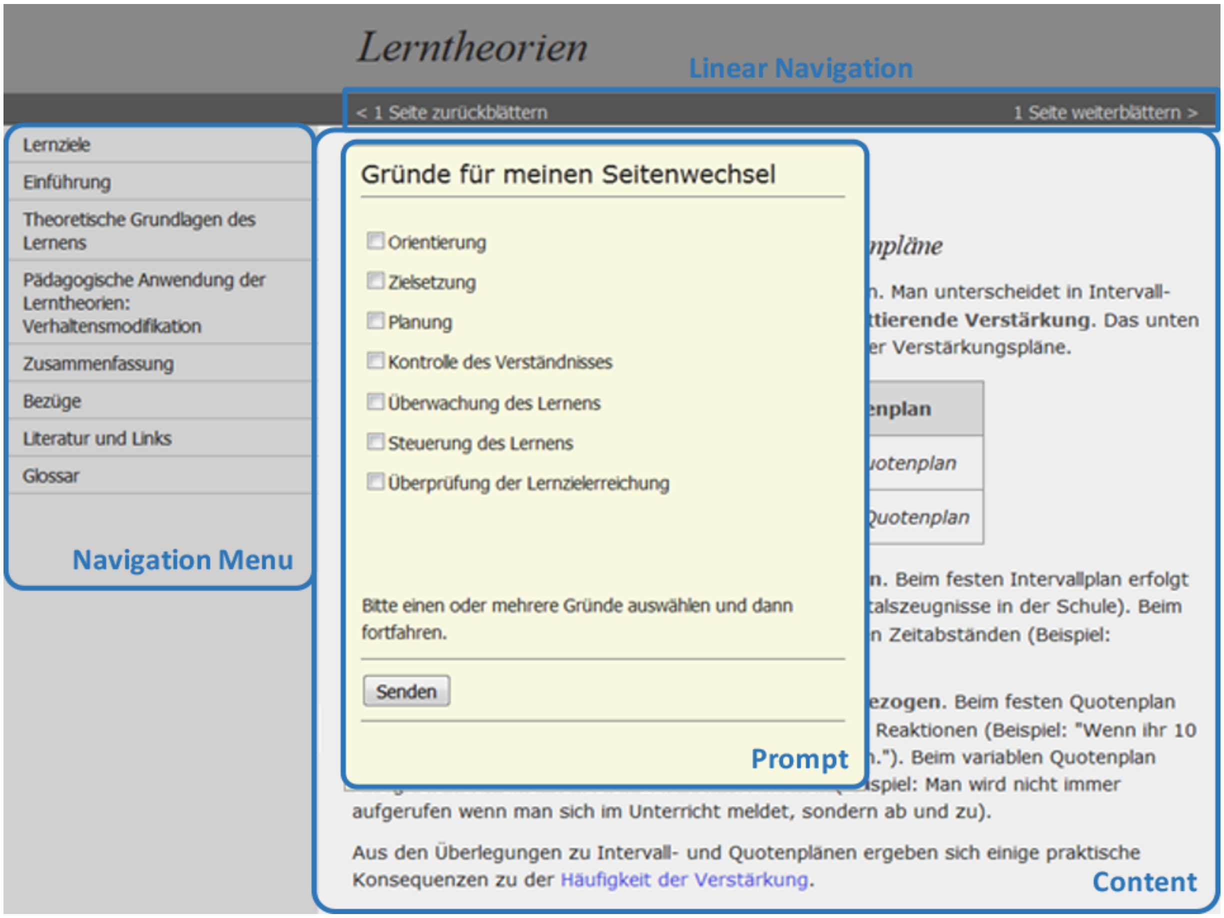

Figure 1: Screenshot of the learning environment including a metacognitive prompt. The main elements are labeled here for a better understanding of the environment, however it was not labeled in the learning sessions. Participants were asked to select at least one reason for node selection by choosing among a list of strategic reasons presented in the prompt window. The list comprised *orientation*, *goal-setting*, *planning*, *checking of understanding*, *monitoring of learning*, *control of learning*, and *evaluation of goal attainment*.

several times during learning. Each pop-up window comprised the same list of strategic reasons for node selection. Examples of these reasons are *orientation*, *goal-specification*, or *evaluation of goal attainment*. The participants had to select at least one reason for node selection before continuing with learning. Figure 1 presents a metacognitive prompt in the form of a pop-up window that was implemented in the learning environment.

Three knowledge tests on different levels based on Bloom's taxonomy of cognitive learning (Bloom, 1956) were used for performance measurement: (i) a free recall test, (ii) a comprehension test, and (iii) a transfer test. During the free recall test, the participants had to write down all of the basic concepts of operant conditioning that they could remember. The comprehension test assessed factual knowledge and comprised 22 multiple-choice items, each with one correct and three incorrect answers (Cronbach's α = .69). For example, the students were asked which of the following terms describes a primary reinforcer: money, praising words, food, or any stimulus directly following the behavior. Finally, transfer performance was measured by instructing the participants to apply their knowledge of the basic concepts and facts to eight prototypical situations in educational settings, which were not explicitly addressed in the learning material. For example, in one task they had to apply the principles of operant conditioning to solve a classroom situation in which a teacher experiences discipline problems. Two research assistants rated the answers to these situations on a researcher-developed rating scale (maximum score = 40 points; Cohen's κ = .84). In the case of disagreement among the raters, one of the authors determined the final score. More examples of the knowledge tests that were used in our studies are available in Bannert and Reimann (2012).

5.3. Procedure

The hypermedia learning session began with an introductory section. First, the experimenter explained the navigation elements of the learning environment. Then, the participant was instructed to perform a series of exercises using a practice lesson while thinking aloud concurrently during the task. More specifically, he or she was asked to verbalize every thought that came to his or her mind, without any interpretation or justification. These instructions refer to level 2 verbalizations, as specified by Ericsson and Simon (1993). If necessary, the experimenter provided feedback to the participant. Additionally, further exercises could be used until the participant firmly mastered the think-aloud technique.

Next, the participants received a short tutorial regarding the use of metacognitive prompts (approximately 10 minutes). This tutorial comprised information on the importance of reflecting on one's learning activities, an explanation of the reasons for strategic node selection that is listed in the prompts, and the desired usage of the prompts. Such a tutorial is necessary to guarantee the adequate application of metacognitive strategies during learning (e.g., Veenman, 2007). We derived the list of reasons for node selection from categories that had been developed in previous work (Bannert, 2006), in which students were asked to name their reasons freely. Following the tutorial, the participants were instructed to configure the prompts by the following arrangements. First, they could change the order of the items in the list of reasons for node selection. The initial order was randomized. Second, they were asked to define eight time stamps when the prompts should appear during learning.

As the next step, the participants engaged in 40 minutes of learning in our hypermedia learning environment. In the beginning, they were instructed about their learning task, that is, to learn the basic concepts of operant conditioning. During learning, the students were supported by metacognitive prompts. All of the participants were completely free to use the navigation elements of the learning environment and to use their learning strategies. During the whole learning phase, the participants had to read and to think aloud as practiced in advance, and their utterances were recorded using a microphone. If a participant stopped his or her verbalizations for more than five seconds, the experimenter prompted him or her to continue by saying "Please think aloud". Following the learning task, the participants worked on the three knowledge tests that are described above. The total duration of the session was approximately two hours.

5.4. Coding Scheme

Concurrent think-aloud protocols were used for the online measurement of the learning activities. The participants' recorded verbal protocols were segmented and coded post-hoc according to a coding scheme that was based on our theoretical framework of self-regulated hypermedia learning (Bannert, 2007). This conceptual framework organizes the student's activities into the three major categories *Metacognition*, *Cognition*, and *Motivation*. Additionally, it comprises several sub-categories, as further described in Table 1. Motivation refers to statements that reflect a beneficial or an obstructive embedding of a metacognitive or cognitive learning activity, particularly according to the concepts of achievement motivation, action control, and self-efficacy. More specifically, we assigned a motivational code when a student made an evaluative statement with respect to the task (e.g., "The task is quite difficult"), to his or her competencies (e.g., "I'm good at finding the relevant information"), and to the situation (e.g., "Thinking aloud isn't as troublesome as I expected"). These utterances reflect motivational states that might affect the self-regulatory behavior, for example, the use of strategies. According to Zimmerman (2008) especially self-motivation beliefs, such as self-efficacy, task interest, and value, might impact the student's engagement during learning.

In general, the coding process followed the procedure that was presented by Chi (1997). We segmented the verbal protocols by units of meaning. Thus, we assigned a segment for every definable learning activity. Furthermore, we did not use multiple nor nested codes. Four trained research assistants coded the verbalizations of all of the participants. We selected a random sample of six participants to compute the inter-rater reliability for our coding scheme. Based on 1,385 segments, the reliability showed substantial agreement (Cohen's κ = 0.78), which is considered to be sufficient for the following analysis.

5.5. Data Preparation

For the purpose of our analysis, we needed to prepare our data as described below. The starting point for the data preparation was the coded verbal protocols of 35 participants during 40 minutes of hypermedia learning. Each one of the students was supported by metacognitive prompts five to eight times during learning. The absolute frequencies and means of all of the coded learning activities, as well as the mean duration times of the events, are presented in Table 2.

The basic idea of our analysis was to use the coded learning activities, which represent the metacognitive, cognitive, and motivational utterances as described in the previous section, to classify each metacognitive prompt as "effective" or "non-effective" in inducing regulatory activities. We consider an increase in metacognitive utterances following a prompt, in relation to a student's individual baseline of metacognitive activities, to be an indicator of effectiveness. Therefore, we started our investigation by using a *Dotted Chart Analysis* (Song & van der Aalst, 2007; Van der Aalst, 2011), which is applicable with the ProM framework Version 5.2 (2008), to obtain an overview of the present process data. A dotted chart illustrates a sequence of events by arranging them as dots in a two-dimensional plane. The horizontal axis represents the occurrence of an event in time, and the vertical axis accounts for a case (i.e., the data of one participant). Figure 2 shows the visualization of the coded learning activities during the first ten minutes of four sample cases (i.e., participants CHRO10, KAHA17, KOGE04, and VAIV20). An inspection of the events following a prompt (i.e., the events following the yellow sections) showed that some prompts are followed by a high number of metacognitive events, whereas other prompts are not. Furthermore, the increase in metacognitive events that follow a prompt—if present—usually takes up to two minutes.

Table 1: Coding scheme for analyzing student's learning activities

Code	Category	Description and Examples
	Metacognition	
ORIENT	Orientation	Task clarification, overviewing the material to prepare strategic learning behavior *At first I read my learning goals to get an overview of my task.*
SETGOAL	Goal Specification	Goal setting and sub-goaling *I have to learn the basic concepts on operant conditioning.*
PLAN	Planning	Planning of proceeding *First I will read the introductory text, then I will decide in which sequence I will proceed.*
SEARCH	Search	Searching information *Now I'm looking for information on reinforcement plans.*
JUDGE	Judgement	Judgements of relevance of information *Skinner's Vita is not relevant for my learning task.*
EVAL	Evaluation	Evaluating the attainment of goals or sub-goals *Did I process all topics according to my learning goals?*
MONITOR	Monitoring	Monitoring and controlling of one's learning *Ah now I understand the principle and I can proceed to my next learning goal.*
	Cognition	
READ	Reading	Reading out loud
REPEAT	Repeating	Repeating in terms of memorizing *Re-reading a paragraph or notes*
ELABORATE	Elaboration	Deeper processing: paraphrasing, connecting, inferring *I already know the Skinner Box from my biology class.*
ORGANIZATION	Organization	Organization of information *Drawing a map, writing down major concepts*
	Motivation	
MOT	Motivation	Evaluative statements regarding the task, the student's competencies, or the situation *The task is very interesting and relevant for my studies.* *I'm good at memorizing this subchapter.* *This section distracts me from my original goal.*
	Residual Category	
OTHER	Others	Off-topic statements, comments on technique, not interpretable statements, pauses *May I make notes? The mouse doesn't work well.*

Table 2: Absolute frequencies and means of all coded learning events, and mean duration times of events during 40 minutes of hypermedia learning (N = 35)

	Absolute Frequency					Duration	
	Min	Max	Frequency	M	SD	M_Dur	SD_Dur
Metacognition	34	242	4075	116.43	45.97	4.54	7.16
Orientation	5	30	501	14.31	7.23	6.10	5.19
Planning	0	5	61	1.74	1.65	5.76	3.21
Goal Specification	0	10	72	2.06	2.36	6.14	3.09
Search	1	32	414	11.83	7.54	8.53	9.18
Judgment	2	23	409	11.69	5.70	4.48	2.61
Evaluation	0	15	127	3.63	3.08	13.41	29.46
Monitoring	11	203	2491	71.17	37.62	3.07	2.74
Cognition	47	201	3760	107.43	36.01	14.72	14.93
Reading	20	84	1423	40.66	15.75	17.56	18.20
Repeating	2	45	647	18.49	11.04	10.82	10.47
Elaborating	3	55	767	21.91	12.92	10.35	11.18
Organizing	3	61	923	26.37	13.87	16.62	13.12
Motivation	0	18	72	2.06	4.14	4.07	2.46
Others	9	76	1124	32.11	14.71	3.25	4.01
Sum of all coded events	127	482	9031	258.03	78.36	8.87	12.08

Note. M(Dur) = mean duration time, SD(Dur) = standard deviation of duration time; duration times in seconds.

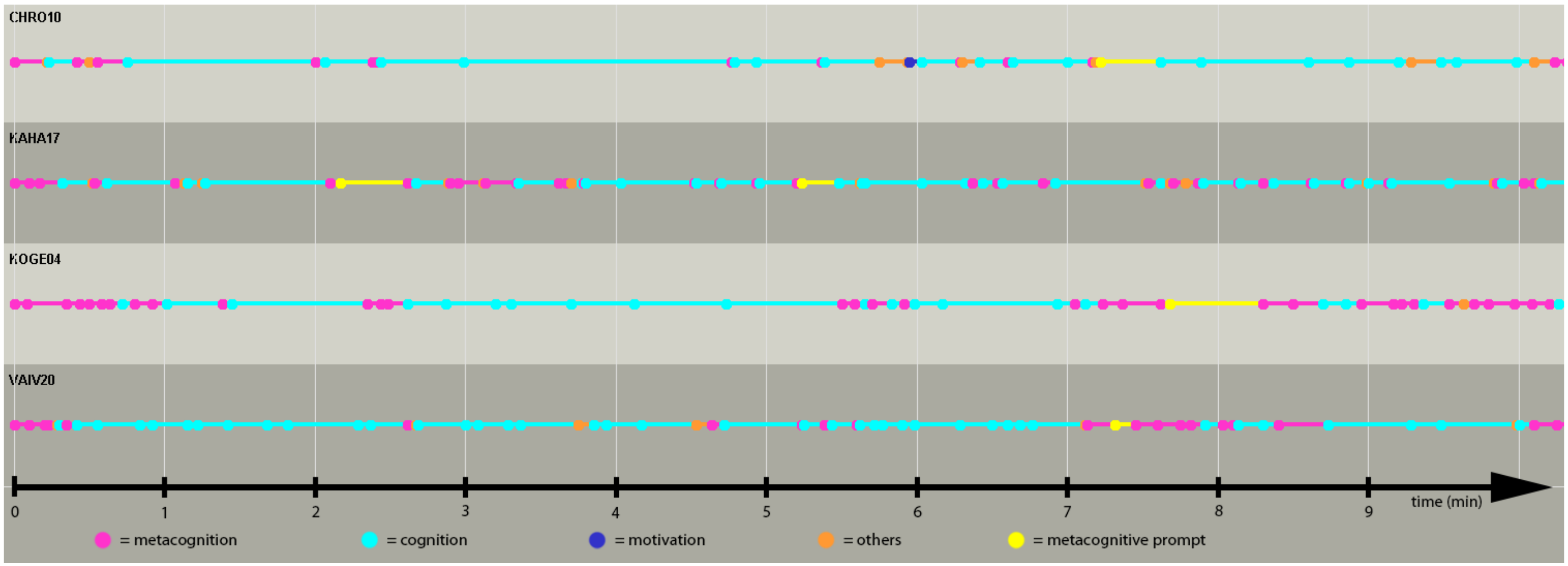

Figure 2: Visualization of four cases using a Dotted Chart Analysis of coded verbal protocols. The horizontal axis represents the first ten minutes of learning with each segment representing one minute. Each dot represents the occurrence of an activity or a prompt presentation. The activities are color-coded and explained at the bottom. The line between two dots represents the duration of an event.

Based on this observation, we decided to use a two-minute time interval following each prompt to determine the successful induction of metacognitive activities. There are additional reasons that support the selection of this measurement unit. A shorter time interval would reduce the sample size to investigate the learning process, that is, the number of investigated events. Here, one must consider that the duration time of learning activities varies from a few seconds (e.g., monitoring one's learning progress) to longer time periods (e.g., reading a paragraph). Table 2 presents the means and standard deviations for the duration times of all of the coded categories. On the other hand, the selection of a longer time interval would be accompanied by a possible overlap of prompt presentations. The participants were instructed to determine eight time stamps for the presentation of prompts, considering a distance of at least two minutes between two presentation times. Each selected time interval comprised all of the coded utterances within this timeframe, including overlapping events, that is, events that started and respectively ended beyond the two-minute interval. Because the duration of events varied from seconds to minutes, the number of events that are included in a time interval also varied (M = 14.16, SD = 5.86, Min = 1, Max = 31).

In addition to selecting the two minutes following each prompt, we also computed an *individual baseline* of metacognitive utterances for each participant. This baseline is necessary to determine if an increase of metacognitive events following a prompt occurred. It represents the amount of spontaneous metacognitive utterances within an interval of two minutes. For the computation of this baseline, we excluded the two-minute time interval following each prompt because this time span is supposed to be affected by the prompted requests. The remaining time span of the total learning time was used to compute a baseline for each participant. For example, one participant showed a total of 127 metacognitive utterances within the total learning time of 40 minutes. He or she received eight prompts with 64 metacognitive utterances within the two-minute time intervals following each prompt. For the individual baseline, we considered the number of remaining metacognitive events (i.e., 127 - 64 = 63) and the remaining time span (i.e., 40 - 16 = 24 minutes). In this case, the computation yields a baseline of 5.25 metacognitive utterances within a time span of two minutes. Overall, the individual baselines for all of the students ranged from 1.47 to 12.68 metacognitive utterances (M = 5.32, SD = 2.29).

Although the observed learning time of 40 minutes was quite short, and, therefore, we did not expect the number of metacognitive utterances to vary significantly by time, we conducted additional analyses to account for a possible influence of time or the number of prompts that were received. First, we checked if the number of prompts that are received predicts the number of metacognitive utterances following a prompt (i.e., a possible additivity of prompting effects). The results of a linear regression showed no significant correlation between prompt number and metacognitive utterances, $F(1, 239) = 3.16$, $p = .077$, $R^2 = .013$, $R^2_{adjusted} = .009$. Figure 3 presents the mean numbers of metacognitive utterances for all of the students (N = 35) in each two-minute time interval following a prompt. Second, we considered whether there was a general upward drift of metacognitive utterances during the total learning time. Again, a linear regression using the time interval as the predictor and the number of metacognitive utterances as the independent variable revealed no significant influence of time, $F(1, 698) = 2.36$, $p = .125$, $R^2 = .003$, $R^2_{adjusted} = .002$. Additionally, Figure 4 shows that no upward trend is present. The high number of metacognitive events during the first time interval can be explained by necessary orientation activities at the beginning of the learning phase. To summarize, the number of metacognitive utterances is not significantly affected by time or by the number of prompts that are received. Therefore, the computation of the baseline as described above does not need to be

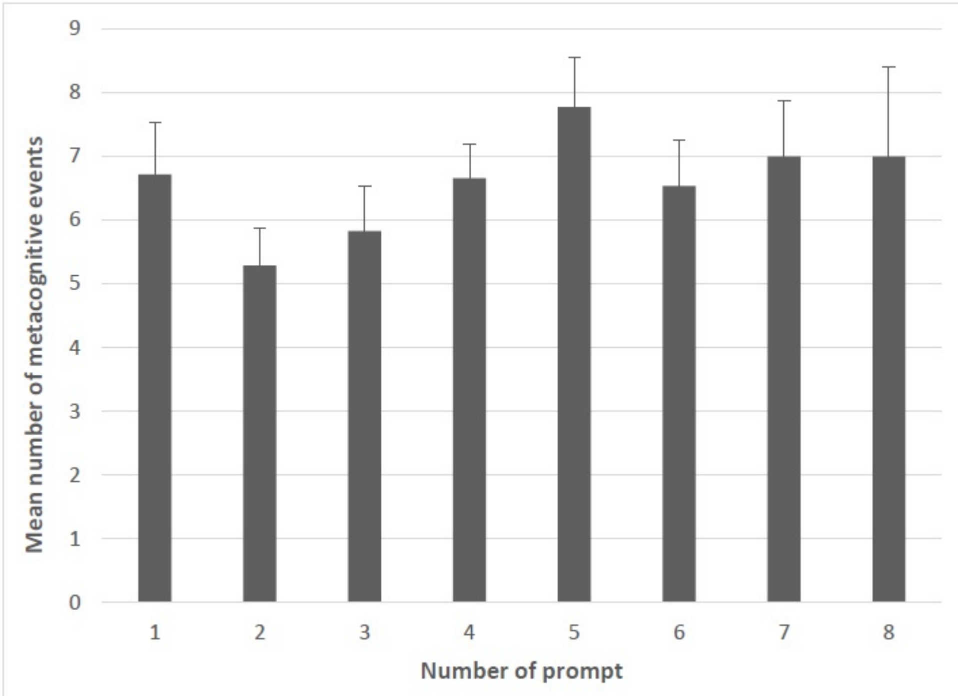

Figure 3: The mean number of metacognitive utterances for all students ($N = 35$) in each two-minute time interval following a prompt. Error bars represent standard errors.

adjusted for these factors.

By comparing the individual baseline with the number of metacognitive utterances in the two-minute time interval following a prompt, it is possible to determine whether there was an increase of metacognitive activities. If the number of metacognitive utterances exceeds the individual baseline, we regard the prompt as effective. Because it may be possible that both values are equal or very close to one another, we decided to set a cutoff value to solve those cases. We selected an absolute value of one metacognitive event because it represents the smallest possible threshold. For example, if the individual baseline of a participant is 5.30, then the number of metacognitive utterances in the time interval following a prompt must be greater than 6.30 to indicate an increase, that is, a successful induction of metacognitive activities based on our criteria.

Finally, we selected two-minute time intervals preceding the presentation of prompts for the purpose of our third research question, that is, an analysis of the conditions of effectiveness. Again, we decided that this time span represents a suitable sample for our analysis. The selection of time intervals that precede and follow a prompt presentation is illustrated in Table 3, using two prompting times of an example case.

5.6. Analysis Techniques

In the first step of our analysis, we determined the number of metacognitive activities within each two-minute time interval following a prompt. Then, we verified the chosen type of classification by comparing the students' mean baseline of metacognitive utterances with the mean number in the time intervals that follow each prompt: If the prompts induced metacognitive activities

Table 3: The selection of two-minute time intervals preceding and following the presentation of the first two prompts for the case EDMI18

Case ID	*Activity (Code)*	*Timestamp*
	...	
EDMI18	READ	00:04:46
EDMI18	PLAN	00:06:08
EDMI18	MONITOR	00:06:09
EDMI18	READ	00:06:14
EDMI18	MONITOR	00:06:48
EDMI18	**PROMPT1**	**00:06:56**
EDMI18	OTHER	00:07:19
EDMI18	READ	00:07:44
EDMI18	MONITOR	00:07:46
EDMI18	ORGANIZATION	00:08:10
EDMI18	OTHER	00:08:13
EDMI18	MONITOR	00:08:19
EDMI18	ELABORATE	00:08:25
EDMI18	MONITOR	00:08:27
EDMI18	ORGANIZATION	00:08:52
EDMI18	MONITOR	00:08:53
EDMI18	READ	00:09:04
EDMI18	MONITOR	00:09:04
EDMI18	ORGANIZATION	00:09:10
	...	

Case ID	*Activity (Code)*	*Timestamp*
	...	
EDMI18	MONITOR	00:11:13
EDMI18	ORGANIZATION	00:11:19
EDMI18	OTHER	00:11:53
EDMI18	ELABORATE	00:11:56
EDMI18	ELABORATE	00:12:38
EDMI18	OTHER	00:12:50
EDMI18	ELABORATE	00:12:52
EDMI18	PLAN	00:12:56
EDMI18	MONITOR	00:13:00
EDMI18	EVAL	00:13:02
EDMI18	MONITOR	00:13:10
EDMI18	**PROMPT2**	**00:13:13**
EDMI18	OTHER	00:13:24
EDMI18	SETGOAL	00:13:26
EDMI18	ORGANIZATION	00:13:27
EDMI18	READ	00:13:35
EDMI18	ORGANIZATION	00:13:59
EDMI18	READ	00:14:25
EDMI18	MONITOR	00:14:31
EDMI18	ORGANIZATION	00:14:32
EDMI18	READ	00:14:38
EDMI18	MONITOR	00:15:09
EDMI18	ORGANIZATION	00:15:11
EDMI18	OTHER	00:15:24
	...	

Note. For an explanation of the codes please see the coding scheme presented in Table 1. The timestamp indicates the starting time of an activity, using the format hh:mm:ss.

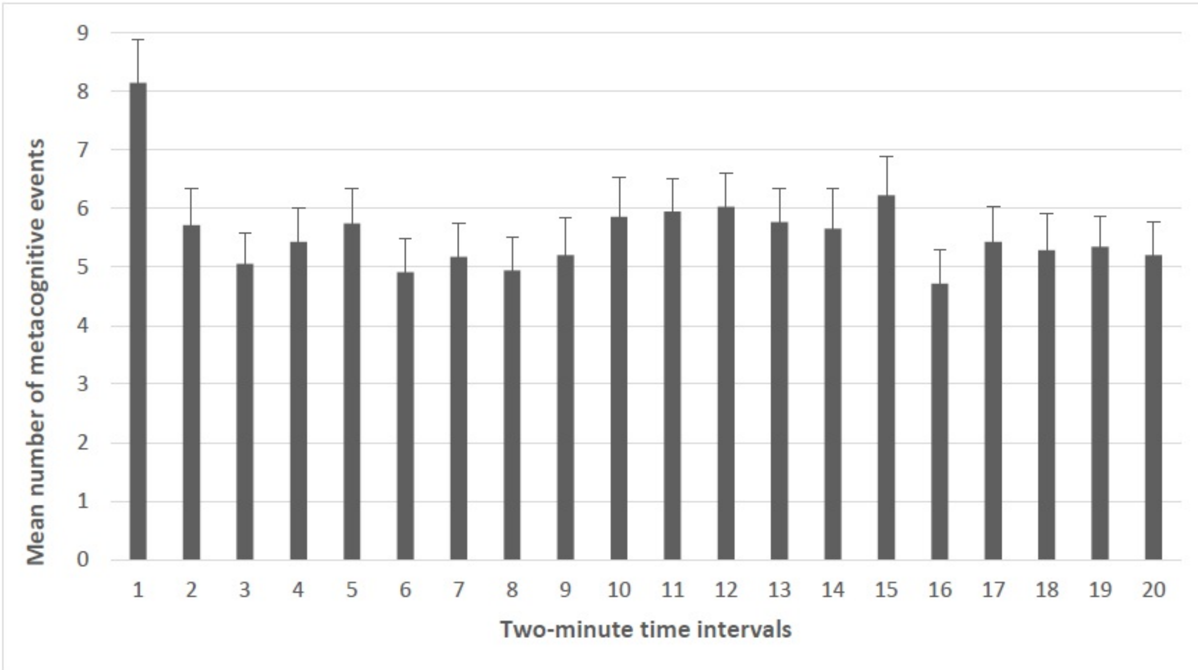

Figure 4: The mean number of metacognitive utterances for all students (N = 35) in each two-minute time interval. The total learning time was 40 minutes, resulting in 20 two-minute time intervals. Error bars represent standard errors.

within the determined time interval successfully, the mean number of metacognitive utterances in these intervals should be higher than the mean baseline. Additionally, we examined the correlation between the number of induced metacognitive activities and learning performance. According to our theoretical framework, we assumed that students who show a higher number of induced metacognitive activities should perform better than students who show less induced regulatory activities.

The second step of our analysis considered the conditions of effectiveness. For this purpose, we first used the RapidMiner Studio Version 6.3 (2015) and a linear regression learner to explore the learning activities that occurred before a prompt appeared. A linear regression can be used as a classification method in the case of exclusive numeric attributes, and it has the advantages of working well on small data sets and of generating simple models to enable a natural interpretation (Hämäläinen & Vinni, 2010). As described above, we decided to use two-minute time intervals for our analysis. The coded learning activities that occurred before each prompt were used as the predictors, and the number of metacognitive events that followed each prompt as the outcome variable. We applied cross-validations to avoid over-fitting of the learning algorithm. By using this data mining (DM) approach, we expected to find conditions, that is, states of the learning process, that indicate when a prompt is more likely to be effective and non-effective, respectively, in inducing regulatory activities.

Finally, we applied PM to gain more detailed insights into the learning process preceding the effective versus non-effective prompts by taking into account the sequential order of the learning activities. For this analysis, it was necessary to classify the prompts into the discrete categories of "effective" and "non-effective" in inducing metacognitive activities, as described in the section titled "Data Preparation". We used the ProM Framework Version 5.2 (2009) and the Fuzzy Miner algorithm (Günther & van der Aalst, 2007) to generate inductively a process model for the learning events that preceded all of the prompts that were classified as effective and a process model for the activities that preceded all of the prompts that were classified as non-effective. A comparison of these process models provides more detailed information on the conditions that may influence the effectiveness of prompts. In the following section, we provide a short introduction to the principle of the Fuzzy Miner algorithm.

Fuzzy Mining (Günther & van der Aalst, 2007) is an approach that was designed to find underlying processes in data that are less structured in appearance, such as our coded learning activities. The algorithm allows a flexible simplification of output models by distinguishing between the important and the less important details of an input event sequence. Thus, it is possible to generate a model that emphasizes the main features of a process and that is easily understandable. More precisely, the Fuzzy Miner transforms an event sequence into a process model that consists of nodes (event classes or categories) and edges (relations between two event classes). The data input comprises several cases with every case including an event sequence, which is ordered by a timestamp. First, the Fuzzy Miner algorithm uses these data to generate a complete model that comprises all of the observed nodes and edges by taking into account the relative importance and the sequential order of all of the events. Then, the algorithm uses two fundamental metrics, which are referred to as *significance* and *correlation*, to compute a simplified model for the given data set. *Significance* measures the relative importance of the occurrence of event classes and relationships between events. For instance, events that occur more frequently are assessed as being more significant. *Correlation* is calculated for edges. It indicates the closeness of two events, measured by their temporal proximity. The basic concepts of significance and correlation are embedded in a metrics framework that calculates three primary types of metrics: unary significance (event classes), binary significance (relationships between event classes), and binary correlation (relationships between event classes). Günther and van der Aalst (2007) describe these metrics in more detail. As a final step, the model is simplified by making decisions regarding the inclusion of nodes and edges in the final model using the following rules: Events that are highly significant are preserved, events that are less significant, but highly correlated, are aggregated, and events that are less significant and lowly correlated are removed. It is possible to influence the model simplification by parameter setting, for example, by specifying cutoff values. To bring structure to the model, the algorithm uses edge filtering and thereby tries to focus only on the most important relationships between event classes. The utility of edges, which is the weighted sum of significance and correlation of an edge, is calculated, and this weighting is then configured by the utility ratio. Moreover, by setting an edge cutoff, an absolute threshold value for filtering edges can be determined. The higher the value at which the edge cutoff is set, the more likely the Fuzzy Miner is to remove an edge. Finally, there is another important mean to simplify the model: node aggregation and abstraction. Nodes are removed based on a parameter referred to as the node cutoff. If the unary significance of a node is below this cutoff, it will be excluded from the resulting model, or it will be aggregated. The latter happens if it is possible to preserve less significant nodes by merging them into a cluster of highly correlated nodes.

6. Results

This section presents the results of our analyses as follows. First, we report the classification results using the learning activities following each prompt and an individual baseline of metacognitive utterances. Additionally, we refer to the correlation between our classification and learning performance. Second, the findings of the investigation of conditions for the effectiveness of prompts are reported. We further present the results of a linear regression learner as well as those of a process mining algorithm.

6.1. Classification of Prompts

In the first step of our analysis, we took into account the number of metacognitive utterances in a two-minute time interval following each prompt as well as an individual baseline to determine each prompt's effectiveness in inducing regulatory activities. To verify our classification criteria, we used a Wilcoxon signed-rank test to compare the participants' mean baseline and the mean number of metacognitive utterances following each prompt presentation. The number following the prompts was significantly higher (Mdn = 6.80) than the individual baseline (Mdn = 5.05), z = -4.18, $p < .001$, d = 2.00. Thus, in general, the participants showed significantly more metacognitive activities within the two-minute time intervals following each prompt compared to their individual baseline (i.e., the mean number of metacognitive utterances within any two minutes). The results support the selection of a two-minute time span following a prompt to evaluate the successful induction of metacognitive activities.

Next, we analyzed the correlation between the induced number of metacognitive utterances through prompts and learning outcome. The alpha level was adjusted according to the procedure of Benjamini and Hochberg (1995) because we conducted multiple tests of statistical significance. The results showed no significant correlation for any of the performance measurements (recall: r = -.23, comprehension: r = .08, transfer: r = .12). Because the total number of metacognitive activities within 40-minutes learning time also showed no significant correlation with learning outcome, we checked the correlations with the sub-categories. The findings revealed that the sub-category Monitoring shows the highest positive correlation with two measurements of learning performance (recall: r = -.17, *n.s.*, comprehension: r = .25, *n.s.*, transfer: r = .27, *n.s.*). Therefore, we additionally analyzed the correlation between the number of monitoring activities within the two-minute intervals following the prompts and learning performance. The number of monitoring activities corresponds positively with transfer performance (r = .26, *n.s.*) and comprehension performance (r = .22, *n.s.*), but not with recall performance (recall: r = -.18, *n.s.*). Possibly due to the small sample size, all of the correlations are non-significant. A regression equation to predict transfer performance ($y = 18.95 + 0.068\ x$) shows that a low number of induced monitoring activities through prompting (M = 24.54, SD = 15.89; $M - SD$ = 8.65) results in a performance score of y = 19.54 points and that a high number ($M + SD$ = 40.43) results in a score of y = 21.70 points. These observations are in line with previous work (Bannert & Mengelkamp, 2013), which indicated that prompting primarily enhances transfer performance. Additionally, a process analysis showed that the positive effect on transfer performance is mainly mediated by inducing monitoring activities (Sonnenberg & Bannert, 2015). In conclusion, the positive correlation between the number of monitoring activities and transfer performance indicates the usefulness of our classification based on a two-minute time interval following each prompt. The students who showed a higher number of induced monitoring activities through prompting performed better at transfer tasks than the students who showed less monitoring events.

Moreover, the classification of prompts into two discrete classes (effective and non-effective, respectively, in inducing metacognitive activities), based on a comparison of the individual baseline and the number of metacognitive utterances following a prompt, resulted in a total of n = 113 effective prompts and n = 127 non-effective prompts for our learner sample (N = 35). Furthermore, the classification shows that each student received a mean of M = 3.23 (SD = 1.46, Min = 1, Max = 6) effective prompts. Except for one student whose prompts were all classified as being effective, each student received both effective and non-effective prompts.

The selection of two-minute time intervals following the presentation of a prompt and the classification of prompts represents the starting point for the following analyses, which aim at exploring conditions for effectiveness. Please note that for the purposes of these analyses, each prompt was regarded as a separate case labeled either as "effective" or "non-effective" (N = 240, $n_{effective}$ = 113, $n_{non\text{-}effective}$ = 127).

6.2. Conditions for the Effectiveness of Prompts

We analyzed the coded events that preceded effective and non-effective prompts by using two approaches. First, we applied a linear regression learner taking into account the absolute frequency of coded events as predictors and the number of metacognitive events following a prompt as the outcome variable. Then, we considered the sequential order of coded learning activities by generating process models for the time interval before a prompt appeared using a PM algorithm.

Linear regression learner. We applied a linear regression learner to our data set using the RapidMiner Studio Version 6.3 (2015). Because we used the same data set for training and for measuring the performance of the linear regression model, we used two cross-validation techniques to avoid over-fitting. First, a ten-fold cross-validation was conducted, randomly splitting the data into a training and a test set at the prompt-level (N = 240 prompts). Second, we conducted a student-level cross-validation by defining a subset for each student (in RapidMiner: Batch-X-Validation) to account for the possible influence of student factors. In this procedure, complete cases are removed iteratively before training the algorithm and are then used to estimate the goodness of the model. Moreover, we activated the automatic feature selection using the M5 prime method. The input predictors were all learning activities that occurred in a two-minute time interval preceding each prompt, whereas the outcome variable was the number of metacognitive utterances that followed each prompt. We used the VIF and tolerance statistics to assess potential multicollinearity within our data. Because VIF values were well below 10 (average VIF = 1.182) and tolerance statistics were well below 0.2, there was no cause for concern. The feature selection method selected the categories Orientation, Search, and Monitoring as relevant predictors for the regression model.

Table 4 presents the weight table of the linear regression model that was trained with our data. The results show the influence of the regression coefficients. Based on these coefficients, the findings indicate that a high occurrence of orientation and monitoring activities preceding the prompt presentation are significant predictors for the successful induction of metacognitive events. The occurrence of all other learning activities has no significant influence on prompt effectiveness in the resulting linear model. The performance measurement of the linear regression model using a ten-fold cross-validation showed a root-mean-squared error of 3.64 +/- 0.55 and a correlation of .46 +/- .19. In case of the student-level cross-validation, the model goodness was slightly lower ($RMSE$ = 3.60 +/- 1.23, R = .22 +/- .26).

Table 4: Weight table of the learned linear regression model

	B	*SE B*	β	*p*
Constant	3.40	0.44		$< .001$
Orientation	0.78	0.20	.24	$< .001$
Search	0.13	0.22	.01	.550
Monitoring	0.61	0.08	.05	$< .001$

Note. Prediction of the metacognitive activities following a prompt by considering all activities in a two-minute time interval preceding the prompt presentation. *N* = 240 prompts. All codes were included, but only three codes were selected by feature selection. Estimation of model goodness using cross-validation: prompt-level, *RMSE* = 3.64, R^2 = .24; student-level: *RMSE* = 3.60, R^2 = .12.

PROCESS MINING USING THE FUZZY MINER ALGORITHM. In a further analysis, we used the approach of PM and an algorithm for process discovery. This algorithm inductively generated process models for the time interval before an effective prompt appeared and before a non-effective prompt appeared. Our analysis was conducted by using the following parameter settings: edge filtering was set with the edge cutoff = 0.20 and the utility ratio = 0.75 (both are default values); the significance cutoff of the node filter was set to 0.20.

Figure 5 shows the resulting models with the event classes and their process relationships. Event classes are represented by the rectangular nodes that include the label and its significance (a value between 0 and 1). The arcs between categories indicate successive events (the upper number displays significance, that is, their relative importance, and the lower number shows correlation, that is, a measure which indicates how closely related two events are). The arcs that point towards a category indicate a repeated occurrence of that category. Less significant and lowly correlated events were discarded from the process model. Thus, the nodes and arcs that fall into this category were not included in the graph.

In Figure 5, the model of the learning process preceding non-effective prompts (left part) comprises nine event classes, and the model of effective prompts (right part) comprises seven categories with two clusters. The model of non-effective prompts does not include the categories SETGOAL and PLAN, whereas in the model of effective prompts, the categories PLAN and SEARCH as well as JUDGE and EVAL were combined into a cluster. This means that these event classes did not reach the significance cutoff value. Here, the main idea of the Fuzzy Miner is realized by abstracting from information that is perceived as being too fuzzy, that is, it does not play a significant role in the process as a whole.

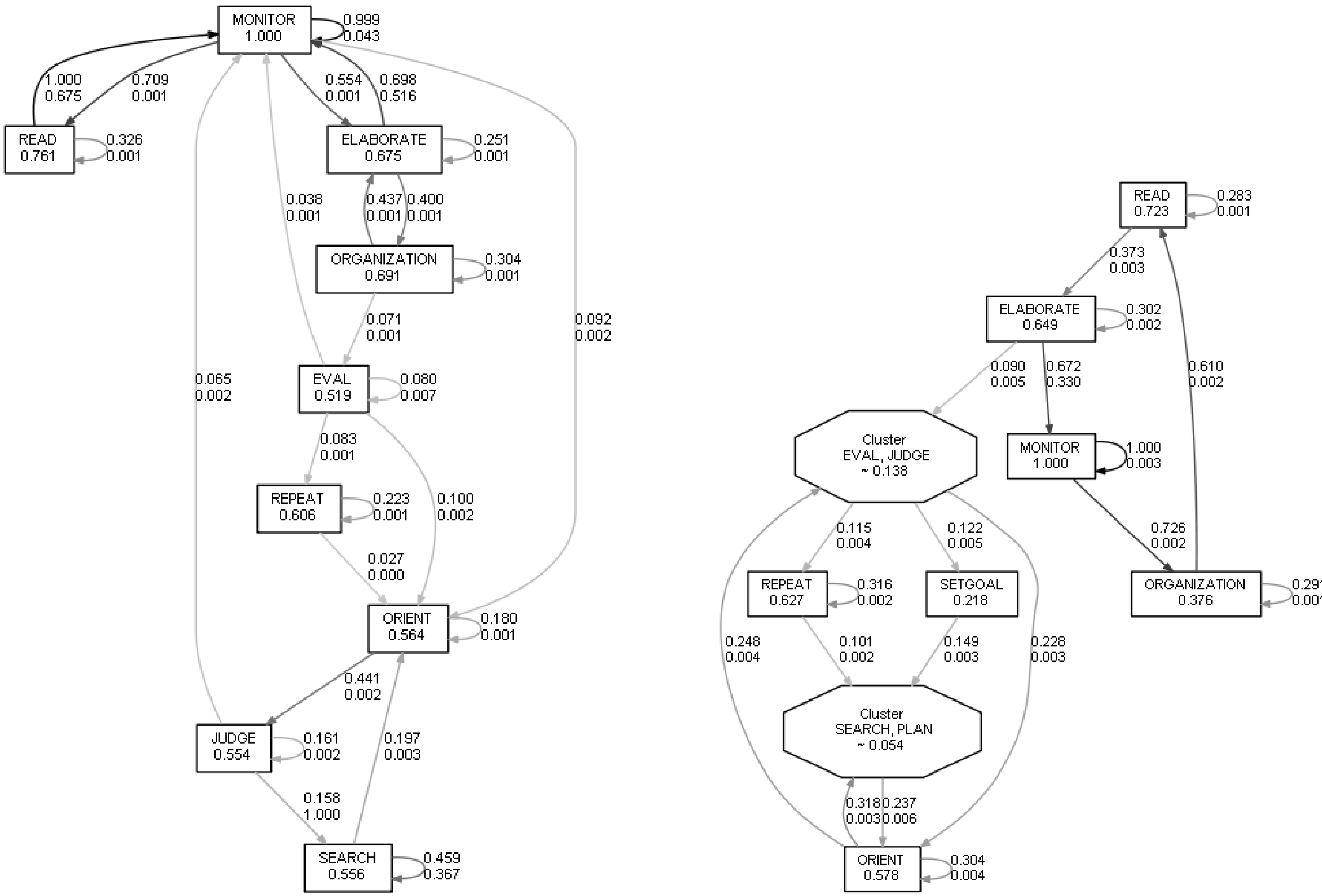

Figure 5: Process models for the time interval preceding non-effective prompts (n = 127, left) and effective prompts (n = 113, right). For an explanation of the codes please see the coding scheme presented in Table 1.

The process model of non-effective prompts comprises five metacognitive event classes. Furthermore, the most important event classes are MONITOR, READ, ORGANIZATION, and ELABORATION. Referring to the edges in the process model, the most important connections among event classes are between READ and MONITOR, between ELABORATE and MONITOR, between ORIENT and EVAL, as well as between ORGANIZATION and ELABORATE. Moreover, the classes MONITOR and ORIENT show several connections to other events, thus representing significant routing points for the process. Finally, there is a triple loop between READ, MONITOR, ELABORATE, and ORGANIZATION, which indicates a very active processing of information which is constantly monitored. In general, the integration of metacognitive event classes in the model of non-effective prompts can be interpreted as a successful regulation of learning activities with respect to SRL assumptions. According to SRL models (e.g., Winne & Hadwin, 2008; Zimmerman, 2008), students ideally perform various metacognitive activities and run through different regulation phases (forethought, performance, and evaluation).

In contrast, the process model of effective prompts only shows the three metacognitive categories, MONITOR, ORIENT, and SETGOAL, but all of the four cognitive event classes previously described. The most important event classes are MONITOR, READ, ELABORATE, and REPEAT. Furthermore, important edges are between MONITOR and ORGANIZATION, between ELABORATE and MONITOR, as well as between ORGANIZATION and READ. There is a cycle that connects READ, ELABORATE, MONITOR, and ORGANIZATION. This cycle can be interpreted as the processing of information that is monitored. However, compared to the model of non-effective prompts, there is a less active interconnection between these four event classes. Finally, the event classes REPEAT, SETGOAL, and ORIENT represent a second cycle, which shows a weak structure between these events. In general, the performance of cognitive activities, which is connected with MONITOR, represents the only clear structure in this process model. Furthermore, there is poor regulation of learning activities in comparison to the model of non-effective prompts.

In summary, the process model of non-effective prompts represents a highly regulated learning process, including various metacognitive categories, and a very active cycle of information processing. In contrast, the model of effective prompts shows a learning process that comprises few metacognitive event classes, and a cycle of monitored information processing. However, this cycle shows a weaker interconnectedness than the active cycle in the model of non-effective prompts. As a result, in the case of the model of non-effective prompts, an intervention through a metacognitive prompt may not be necessary, or it may even be disturbing considering the already high state of regulation. Moreover, interventions in a process that comprises very active information processing may not be the right context to foster metacognition because the learner has to interrupt his or her cognitive activities to interact with the prompt and follow its request. Consequently, this can lead to ignoring the prompts and to a continuation of one's information processing activities. On the other hand, the presentation of a metacognitive prompt can be more effective if the learning process comprises a weak regulatory behavior and a less intensive cycle of cognitive activities. With respect to the findings of the linear regression, the precursor states that are likely to make a prompt effective may be orientation and monitoring events that are, however, not well embedded in the course of learning, as displayed in the structure of the process model of the effective prompts. Moreover, an effective prompt may foster the learning process by supporting the interconnectedness between the two cycles, especially by stimulating the connections that lead back from the regulatory behavior to the information processing cycle. As is shown above, particularly induced monitoring activities seem to cause better learning.

This may be due to the fact that effective prompts helped to cycle back to monitoring and the cognitive loop.

7. Discussion

In this contribution, we evaluated the effectiveness of metacognitive prompts on the micro level to investigate the conditions for effective scaffolding during hypermedia learning. Because the research on SRL and scaffolding hypermedia learning has noted the importance of tailoring support to the needs of the learner (e.g., Azevedo & Hadwin, 2005), and learning systems offer the possibility of implementing adaptive features (e.g., Brusilovsky, 2001; Molenaar & Roda, 2008), we attempted to explore the reasons why a provided metacognitive prompt is sometimes effective in stimulating regulatory activities and why sometimes it is not. In general, this type of research is important for the improvement of scaffold design because information on the conditions for effectiveness can inform the positioning of support in the course of learning (e.g., Thillmann et al., 2009). For our purpose, we analyzed fine-grained process data, which was measured by coded think-aloud protocols, using DM and PM. With this approach, we sought to classify the effectiveness of prompts, to investigate the correspondence of the classification with learning outcome, and to discover the conditions under which prompts induce regulatory activities (i.e., the proper temporal positioning).

Our findings indicate that it is possible to distinguish between effective and non-effective prompts by considering the increase of metacognitive utterances following the prompt presentation. Overall, approximately half of the prompts that were provided to the students were classified as effective according to our chosen classification criteria. Because we had to make a decision about how to determine the time intervals to investigate the increase of metacognitive utterances, it is possible that a different selection of measurement units would result in a different prompt classification. The determination of measurement units is a general challenge in analyzing fine-grained process data (Johnson et al., 2011; Winne, 2014). Therefore, we validated our decision by showing that the mean number of metacognitive events following the prompts is significantly higher than the baseline of metacognitive utterances. Moreover, we related the selection of time intervals to learning performance. Ideally, the students who showed a higher number of induced metacognitive activities in the two-minute time intervals following a prompt would also show a better learning performance. The results revealed that there is no significant correlation between the total number of induced metacognitive activities and learning performance. A closer look at the sub-categories of metacognition showed that the highest correlation was between monitoring and transfer performance. However, this small effect was not significant, possibly because of the small sample size. Nevertheless, the positive correspondence between monitoring and transfer performance is in line with the previous research (Sonnenberg & Bannert, 2015).

Based on the classification of effective and non-effective prompts, it was our next aim to compare these two types by considering the students' current learning process preceding a prompt. Depending on the learning activities that were performed by the students (e.g., reading, the processing of information, or monitoring and controlling their learning progress), it is possible that the presentation of a scaffold affects the learning process differently. For example, one reason for the non-effectiveness of a provided prompt in our study may be a suboptimal positioning within the student's learning process. By applying DM and PM techniques to the coded learning activities, we tried to discover indicators of the effectiveness of a prompt (i.e., the conditions under which it might be more likely to be effective and non-effective, respectively) to improve the positioning of metacognitive support in future scaffold design. First, we analyzed the frequency of learning activities in the time intervals that preceded each prompt by applying a linear regression learner. The findings indicate that a high occurrence of orientation and monitoring activities fosters the desired prompting effects (i.e., the activation of regulatory activities).

These findings were supported by applying a PM algorithm that not only takes into account the frequency of coded learning activities but also their relative positioning (i.e., their sequential order). We compared two process models, both inductively generated by the Fuzzy Miner algorithm, that represent the learning processes (more precisely, the workflow of learning activities) that precede the appearance of an effective and a non-effective prompt. The sequence of learning activities in the process model of non-effective prompts resembles the performance of successful regulation patterns as described in SRL models (e.g., Winne & Hadwin, 2008; Zimmerman, 2008). Additionally, this process model shows a very active sequence of monitored information processing. In contrast, the process model of effective prompts comprises the execution of cognitive activities, which are only connected with one metacognitive activity, namely monitoring. This could be interpreted as a poor regulation of the learning process.

In summary, the findings of the learned linear regression function and the process models indicate that the occurrence of orientation and monitoring activities, if they are not yet well embedded in the course of learning, increase the likelihood of effectiveness. In this case, metacognitive prompting may be successful in fostering further regulatory activities and in structuring the regulating of the learning process. If students already show a highly regulated process and an intense sequence of cognitive processing, the intervention of a metacognitive scaffold may not be necessary, and it may even be disruptive. These results relate to the implications for the design of adaptive systems based on attention management of Molenaar and Roda (2008). The authors note that one has to evaluate the cost of switching attentional focus and that learners should be able to predict interruption times.

Our results can be considered for future scaffold design, for example, for adapting the presentation of a metacognitive prompt to the current learning process. In general, our analysis shows how to use process data to evaluate the effectiveness of provided support on the micro level by applying DM and PM techniques. Furthermore, the findings strengthen the importance of using process data in the investigation of the effectiveness of instructional support, which in turn provides implications for improving its design (e.g., by providing conditions for adaptive scaffolds).

With regard to the limitations of our study, it should be noted that the results depend on our learning setting (e.g., learning material and learning environment) and on our participants. Therefore, our findings may be task-specific, and they probably do not represent general patterns. For different learning settings, metacognitive regulatory processes may look entirely different. Additionally, because the sample group of this study was predominately female, the results may be influenced by gender differences. The question of possible differences between male and female students could be addressed in future studies.

Moreover, our coding scheme, which comprises several metacognitive, cognitive, and motivational activities, determines the level of granularity for the present analysis. Even more importantly, the applied DM and PM techniques are inductive approaches, and thereby the findings rely on the representativeness and quality of the underlying data sources (e.g., Bannert et al., 2014). Therefore, the resulting patterns need to be validated in future research. Furthermore, the process data were measured by concurrent think-aloud protocols. The results should also be replicated using process data on different levels (e.g., log files or eye tracking), ideally advancing towards a temporal alignment of different data channels (Azevedo, 2014), and towards an integration of findings across several SRL studies (Dent & Hoyle, 2015). However, approaches that use log files to track students' cognitive and metacognitive strategies in learning environments imply a more difficult interpretation of awareness and intent regarding students' actions and behaviors than coded think-aloud protocols (e.g., Kinnebrew et al., 2014). Finally, we did not consider the impact of previously presented prompts in detail, but there seems to be no additive effect in our data. Additivity means that it might be possible that the success of a scaffold depends on the effectiveness of previous scaffolds. For example, Molenaar and Roda (2008) recommend considering the history of a learner's interactions with previously provided scaffolds to calibrate the presentation of support. Moreover, without this type of history, the challenge of fading support when the learner's regulative behavior progresses cannot be addressed.

Future directions should address the following issues. In our study, we investigated the learning process that precedes a prompt by analyzing coded learning activities to discover the conditions for effectiveness. However, there may be further conditions that influence the effectiveness of metacognitive prompts that were not considered in the present analysis. For example, the students' level of expertise could be another important factor that affects scaffolding effects (e.g., Yeh et al., 2010). Moreover, questions with regard to the role of the presentation time (e.g., is a prompt that appears during the first minutes of the learning session more effective than a prompt that appears in the middle of the learning time?) and of the number of prompts (e.g., is the first prompt already effective or are two or more prompts needed to affect the regulatory behavior?) still have to be investigated in future research. Furthermore, the findings on possible conditions for the effectiveness of scaffolds should be related to models of SRL and metacognition. For example, the COPES model (Winne & Hadwin, 2008) could be specified by including positions for intervening the SRL cycle. These findings could contribute to a micro-level theory of SRL and to the impact of scaffolds on learning processes that is needed to implement optimally adaptive support in learning environments, and that is currently missing in the field of SRL (Molenaar & Järvelä, 2014). Additionally, more research on the automatic detection of learning processes (e.g., Cocea & Weibelzahl, 2009) is needed to improve the diagnosis of adaptive learning systems. The next step in our future research will be the validation of findings on the effectiveness of metacognitive prompts by analyzing think-aloud data from another prompting experiment. Additionally, we will aim to derive standards and guidelines for the application of DM and PM techniques in SRL settings.

Acknowledgements

This research was supported by funds from the German Science Foundation (DFG: BA 2044/7-1). We would like to thank our student research assistants, Ferdinand Guggenberger, Valentina Haspert, Michaela Hofbauer, and Lisa Müsse, for their assistance with regard to coding the think-aloud data.

Correspondence concerning this manuscript should be sent to Christoph Sonnenberg, University of Wuerzburg, Instructional Media, Oswald-Kuelpe-Weg 82, 97074 Wuerzburg, Germany, email: christoph.sonnenberg@uni-wuerzburg.de; phone: +49931/3189141

References

Aleven, V. (2013). Help seeking and intelligent tutoring systems: Theoretical perspectives and a step towards theoretical integration. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 311–335). New York: Springer. doi:10.1007/978-1-4419-5546-3_21

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205–223. doi:10.1007/s40593-015-0089-1

Azevedo, R. (2005). Using hypermedia as a metacognitive tool for enhancing student learning? The Role of self-regulated learning. *Educational Psychologist*, *40*(4), 199–209. doi:10.1207/s15326985ep4004_2

Azevedo, R. (2009). Theoretical, conceptual, methodological, and instructional issues in research on metacognition and self-regulated learning: A discussion. *Metacognition and Learning*, *4*, 87–95. doi:10.1007/s11409-009-9035-7

Azevedo, R. (2014). Issues in dealing with sequential and temporal characteristics of self- and socially-regulated learning. *Metacognition and Learning*, *9*, 217–228. doi:10.1007/s11409-014-9123-1

Azevedo, R., Cromley, J. G., Moos, D. C., Greene, J. A., & Winters, F. I. (2011). Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Testing and Assessment Modeling*, 53(1), 106–140.

Azevedo, R., Cromley, J. G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344–370. doi:10.1016/j.cedpsych.2003.09.002

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional Science*, *33*, 381–412. doi:10.1007/s11251-005-1273-8

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition - Implications for the design of computer-based scaffolds. *Instructional Science*, *33*, 367–379. doi:10.1007/s11251-005-1272-9

Azevedo, R., Landis, R. S., Feyzi-Behnagh, R., Duffy, M., Trevors, G., Harley, J. M., ... Hossain, G. (2012). The effectiveness of pedagogical agents' prompting and feedback in facilitating co-adapted learning with MetaTutor. *Lecture Notes in Computer Science*, 212–221. doi:10.1007/978-3-642-30950-2_27

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375. doi:10.2190/94V6-R58H-3367-G388

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia. Erfassung, Beschreibung und Vermittlung wirksamer metakognitiver Lernstrategien und Regulationsaktivitäten [Metacognition and hypermedia learning]*. Münster: Waxmann.

Bannert, M. (2009). Promoting self-regulated learning through prompts: A discussion. *Zeitschrift für Pädagogische Psychologie*, *23*(2), 139–145. doi:10.1024/1010-0652.23.2.139

Bannert, M., & Mengelkamp, C. (2013). Scaffolding hypermedia learning through metacognitive prompts. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 171–186). New York: Springer. doi:10.1007/978-1-4419-5546-3_12

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. doi: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. doi: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. doi:10.1016/j.chb.2015.05.038

BEN-ELIYAHU, A., & BERNACKI, M. L. (2015). Addressing complexities in self-regulated learning: A focus on contextual factors, contingencies, and dynamic relations. *Metacognition and Learning*, *10*, 1–13. doi:10.1007/s11409-015-9134-6

BENJAMINI, Y., & HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1), 289-300.

BLOOM, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: D. McKay.

BOUCHET, F., HARLEY, J. M., TREVORS, G. J., & AZEVEDO, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 104–146.

BRUSILOVSKY, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11, 87–110. doi:10.1023/A:1011143116306

BRUSILOVSKY, P. (2007). Adaptive navigation support. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web. Methods and strategies of web personalization* (pp. 263–290). Berlin Heidelberg: Springer. doi:10.1007/978-3-540-72079-9

CHI, M. T. H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences*, *6*(3), 271–315. doi:10.1207/s15327809jls0603_1

CLAREBOUT, G., & ELEN, J. (2006). Tool use in computer-based learning environments: Towards a research framework. *Computers in Human Behavior*, *22*, 389–411. doi:10.1016/j.chb.2004.09.007

CLAREBOUT, G., ELEN, J., COLLAZO, N. A. J., LUST, G., & JIANG, L. (2013). Metacognition and the use of tools. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 187–195). Springer Science. doi: 10.1007/978-1-4419-5546-3_13

COCEA, M., & WEIBELZAHL, S. (2009). *Log file analysis for disengagement detection in e-Learning environments*. *User Modelling and User-Adapted Interaction, 19*, 341–385. doi:10.1007/s11257-009-9065-5

DENT, A. L., & HOYLE, R. H. (2015). A framework for evaluating and enhancing alignment in self-regulated learning research. *Metacognition and Learning, 10*, 165–179. doi:10.1007/s11409-015-9136-4.

ERICSSON, K. A., & SIMON, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, Mass.: MIT Press.

GE, X. (2013). Designing learning technologies to support self-regulation during III-structured problem-solving processes. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 213–228). New York: Springer. doi:10.1007/978-1-4419-5546-3_15

GREENE, J. A., DELLINGER, K. R., TÜYSÜZOGLU, B. B., & COSTA, L.-J. (2013). A two-tiered approach to analyzing self-regulated learning data to inform the design of hypermedia learning environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 117–128). New York: Springer. doi:10.1007/978-1-4419-5546-3_8

GÜNTHER, C. W., & VAN DER AALST, W. M. P. (2007). Fuzzy mining – Adaptive process simplification based on multi-perspective metrics. *Business Process Management - Lecture Notes in Computer Science*, *4714*, 328–343. doi:10.1007/978-3-540-75183-0

HÄMÄLÄINEN, W., & VINNI, M. (2010). Classifiers for educational data mining. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 57–74). Boca Raton, FL: Chapman&Hall/CRC.

HATTIE, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

JEONG, H., GUPTA, A., ROSCOE, R., WAGSTER, J., BISWAS, G., & SCHWARTZ, D. (2008). Using hidden markov models to characterize student behaviors in learning-by-teaching environments. *Lecture Notes in Computer Science*, *5091*, 614–625. doi:10.1007/978-3-540-69132-7_64

JOHNSON, A. M., AZEVEDO, R., & D'MELLO, S. K. (2011). The temporal and dynamic nature of self-regulatory processes during independent and externally assisted hypermedia learning. *Cognition and Instruction*, *29*(4), 471–504. doi:10.1080/07370008.2011.610244

KINNEBREW, J. S., SEGEDY, J. R., & BISWAS, G. (2014). Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and Learning*, *9*, 187–215. doi:10.1007/s11409-014-9112-4

KRAMARSKI, B., & GUTMAN, M. (2006). How can self-regulated learning be supported in mathematical e-learning environments? *Journal of Computer Assisted Learning*, *22*, 24–33. doi:10.1111/j.1365-2729.2006.00157.x

LAJOIE, S. P., NAISMITH, L., POITRAS, E., HONG, Y.-J., CRUZ-PANESSO, I., RANELLUCCI, J., ... WISEMAN, J. (2013). Technology-rich tools to support self-regulated learning and performance in medicine. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 229–242). New York: Springer. doi:10.1007/978-1-4419-5546-3_16

LEHMANN, T., HÄHNLEIN, I., & IFENTHALER, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. *Computers in Human Behavior*, *32*, 313–323. doi:10.1016/j.chb.2013.07.051

LESTER, J. C., MOTT, B. W., ROBISON, J. L., ROWE, J. P., & SHORES, L. R. (2013). Supporting self-regulated science learning in narrative-centered learning environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 471–483). New York: Springer. doi:10.1007/978-1-4419-5546-3_30

LIN, X., HMELO, C., KINZER, C. K., & SECULES, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, *47*(3), 43–62. doi:10.1007/BF02299633

MALMBERG, J., JÄRVELÄ, S., JÄRVENOJA, H., & PANADERO, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, *52*, 562–572. doi:10.1016/j.chb.2015.03.082

MARTIN, T., & SHERIN, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences*, *22*(4), 511–520. doi:10.1080/10508406.2013.840466

MOLENAAR, I., & CHIU, M. M. (2014). Dissecting sequences of regulation and cognition: Statistical discourse analysis of primary school children's collaborative learning. *Metacognition and Learning*, *9*(2), 137–160. doi:10.1007/s11409-013-9105-8

MOLENAAR, I., & JÄRVELÄ, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, *9*(2), 75–85. doi:10.1007/s11409-014-9114-2

MOLENAAR, I., & RODA, C. (2008). Attention management for dynamic and adaptive scaffolding. *Pragmatics & Cognition, 16*(2), 224-271. doi: 10.1075/pc.16.2.04mol

REIMANN, P., MARKAUSKAITE, L., & BANNERT, M. (2014). E-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. doi:10.1111/bjet.12146

REIMANN, P., & YACEF, K. (2013). Using process mining for understanding learning. In R. Luckin, P. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood, & N. Winters (Eds.), *Handbook of Design in Educational Technology* (pp. 472–481). New York: Routledge.

ROLL, I., ALEVEN, V., MCLAREN, B. M., & KOEDINGER, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction*, *21*(2), 267–280. doi:10.1016/j.learninstruc.2010.07.004

SCHOOR, C., & BANNERT, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*, 1321–1331. doi:10.1016/j.chb.2012.02.016

SCHWONKE, R., HAUSER, S., NÜCKLES, M., & RENKL, A. (2006). Enhancing computer-supported writing of learning protocols by adaptive prompts. *Computers in Human Behavior*, *22*, 77–92. doi:10.1016/j.chb.2005.01.002

SITZMANN, T., BELL, B. S., KRAIGER, K., & KANAR, A. M. (2009). A multilevel analysis of the effect of prompting self-regulation in technology-delivered instruction. *Personnel Psychology*, *62*, 697–734. doi:10.1111/j.1744-6570.2009.01155.x

SONG, M., & VAN DER AALST, W.M.P. (2007). Supporting process mining by showing events at a glance. In K. Chari & A. Kumar (Eds.), *Proceedings of 17th Annual Workshop on Information Technologies and Systems (WITS 2007)* (pp. 139–145). Montreal, Canada.

SONNENBERG, C., & BANNERT, M. (2015). Discovering the effects of metacognitive prompts on the sequential structure of SRL-processes using process mining techniques. *Journal of Learning Analytics*, *2*(1), 72–100.

THILLMANN, H., KÜNSTING, J., WIRTH, J., & LEUTNER, D. (2009). Is it merely a question of "what" to prompt or also "when" to prompt? *Zeitschrift für Pädagogische*, *23*(2), 105–115. doi:10.1024/1010-0652.23.2.105

TRČKA, N., PECHENIZKIY, M., & VAN DER AALST, W. M. P. (2010). Process mining from educational data. In C. Romero, S. Ventura, M. Pechenizkiy, & R. Baker (Eds.), *Handbook of Educational Data Mining* (pp. 123–142). Boca Raton, FL: Chapman&Hall/CRC.

VAN DER AALST, W. (2011). *Process mining: Discovery, conformance and enhancement of business processes*. Berlin, Heidelberg: Springer.

VAN DER AALST, W., WEIJTERS, T., & MARUSTER, L. (2004). Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, *16*(9), 1128–1142. doi:10.1109/TKDE.2004.47

VEENMAN, M. (1993). *Metacognitive ability and metacognitive skill: Determinants of discovery learning in computerized learning environments*. University of Amsterdam.

VEENMAN, M. V. J. (2007). The assessment and instruction of self-regulation in computer-based environments: A discussion. *Metacognition and Learning*, *2*(2-3), 177–183. doi:10.1007/s11409-007-9017-6

WALKER, E., RUMMEL, N., & KOEDINGER, K. (2011). Designing automated adaptive support to improve student helping behaviors in a peer tutoring activity. *International Journal of Computer Supported Collaborative Learning, 6*(2), 279-306. doi: 10.1007/s11412-011-9111-2

WINNE, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, *8*(4), 327–353.

WINNE, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, *9*, 229–237. doi:10.1007/s11409-014-9113-3

WINNE, P. H., & BAKER, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 1–8.

WINNE, P. H., & HADWIN, A. F. (2008). The weave of motivation and self-regulated learning. In D. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning: Theory, research, and applications* (pp. 297–314). NY: Taylor & Francis.

WIRTH, J. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie*, *23*(2), 91–94. doi:10.1024/1010-0652.23.2.91

WITTWER, J., & RENKL, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*, 49–64. doi:10.1080/00461520701756420

YEH, Y. F., CHEN, M. C., HUNG, P. H., & HWANG, G. J. (2010). Optimal self-explanation prompt design in dynamic multi-representational learning environments. *Computers and Education*, *54*, 1089–1100. doi:10.1016/j.compedu.2009.10.013

ZIMMERMAN, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. doi:10.3102/0002831207312909