The Internet and Higher Education 49 (2021) 100791

Contents lists available at ScienceDirect

The Internet and Higher Education

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

Investigating prompts for supporting students' self-regulation – A remaining challenge for learning analytics approaches?[☆]

Clara Schumacher[a,*], Dirk Ifenthaler[b,c]

[a] Humboldt-Universität zu Berlin, Department of Computer Science, Unter den Linden 6, 10099 Berlin, Germany
[b] University of Mannheim, Learning Design and Technology, L 4, 1, 68161 Mannheim, Germany
[c] Curtin University, UNESCO Deputy Chair of Data Science in Higher Education, Learning and Teaching, Kent Street, Bentley, WA, Australia

ARTICLE INFO

Keywords:
Prompting
Self-regulated learning
Higher education
Learning analytics

ABSTRACT

To perform successfully in higher education learners are considered to engage in self-regulation. Prompts in digital learning environments aim at activating self-regulation strategies that learners know but do not spontaneously show. To investigate such interventions learning analytics approaches can be applied. This quasi-experimental study ($N = 110$) investigates whether different prompts based on theory of self-regulated learning (e.g., cognitive, metacognitive, motivational) impact declarative knowledge and transfer, perceptions as well as online learning behavior, and whether trace data can inform learning performance. Findings indicate small effects of prompts supporting the performance in a declarative knowledge and transfer test. In addition, the prompted groups showed different online learning behavior than the control group. However, trace data in this study were not capable of sufficiently explaining learning performance in a transfer test. Future research is required to investigate adaptive prompts using trace data in authentic learning settings as well as focusing on learners' reactions to distinct prompts.

1. Introduction

Learning in higher education increasingly takes place in digital learning environments, allowing advanced approaches to capture learner behavior when learning actually occurs. This can be used to support learning and further to reconstruct its processes, thus allowing further insights on students' actions without intrusion (Vieira, Parsons, & Byrd, 2018; Winne & Baker, 2013).

Self-regulated learning is considered to be key for successful learning in higher education and likewise in less structured environments, such as digital learning environments (Azevedo, Cromley, & Seibert, 2004; Bannert & Mengelkamp, 2013; Broadbent & Poon, 2015; Cassidy, 2011; Nussbaumer, Dahn, Kroop, Mikroyannidis, & Albert, 2015). Self-regulated learning is conceptualized as "an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, motivation, and behavior, guided and constrained by their goals and the contextual features in the environment" (Pintrich, 2000, p. 453). However, self-regulating one's learning demands high efforts and skills of learners (Azevedo et al., 2004; Boekaerts, 1999; Schmitz, 2001; Zimmerman, 2000; Lehmann, Hähnlein, & Ifenthaler, 2014). Learners often do not show self-regulatory behavior spontaneously without guidance (Moos & Bonde, 2016; Sonnenberg & Bannert, 2016). Hence, effective means of supporting students' regulation of learning processes and motivation are required, such as the utilization of prompts. Research on prompting focuses on how to design prompts to support self-regulated learning and specifically, on which learning activities should be prompted (Ifenthaler, 2012; Bannert, 2009; Wirth, 2009). It is also relevant to gain insights into how prompts impact learning behavior. Thus, combining means of supporting self-regulated learning with learning analytics approaches could enable a better understanding of learning processes as a prerequisite to design and develop adaptive prompts for digital learning environments.

Research on prompts to support (self-regulated) learning in digital learning environments showed varying findings. For example, in an experimental study investigating the use of prompts based on theory of self-regulated learning in a flipped classroom setting, the learners who received the prompts showed significantly higher learning performance than the control group in a pre-post-test plus they used more self-regulation strategies compared to the control group (Moos & Bonde,

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
* Corresponding author.
E-mail addresses: clara.schumacher@hu-berlin.de (C. Schumacher), dirk@ifenthaler.info (D. Ifenthaler).

https://doi.org/10.1016/j.iheduc.2020.100791
Received 12 July 2019; Received in revised form 12 November 2020; Accepted 7 December 2020
Available online 13 December 2020

2016). An experimental study (Prieger & Bannert, 2018) investigating the impact of metacognitive prompts on learning behavior and learning performance found that students receiving prompts showed significantly different and presumably more systematic navigation patterns within the hypermedia environment than those in the control group but without having an effect on learning performance. Furthermore, the authors found that the effects of prompts are dependent on learner characteristics as participants with lower learning-related competencies profited from metacognitive prompts in terms of their learning behavior and learning performance, whereas students with higher learning-related competencies did not benefit, or were even hampered by the prompts in a hypermedia environment (Prieger & Bannert, 2018). Daumiller and Dresel (2018) found that prompts referring to students' motivation regulation (i.e., strategies used to initiate and persist in learning processes or to raise effort by increasing task value or self-efficacy beliefs (Daumiller & Dresel, 2018)) were effective instructional means, leading to higher task value as part of learning motivation, higher metacognitive control, more task-related learning activities (e.g., cognitive strategies, persistence), and higher learning performance as well as memorization.

While current research identified several potentials of prompts for self-regulated learning, learning analytics approaches have not be fully studied in experimental settings. Accordingly, the focus of this quasi-experimental study is on examining whether cognitive, metacognitive, motivational, or a combination of these plus resource-related prompts affect learning performance, participants' perceptions, as well as online learning behavior, and further, if trace data in a digital learning environment can be used as predictors of learning performance in a knowledge transfer test.

The first section of this paper focuses on prompts to support self-regulated learning (1.1), and on learning analytics approaches for gaining additional insights into learning processes (1.2). Related to the derived hypotheses (1.3), the design of the quasi-experimental study and instruments are described in Section 2. The findings of the study are reported (3), discussed (4) and concluded (5) by pointing out the findings' implications, further research needs, as well as limitations of the study.

1.1. Prompts supporting self-regulated learning

Self-regulated learning is conceptualized as a recursive process in which learners adapt cognitive, metacognitive and motivational processes according to task requirements (Winne, 2017a; Winne & Hadwin, 1998; Zimmerman, 2002). Prompts can be described as "short hints or questions presented to students in order to activate knowledge, strategies, or skills that students have already available but do not use spontaneously" (Wirth, 2009, p. 92). Prompts are a non-directive external support, not providing new information but stimulating the application of known cognitive, metacognitive, motivational or resource management-related strategies during learning (Bannert, 2009). Thus, instructional support on self-regulated learning, such as prompts, should be aligned with learners' strategy knowledge (Thillmann, Künsting, Wirth, & Leutner, 2009). In general, prompts guide learners to reflect on specific aspects of the learning material/task or on their cognitive activities during the learning process, and might further ask them to express these thoughts (Bannert, 2009). Prompts can be designed as questions, incomplete sentences or instructions (Ifenthaler, 2012; Kramarski & Kohen, 2017).

Wirth (2009) proposes a framework to classify prompts according to their (a) *content*: the activities that should be stimulated through prompts (e.g., cognitive or metacognitive learning strategies), (b) the *condition* that must be fulfilled in order that the prompt is presented to the learners: a certain amount of *time*; related to the *task* or based on *previous activities*, and (c) the *method* used for presenting the prompt: *feed forward* prompts – directly referring to the upcoming activities learners are expected to perform – or *feedback* prompts – an indirect method of guiding learners through feedback based on their previous behavior.

Referring to the concept of self-regulated learning and learning strategies (Boekaerts, 1999; Weinstein & Mayer, 1986), c*ognitive prompts* aim to support students' information processing, whereas *metacognitive prompts* focus on activating students' monitoring and controlling of their cognitive activities, such as planning, goal-setting, and evaluating their learning processes and outcomes. Furthermore, *motivational prompts* aspire to enhance motivation to learn, by highlighting targets or giving hints on how to regulate one's motivation, and *resource-related prompts* aim to support students in setting up a supportive learning environment or initiating help-seeking behavior.

Prompts need to be aligned with learning theory and instructional intentions (Moos & Bonde, 2016) and presented at the time the learner needs the support in order to avoid additional cognitive processing (Thillmann et al., 2009). Sonnenberg and Bannert (2016) propose using process data to develop effective instructional means. In addition, using trace data of learners allows insights into their behavior and strategy use after receiving an intervention, such as a prompt (Thillmann et al., 2009; Winne & Baker, 2013). Furthermore, Prieger and Bannert (2018) argue that fixed prompts, which are pre-defined in terms of timing and content, might interrupt the learning process. For example, Backhaus, Jeske, Pointstingl, and Koenig (2017) presented different prompts to the students related to their self-reported learner characteristics such as work effort and strategy use. However, in their study only an assessment prompt, which asked the participants to assess their own progress, significantly improved learning performance in comparison to the control group. But this study did not include trace data for presenting the prompts or understanding the learning processes. However, to provide adaptive support further evidence on the relation of trace data plus learner characteristics and learning performance are required.

From a methodological point of view, current research on prompts in higher education predominantly investigate learning processes by using think-aloud protocols sometimes enhanced with screen recording (e.g., Bannert, Sonnenberg, Mengelkamp, & Prieger, 2015; Engelmann & Bannert, 2019; Moos & Bonde, 2016; Sonnenberg & Bannert, 2016). To date, the use of trace data from digital learning environments for investigating prompts is realized in a few studies (e.g., Bannert et al., 2015; Müller & Seufert, 2018; Prieger & Bannert, 2018). As outlined in the introduction, findings on the impact of prompting on learning performance are ambiguous which is also the case for learning behavior or strategy use (e.g., Engelmann & Bannert, 2019; Moos & Bonde, 2016; Müller & Seufert, 2018; Prieger & Bannert, 2018). Thus, referring to the current state of research on prompts supporting self-regulated learning in digital learning environments further empirical evidence is required on how different prompts impact learning performance and online learning behavior as well as how the trace data can explain and inform learning performance. As think-aloud methods are not applicable for authentic learning scenarios it needs to be further investigated if trace data can offer sufficient additional insights into learning processes to serve as a basis for providing support through adaptive prompts in advanced digital learning environments.

1.2. Using learning analytics approaches for understanding learning processes

Learning analytics offer a promising approach for digital and adaptive learning environments (Aguilar, 2018; Greller & Drachsler, 2012; Ifenthaler & Widanapathirana, 2014). Therefore, learning analytics use static and dynamic information about learners and learning environments, assessing, eliciting, and analyzing them for real-time modeling, prediction, and optimization of learning processes, learning environments, and educational decision-making (Ifenthaler, 2015). The aim is to better meet students' needs by offering individual learning paths, adaptive assessments and recommendations, or adaptive and just-in-time feedback (Corrin & de Barba, 2014; Gašević, Dawson, & Siemens, 2015; McLoughlin & Lee, 2010) according to learners' motivational states, individual characteristics, and learning goals. However, a better understanding is required of how learning processes are related to and can be captured via data available in current digital learning environments (Greller & Drachsler, 2012; Wilson, Watson, Thompson, Drew, & Doyle, 2017). Therefore, learning analytics approaches might be suitable as they enable additional insights into online learning behavior without being intrusive (Vieira et al., 2018; Winne, 2017b).

Current learning analytics approaches focus on indicators based on the behavior in the digital learning environment, such as time spent online, access to various types of resources, or reading and writing posts to relate them to learning performance (Mah, 2016). In addition, few other approaches are enriched with learner characteristics such as demographic data or results of assessments, to predict study success (Costa, Fonseca, Santana, de Araújo, & Rego, 2017; Vieira et al., 2018). In a literature review focusing on visual learning analytics, Vieira et al. (2018) found that most studies analyze usage of resources in particular, with only a few studies having a processual approach by trying to understand learning paths or students' learning progress. For learning analytics to understand and ideally support self-regulated learning, Winne (2017b) proposes that: (a) every operation during learning is tracked; (b) the information operated on by a learner is identifiable; (c) the traces are time-stamped; and (d) the results of the operations are recorded.

However, not all collected indicators are (pedagogically) valid and learning analytics only have a limited insight into students' learning as not all learning processes take place in the digital learning environment or can be captured with trace data (Ifenthaler & Schumacher, 2016; Eradze, Väljataga, & Laanpere, 2014; Ferguson, 2012; Wilson et al., 2017; Winne, 2017b). Thus, this study applies a quasi-experimental design controlling for external learning behavior and uses trace data to gain additional insights into online learning behavior and how this relates to learning performance.

1.3. Purpose of the study and hypotheses

It is suggested that prompts are capable of supporting learners by providing them with additional hints to apply relevant learning strategies (Bannert, 2009). Trace data are considered to provide further insights into learners' behavior (Winne & Baker, 2013). Consequently, this quasi-experimental study focuses on (a) investigating how prompts impact learning performance, (b) learning behavior, and (c) perceptions, and (d) if online learning behavior enables an understanding of learning performance.

The assumption that using cognitive, metacognitive, motivational and resource-related strategies is associated with successful learning processes and thus better performance (Pintrich, 2000; Weinstein & Mayer, 1986; Zimmerman, 2001) guided our first two hypotheses.

Hypothesis 1. It is assumed that learners in different prompting conditions vary regarding their learning performance in a *knowledge test* (Hypothesis 1a) and especially *over time* (Hypothesis 1b).

Hypothesis 2. It is furthermore assumed that learners in different prompting conditions vary regarding their learning performance in a *knowledge transfer test* (Hypothesis 2a), and especially *over time* (Hypothesis 2b).

As Prieger and Bannert (2018) argue that predefined prompts might impact learning processes through interruptions participants' evaluation of the prompts with regard to perceived learning support, usefulness, and negative perceptions was investigated.

Hypothesis 3. It is assumed that the participants in the three prompting conditions rated the prompts differently with regard to *perceived learning support* (Hypothesis 3a), *perceived usefulness* (Hypothesis 3b), and *negative perceptions* (Hypothesis 3c).

Prior studies found that prompts affected learners' navigation patterns within digital learning environments (Bannert et al., 2015; Prieger & Bannert, 2018), and that trace data would enable insights into this behavior (Winne & Baker, 2013).

Hypothesis 4. It is hypothesized that the different prompting conditions and the control group differ with regard to their *behavior in the digital learning environment* as indicated by trace data (e.g., views of handout, additional learning material and videos, and their overall interaction; Hypothesis 4a) and that the experimental groups differ regarding the *length of the notes taken* (Hypothesis 4b).

Furthermore, students' prerequisites such as prior knowledge, motivation, and perceptions lead to differences in their learning behavior and outcomes, thus such information can be considered for learning analytics analyses (Ifenthaler & Widanapathirana, 2014; Clow, 2013; Nadasen & List, 2017).

Hypothesis 5. It is hypothesized that *academic characteristics*, such as semester, current study grade, prior domain knowledge, perceived confidence and difficulty (Hypothesis 5a), and the *online learning behavior* as indicated by number of views of the handout, additional learning material, video and overall interaction (Hypothesis 5b) significantly predict participants' *learning performance in a knowledge transfer test*.

Hence, the purpose of this quasi-experimental study is examining if different prompts have an impact on learning performance and online learning behavior. In addition, learners' perceptions of the prompts with regard to learning support, usefulness, and negative perceptions are investigated. Furthermore, to use learning analytics approaches to offer adaptive support trace data need to provide valid insights into learning processes and outcomes. Thus, the study investigates whether trace data can be used for predicting learning performance when controlling for learning behavior outside the digital learning environment.

2. Methods

2.1. Participants

Initially 135 students from a European university participated in the study. After deleting incomplete or discontinued data sets, a total of N = 110 (74 female, 36 male) remained and were used for the hypothesis testing. Participants' average age was 22.68 years (SD = 2.82). They were enrolled in either the Bachelor's (65.5%) or Master's (34.5%) program of economic and business education. The participants had studied for an average of 4.86 semesters (SD = 2.91). The participants received two credits for participating in the study. Consent for tracking experimental and trace data was given by each individual participant.

2.2. Design

In the university's digital learning environment, a laboratory environment consisting of four classes was implemented. Participants were assigned to the four experimental conditions based on their date of participation (see Fig. 1 for details). The experimental conditions were assigned to the components of self-regulated learning (Boekaerts, 1992, 1999; Pintrich, 2000): *cognitive* (CP; $n_1 = 30$), *metacognitive* (MP; $n_2 = 31$), *cognitive, metacognitive, motivational and resource-related* (AP; $n_3 = 28$), and *control group* (CG; $n_4 = 21$). Participants in the *CP group* received prompts related to cognitive learning strategies (see materials for further details). Participants in the *MP group* received prompts related to metacognitive learning strategies. Participants in the *AP group* received prompts related to all learning strategies self-regulated learners are assumed to perform: cognitive, metacognitive, motivational and resource-related. The *control group* did not receive prompts. As indicated by the QQ-plots the data were distributed normally, and Levene's test revealed homoscedasticity of the data ($p > .005$). ANOVA was used to test the four groups for differences in terms of pre-knowledge, study

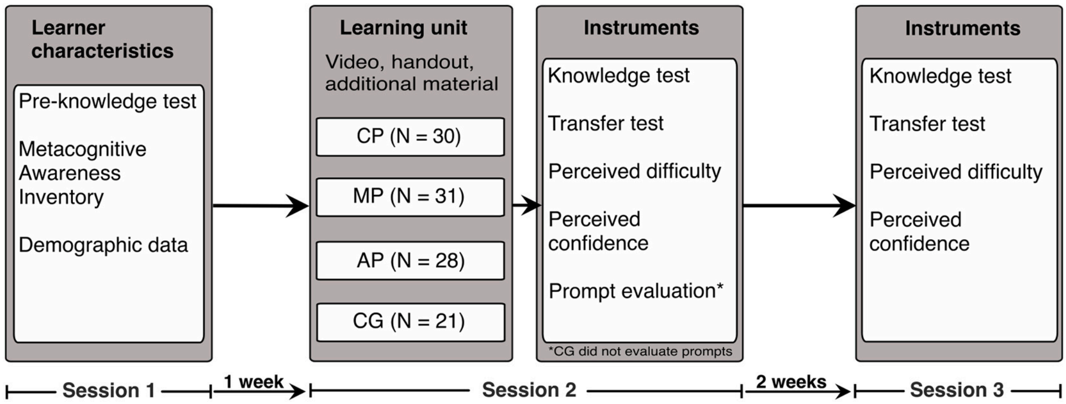

Fig. 1. Overview about the study design

program, age, and study grade. ANOVA revealed that the groups did not differ with regard to their pre-knowledge related to the learning content $F(3,106) = 1.527$, $p = .212$, study program $F(3,106) = .273$ $p = .845$, age $F(3,106) = 1.241$, $p = .299$, and study grade $F(3,97) = 1.568$, $p = .202$.

2.3. Materials and instruments

2.3.1. Learning unit

Participants navigated through a learning unit in the digital learning environment of the university. The set-up was comparable with online lectures, such as in flipped or blended classroom settings (see Fig. 2). Students entered the course and were presented a marketing lecture of a value-based management course. The course folder contained the corresponding video lecture, the related handout and material with additional information. The video lecture showed the lecturer and relevant visualizations. The duration of the lecture video was 13:29 min.

2.3.2. Cognitive, metacognitive, motivational and resource-related prompts

Based on self-regulated learning theory (Boekaerts, 1992, 1999; Pintrich, 2000; Zimmerman, 2002) prompts were designed as shown in Table 1. The prompts were either embedded in the digital learning environment interface or during the videos. Referring to Wirth's (2009) framework on prompts the *content* consisted of the components of self-regulated learning (cognitive, metacognitive, motivational, resource-related); the *condition* under which a prompt was presented was either based on a previous activity (navigation decision, viewed content), a certain point of time in the video or learning period or related to the task; the *method* of the prompts used was feedforward as the prompts referred to behavior the participants were expected to show in the future. The prompts were shown in form of a pop-up window as an overlay in the digital learning environment (see Fig. 2), with some showing optional text boxes or answers on a rating scale. Depending on the experimental group participants were facing a different number of prompts embedded in the learning unit, the *CP group* received four prompts, the *MP group* received five prompts, and the *AP group* received six prompts plus one in the middle of the study. However, it needs to be noted that learners did not have encountered all prompts embedded as they only had limited time for learning and received most prompts only when showing the required navigation or learning behavior (e.g.,

Table 1
Sample prompts for each prompting condition including time of presentation.

Prompts	Time of presentation
Cognitive prompts	
Take notes on the content of the video. Write down your notes in the comment field.	When opening the video
Think about the concepts presented and if they appear to be coherent and reasonable. Write down your critical thoughts and questions.	During the video
Metacognitive prompts	
Stop and reflect how well you understood the content so far. If you have any difficulties revise the corresponding passage.	After 20 min
Reflect on the main contents of the video. Write down your thoughts.	After the video was finished
All prompts	
For this learning unit you have about 25 min, the video takes about 14 min. Please allot your time accordingly.	When opening the learning unit
Try to focus your attention on the content.	During the video
There are additional resources in the digital learning environment you can use to better prepare yourself for the final test.	After video was finished

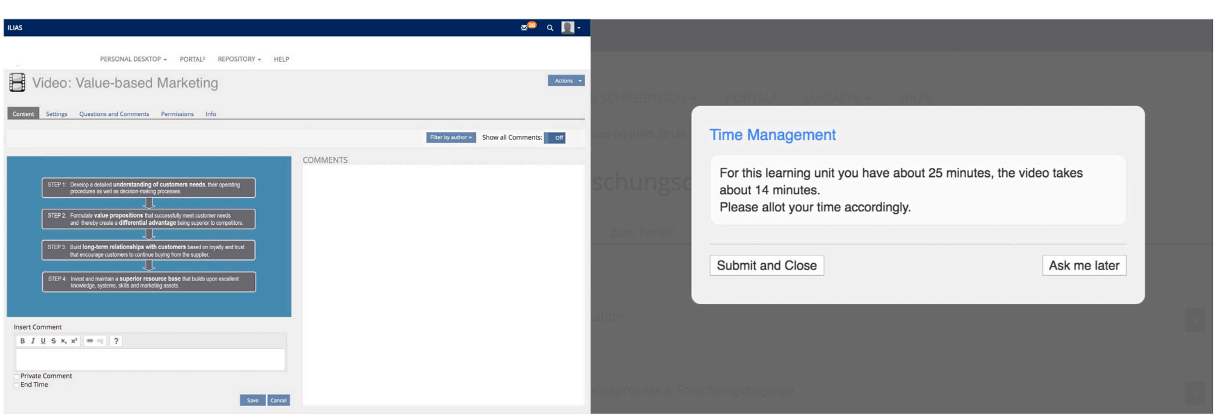

Fig. 2. Digital learning environment and sample prompt

accessing a resource connected to a prompt, watching the video to a certain point of time). After the learning period of 25 min, all groups, including the control group, were guided by a prompt to come to an end and to proceed to the survey.

2.3.3. Knowledge test

To measure participants' declarative knowledge that is referring to factual knowledge (Schunk, 2012) a knowledge test was administered. The pre-knowledge test consisted of four single-choice questions related to the upcoming learning content. At the point of measurement two, the initial questions were used again and supplemented with four additional single-choice questions. The same 8 single-choice questions were used for the third measurement point. A sample question was "Which general rule should a company follow in order to establish a substantial differential advantage? – Focus on one specific dimension of value and provide customers the best available offering with regard to this particular dimension." For each correct single-choice question, one point was scored. For analyses, the overall knowledge test results were used percentagewise.

To assess participants' knowledge transfer that refers to applying (declarative) knowledge in new ways, to other problems or contexts (Schunk, 2012) a writing assignment was administered at t_2 (350 words expected, max. 3 points) and again at t_3 (250 words expected, max. 3 points). A sample task was "Please describe in your own words how value-based marketing is characterized and how it can be realized in a company. Please refer to constructs of the learning material.". In addition, at t_2 and t_3, participants rated the perceived difficulty of the learning unit and how well prepared they felt for the upcoming assessment.

Two independent raters scored the responses to the transfer tasks which participants answered at measurement points t_2 and t_3. Points were assigned based on the quality of responses (0 = not sufficiently described, 1 = only a short description or with significant mistakes, 2 = a good description of the concepts, 3 = a very good description, supplemented with additional examples). In case of non-uniformly rated transfer tasks, the two raters discussed the scoring and either adjusted or kept their score. Among the two raters an interrater reliability of $K = .94$ for the *transfer test* at t_2, and $K = .97$ for the *transfer test* at t_3 was found.

2.3.4. Learner characteristics

Learner characteristics include *personal* characteristics about learners such as age, gender, socio-demographic information, *academic* characteristics such as prior knowledge, learning goals, learning strategies, *social/emotional* characteristics referring to group dynamics or individual emotions (e.g., self-efficacy, motivation), and *cognitive* characteristic such as mental procedures or attention span (Drachsler & Kirschner, 2012).

To investigate participants' metacognitive awareness, the *Metacognitive Awareness Inventory* (Schraw & Dennison, 1994), containing 52 items answered on a Thurstone scale (0 = no; 1 = yes) was used. The two dimensions of the inventory include 1) *knowledge about cognition* (17 items, Cronbach's $\alpha = .644$), and 2) *regulation of cognition* (35 items, Cronbach's $\alpha = .800$). Knowledge about cognition refers to declarative knowledge, procedural knowledge and conditional knowledge. Regulation of cognition includes planning, information management, comprehension monitoring, debugging strategies and evaluation.

Participants further stated demographic information such as age, study program (Bachelor's or Master's program), semester, course load, current study grade (current GPA), etc.

2.3.5. Rating of prompts

Participants who received prompts rated them by answering 14 items including three subscales: evaluation of *perceived learning support* through the prompts (5 items, Cronbach's $\alpha = .836$), *perceived usefulness* of the prompts (5 items, Cronbach's $\alpha = .851$), and *negative perceptions* associated with the prompts (4 items, Cronbach's $\alpha = .914$). Sample items to investigate learning support were: "The prompts encouraged me for reflection" or "The prompts supported my learning processes". To assess the perceived usefulness participants were for example asked to evaluate: "The prompts have increased my effectivity." Sample items investigating if learners perceived the prompts negatively such as distracting or too often were "I perceived the prompts as disturbing" or "Prompts were too often". All items were answered on a 5-point Likert scale with 1 = "I do not agree at all" and 5 = "I fully agree". Hence, high numbers in perceived learning support, and usefulness would indicate that learners perceive high learning support or usefulness whereas high numbers in negative perceptions would indicate that learners evaluated the prompts highly negative.

2.3.6. Trace data

While interacting with the digital learning environment, participants' navigation was tracked. For this research paper the following indicators were used: interaction with the digital learning environment indicated by number of views of resources (handout, additional learning material, video views, overall interaction), and the number and length of written notes taken during the learning unit.

2.4. Procedure

The participants were assigned to the four experimental groups. The study consisted of three measurement points: t_1 as an on-site investigation, t_2 took place on-site and one week later, and t_3 was implemented as an online investigation accessible for one week, two weeks after t_2 occurred. At t_1 participants received an introduction and completed a *pre-knowledge test* (4 single-choice questions; 6 min), the *metacognitive awareness inventory* (52 items; 12 min), and *demographic data* (14 items; 5 min). At t_2 participants completed a *learning unit* in the domain of marketing (25 min). The learning unit consisted of a video lecture (13:29 min), the related handout and additional material. Participants were instructed to prepare themselves for a subsequent knowledge test with the material provided. Hereafter, a *knowledge test* followed that included the questions of t_1 and additional four questions (8 single-choice questions; 10 min). Furthermore, participants had to pass a *transfer task* related to the learning unit (15 min). In addition, the participants rated the *perceived difficulty* of the learning content and their *confidence* (5 items; 2 min) plus if being in an experimental condition *rated the prompts* they received (14 items, 5 min). In t_3 participants again completed the *knowledge test* used in t_2 (8 single-choice questions; 10 min) and answered a *transfer task* related to the learning material (10 min) as well as reporting the *perceived difficulty* and *their confidence* (5 items; 2 min).

3. Results

An alpha level of .05 was used for statistical tests and partial η^2 (small effect: $\eta^2 < .06$, medium effect $.06 \leq \eta^2 \leq .13$, strong effect $\eta^2 > .13$).

Hypothesis 1. Declarative knowledge

A mixed analysis of variance (ANOVA) was computed with dependent variable declarative knowledge, within-subject factor time (t_1, t_2, t_3) and the experimental conditions of the prompting groups (CP, MP, AP, CG) as between-subject factor (see Table 2 for descriptive statistics and Fig. 3). Mauchly's test indicated that the assumption of sphericity has been violated, χ^2 (2) = 6.298, $p = .043$. Results of Levene's test indicated that homogeneity of the error variances was given for the variables ($p > .005$). Box's test further revealed homogeneity of covariance metrices ($p = .500$). ANOVA with Greenhouse-Geisser correction showed a significant within-subject effect for time, $F(1.89,200.34) = 105.177$, $p < .001$, $\eta^2 = .498$ but no interaction effect of time and the experimental conditions $F(5.67,200.34) = .991$, $p = .430$, $\eta^2 = .027$.

Table 2
Descriptive statistics for declarative knowledge (percentagewise), knowledge transfer, perceived confidence and perceived difficulty of the learning unit at each measurement point (t).

t	Variables	CP (*N* = 30)		MP (*N* = 31)		AP (*N* = 28)		CG (*N* = 21)	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1	Declarative knowledge	32.50	18.74	34.67	15.38	30.35	15.74	25.00	15.81
2	Declarative knowledge	53.33	18,25	65.32	16.68	54.91	17.78	52.38	20.00
	Knowledge transfer	.90	.66	.97	.84	.86	.97	1.24	.83
	Confidence	2.47	.86	2.74	.96	2.46	.96	2.29	.85
	Difficulty	2.77	.90	2.74	.89	2.96	.69	3.33	.73
3	Declarative knowledge	51.25	17.78	55.24	16.79	50.44	13.81	50.59	19.95
	Knowledge transfer	.63	.72	1.03	.84	.57	.84	.48	.93
	Confidence	2.13	.82	2.58	.99	2.54	.84	2.24	1.14
	Difficulty	2.73	.78	2.97	.80	3.04	.69	2.67	1.11

Note: CP = cognitive prompt group, MP = metacognitive prompt group, AP = all prompt group, CG = control group, declarative knowledge (percentage-wise), knowledge transfer (measured 0 to 3), perceived confidence (confidence per measurement point, 5-point Likert scale), and perceived difficulty (difficulty per measurement point, 5-point Likert scale).

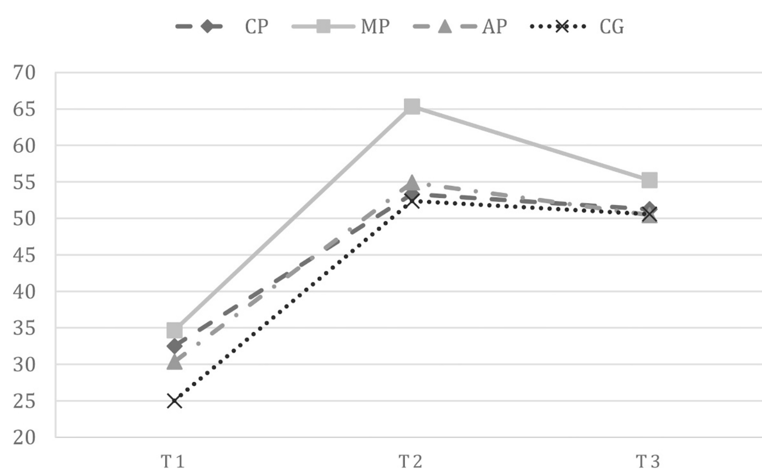

Fig. 3. Declarative knowledge for each experimental condition over time.

Pairwise comparisons using Bonferroni correction showed significant differences between measurement point t_1 ($M = 31.13$; $SD = 16.65$) and t_2 ($M = 56.93$; $SD = 18.59$), $p < .001$, between t_1 ($M = 31.13$; $SD = 16.65$) and t_3 ($M = 52.04$; $SD = 16.89$), $p < .001$ as well as between t_2 ($M = 56.93$; $SD = 18.59$) and t_3 ($M = 52.04$; $SD = 16.89$), $p = .019$.

To check for differences between the experimental conditions post hoc univariate ANOVA were conducted for each measurement point. As indicated before, no significant differences for pre-knowledge at t_1 between the experimental conditions was found $F(3,106) = 1.527$, $p = .212$, $\eta^2 = .041$. Furthermore, no significant differences between the experimental conditions were found for t_3 $F(3,106) = .520$, $p = .669$, $\eta^2 = .014$. Significant differences between the groups were found for t_2 $F(3,106) = 3.190$, $p = .027$, $\eta^2 = .083$. However, Tukey post-hoc tests did not reveal significant differences between the groups. Only the differences between *MP group* ($M = 65.32$; $SD = 16.68$) and *CP group* ($M = 53.33$; $SD = 18.52$), $p = .052$ as well as *MP group* ($M = 65.32$; $SD = 16.68$) and *CG* ($M = 52.83$; $SD = 20.00$), $p = .060$ were slightly above significance level.

To test for changes over time for each group, repeated-measures ANOVA was used. Significant difference in terms of declarative knowledge over time were found for the *CP group* $F(2,58) = 22.868$, $p < .001$, $\eta^2 = .441$. Post-hoc comparisons using Bonferroni correction revealed significant differences between t_1 ($M = 32.50$; $SD = 18.74$) and t_2 ($M = 53.33$; $SD = 18.25$) $p < .001$, between t_1 ($M = 32.50$; $SD = 18.74$) and t_3 ($M = 51.25$; $SD = 17.78$) $p < .001$, but not between t_2 ($M = 53.33$; $SD = 18.25$) and t_3 ($M = 51.25$; $SD = 17.78$) $p > .99$. For the *MP group* significant differences regarding declarative knowledge over time were found $F(2,60) = 32.798$, $p < .001$, $\eta^2 = .522$. Post-hoc comparisons using Bonferroni correction revealed significant differences between t_1 ($M = 34.67$; $SD = 15.38$) and t_2 ($M = 65.32$; $SD = 16.68$) $p < .001$, between t_1 ($M = 34.67$; $SD = 15.38$) and t_3 ($M = 55.24$; $SD = 16.70$) $p < .001$, as well as between t_2 ($M = 65.32$; $SD = 16.68$) and t_3 ($M = 55.24$; $SD = 16.70$) $p = .011$. For *AP group* significant changes of declarative knowledge over time were found $F(2,54) = 26.979$, $p < .001$, $\eta^2 = .500$. Post-hoc comparisons using Bonferroni correction showed significant differences between t_1 ($M = 30.35$; $SD = 15.74$) and t_2 ($M = 54.91$; $SD = 17.78$) $p < .001$, between t_1 ($M = 30.35$; $SD = 15.74$) and t_3 ($M = 50.44$; $SD = 13.81$) $p < .001$, but not for t_2 ($M = 54.91$; $SD = 17.78$) and t_3 ($M = 50.44$; $SD = 13.81$) $p = .509$. For the *control group* significant differences of declarative knowledge over time were found $F(2,40) = 25.442$, $p < .001$, $\eta^2 = .560$. Post-hoc comparisons using Bonferroni correction showed significant differences between t_1 ($M = 25.00$; $SD = 15.81$) and t_2 ($M = 52.38$; $SD = 20.00$) $p < .001$, between t_1 ($M = 25.00$; $SD = 15.81$) and t_3 ($M = 50.59$; $SD = 19.95$) $p < .001$, but not between t_2 ($M = 52.38$; $SD = 20.00$) and t_3 ($M = 50.59$; $SD = 19.95$) $p > .99$.

Given these findings for declarative knowledge, Hypothesis 1a is rejected for t_1 and t_3 and partly accepted for t_2. Hypothesis 1b is accepted for *MP group,* and partly accepted for *CP group, AP group*, and *CG.*

Hypothesis 2. Knowledge transfer test

A mixed ANOVA was computed with dependent variable knowledge transfer test result, within-subject factor time (t_2 and t_3), and the experimental conditions of the prompting groups (CP, MP, AP, CG) as between-subject factor (see Table 2 for descriptive results and Fig. 4). Results of Levene's test indicated that homogeneity of the error variances was given for the variables ($p > .005$). Box's test further revealed homogeneity of covariance metrices ($p = .191$). ANOVA showed a significant interaction effect for time and group $F(3,106) = 3.242$, $p = .025$, $\eta^2 = .084$, and a significant within-subject effect for time $F(1,106) = 9.299$, $p = .003$, $\eta^2 = .081$.

Post-hoc univariate ANOVA was used to investigate differences between the groups at each measurement point. However, no significant differences between the groups were found for knowledge transfer at t_2 $F(3,106) = .975$, $p = .407$, $\eta^2 = .026$, and t_3 $F(3,106) = 2.487$, $p = .065$, $\eta^2 = .066$.

For changes over time on a group level, paired *t*-tests were applied for each group. For the *CP group* significant difference was found for knowledge transfer at t_2 ($M = .90$; $SD = .662$) and t_3 ($M = .63$; $SD = .718$), $t(29) = 2.28$, $p = .030$. For the *MP group* no significant differences for knowledge transfer between t_2 ($M = .97$; $SD = .836$) and t_3 ($M = 1.03$; $SD = .836$), $t(30) = -.338$, $p = .738$. For the *AP group* no significant differences were found for knowledge transfer between t_2 ($M = .86$; $SD = .970$) and t_3 ($M = .57$; $SD = .836$), $t(27) = 1.68$, $p = .103$. For the *control group* significant differences were found for knowledge transfer between t_2 ($M = 1.24$; $SD = 831$) and t_3 ($M = .48$; $SD = .928$), $t(20) = 3.07$, $p = .006$.

Based on these findings for knowledge transfer, Hypothesis 2a is rejected, and Hypothesis 2b is accepted for *CP group* and for the *control group* but not for the *MP* and *AP group*.

Hypothesis 3. Differences in perceptions of the prompts

To test whether the experimental groups differ with regard to their perceptions of the prompts with regard to perceived learning support, perceived usefulness, and negative perceptions multivariate analysis of variance (MANOVA) was used (see Table 3 for descriptive statistics). As indicated by Levene's test the homogeneity of the error variances was met for all variables ($p > .005$). Box's test revealed homogeneity of covariances ($p = .620$). Multivariate analyses revealed significant differences between the groups Wilk's Lambda = .828 $F(6,168) = 2.772$, $p = .014$, $\eta^2 = .090$. Post-hoc univariate ANOVA revealed significant differences between the groups for *perceived learning support* $F(2,86) = 3.843$, $p = .025$, $\eta^2 = .082$, and *negative perceptions of the prompts* $F(2,86) = 3.525$, $p = .034$, $\eta^2 = .076$. However, no significant differences between the groups were found for *perceived usefulness* $F(2,86) = 1.626$, $p = .203$, $\eta^2 = .036$. Post-hoc comparisons using Bonferroni correction showed significant differences between the *AP group* ($M = 2.71$, $SD = .815$) and the *MP group* ($M = 3.29$, $SD = .846$), $p = .025$ with regard to *perceived learning support*. Regarding the *negative perceptions of the prompts* significant differences between the *AP group* ($M = 2.07$, $SD = .863$) and the *CP group* ($M = 1.41$, $SD = .893$), $p = .029$ were found.

Based on these findings, Hypothesis 3a is accepted for *AP* and *MP group*, Hypothesis 3b is rejected, and Hypothesis 3c is accepted for *AP and CP group.*

Table 3
Descriptive statistics for perceptions of the prompts for the experimental groups.

Variables	CP ($N = 30$)		MP ($N = 31$)		AP ($N = 28$)	
	M	*SD*	*M*	*SD*	*M*	*SD*
1) Perceived learning support	2.90	.81	3.29	.84	2.71	.81
2) Perceived usefulness	2.84	.90	3.00	.95	2.58	.82
3) Negative perceptions	1.41	.89	1.69	1.04	2.07	.86

Hypothesis 4. Differences in trace data for the prompting conditions

To determine whether the different experimental conditions vary regarding their online behavior within the learning unit (views of handout and additional material, views of the video, and overall interaction) MANOVA was used (see Table 4 for descriptive statistics). Results indicate that there are significant differences between the groups Wilk's Lambda = .665 $F(12,272.80) = 3.794$, $p < .001$, $\eta^2 = .127$. ANOVA revealed significant differences between the groups for *views of the handout*, $F(3,106) = 3.084$, $p = .032$, $\eta^2 = .079$, and for *views of the additional learning material* $F(3,106) = 8.418$, $p < .001$, $\eta^2 = .192$. No significant differences were found for *video views* $F(3,106) = 1.097$, $p = .354$, $\eta^2 = .030$ and for the *overall interaction* in the learning unit $F(3,106) = 2.117$, $p = .102$, $\eta^2 = .057$. Post-hoc comparisons using Bonferroni correction showed significant differences for views of the handout between the *AP group* ($M = .50$, $SD = .694$) and *CG* ($M = 1.05$, $SD = .384$), $p = .034$. With regard to views of the additional learning material significant differences were found between *AP group* ($M = .18$, $SD = .390$) and *CG* ($M = .90$, $SD = .436$), $p < .001$, between *CP group* ($M = .43$, $SD = .568$) and *CG* ($M = .90$, $SD = .436$), $p = .008$, and between *MP group* ($M = .52$, $SD = .570$) and *CG* ($M = .90$, $SD = .436$), $p = .045$.

As participants in the control group, not receiving any prompts, did not take notes during the learning unit, further analyses were conducted to test for differences between the experimental groups with regard to the length of notes taken within the learning unit. ANOVA revealed significant differences for the length of notes taken $F(2,86) = 3.126$, $p = .049$, $\eta^2 = .068$. Post-hoc comparisons using Bonferroni corrections showed significant differences for the length of notes taken between *AP*

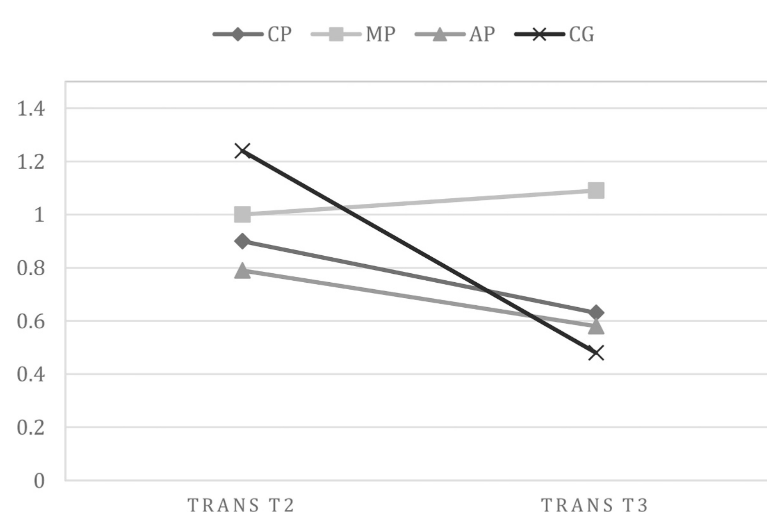

Fig. 4. Knowledge transfer test result for each experimental condition over time.

Table 4
Descriptive statistics for views of handout, additional material, videos, overall interaction, number and length of notes taken.

Variables	CP ($N = 30$)		MP ($N = 31$)		AP ($N = 28$)		CG ($N = 21$)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
1) Number of views of handout	.70	.75	.87	.72	.50	.69	1.05	.38
2) Number of views of additional material	.43	.57	.52	.57	.18	.39	.90	.43
3) Number of views of video	1.30	.95	1.74	2.21	1.29	.98	1.10	.30
4) Overall interaction	6.87	4.14	8.71	4.85	8.61	5.65	6.14	1.49
5) Number of notes taken	2.07	2.64	1.10	2.36	3.46	4.67	.00	.00
6) Length of notes taken	310.10	362.22	171.42	378.03	491.89	686.92	.00	.00

group ($M = 491.89$, $SD = 686.92$) and *MP group* ($M = 171.42$, $SD = 378.03$), $p = .043$, but not for the *CP group* ($M = 310.10$, $SD = 362.22$) and the other groups.

Hence, Hypothesis 4a is accepted for the *number of views of the handout* and the *additional material*, and rejected for *number of views of the video* and the *overall interaction* within the learning unit. With regard to the *length of notes* taken Hypothesis 4b is accepted for *AP and MP group*.

Hypothesis 5. Predicting knowledge transfer test result

Table 5 shows the descriptive statistics and zero-order correlations of the predictors used in the regression analysis. To investigate whether (a) *students' study related characteristics* (semester, current study grade, prior knowledge, perceived difficulty and confidence of the learning unit) and (b) *their online learning behavior* (views of handout, additional material, video, and overall interaction) could significantly predict their *learning performance in the knowledge transfer test*, linear regression analysis (see Table 6) was used, yielding a ΔR^2 of .334 $F(9,91) = 6.571$, $p < .001$. With regard to *academic characteristics* participants' *semester* was a significant positive predictor, current *study grade* was negatively predicting, and their *perceived confidence* was a positive predictor of participants' knowledge transfer test result. Regarding *trace data* only participants' number of *views of the handout* was a significant positive predictor of learning performance.

Based on these results, Hypothesis 5a is accepted for *semester*, *current study grade*, and *perceived confidence* and rejected for *prior knowledge* and *perceived difficulty* of the learning unit. Further, Hypothesis 5b is only accepted for *number of views of the handout* and rejected for *number of views of the additional material* and the *video* as well as the *overall interaction* within the learning unit.

4. Discussion

The purpose of this study was to investigate (a) if different prompting conditions had an impact on learning performance in declarative knowledge and knowledge transfer tests, (b) if the prompts entail different learning behavior and (c) perceptions, and (d) if trace data can inform learning performance. Therefore, a quasi-experimental design was administered and variance as well as linear regression analyses were used.

Table 6
Regression analysis for academic characteristics, and online learning behavior predicting results of the knowledge transfer test ($N = 101$).

Knowledge transfer test result	*B*	*SE B*	*β*
Academic characteristics			
Semester	.081	.025	.273**
Study grade[1]	−.343	.116	−.247**
Prior knowledge	.001	.004	.025
Perceived difficulty	−.099	.085	−.100
Perceived confidence	.389	.076	.438***
Online learning behavior			
Views of handout	.274	.136	.218*
Views of additional material	.126	.152	.085
Views of video	.004	.057	.006
Overall interaction	−.023	.019	−.130

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.
[1] Due to German grading system, a smaller value indicates a better grade.

4.1. Findings on prompting and learning performance

Findings indicate that all participants had a significant increase of *declarative knowledge* between t_1 and t_2 and an expected decrease in t_3. As there was no interaction effect the groups did not change differently over time. At t_2 significant differences between the prompting conditions were found for declarative knowledge. However, the post-hoc tests did not reveal any significant results. Thus, based on the results of mixed ANOVA the prompts might have only limitedly impacted participants' declarative knowledge directly after they have received them. Referring

Table 5
Descriptive statistics and zero-order correlations of predictors used for linear regression analysis predicting the results of the knowledge transfer test.

Variable	1	2	3	4	5	6	7	8	9	10
1) Knowledge transfer test result	–									
2) Semester	.265**	–								
3) Study grade	.309**	−.033	–							
4) Prior knowledge	.049	.096	−.071	–						
5) Perceived difficulty	−.128	.080	.044	−.110	–					
6) Perceived confidence	.406***	−.032	−.028	−.018	−.134	–				
7) Views of handout	.223*	.003	−.136	−.137	.009	−.055	–			
8) Views of additional material	.140	.086	.001	−.057	.205*	−.091	.557***	–		
9) Views of video	−.029	−.224*	−.020	.033	−.029	.148	.059	.010	–	
10) Overall interaction	.052	.020	.032	−.089	−.021	.186*	.378***	.214*	.471***	–
N	101	101	101	101	101	101	101	101	101	101
M	1.01	5.07	2.35	31.43	2.89	2.50	.77	.50	1.42	7.91
SD	.831	2.81	.598	16.82	.835	.934	.662	.559	1.43	4.63

* $p < .05$.
** $p < .01$.
*** $p < .001$.

to the descriptive statistics, the *MP group* showed the highest results at t_2 compared to the other three groups.

With regard to the effects of the prompts on *knowledge transfer test results* ANOVA revealed a significant interaction effect of group and time, indicating that the groups changed differently over the time. However, no significant differences were found for the groups at each measurement point. Referring to Fig. 4 and descriptive statistics (see Table 2) the *control group* not receiving any prompts outperformed the experimental conditions at t_2. However, the *control group* also had the highest decrease from t_2 to t_3. *CP* and *AP group* had a very similar progression. Only the *MP group* had a slight increase between t_2 and t_3 or at least no substantive change over time. Based on these findings the prompts might have been counterproductive directly after the intervention as the *control group* using their own learning strategies without interruptions performed better. However, the metacognitive prompts might have a long-term effect on knowledge transfer as the *MP group* showed similar performance at t_2 and t_3. Thus, metacognitive prompts might be somehow related to learning performance, a finding mirrored by other studies (e.g., Kauffman, 2004; Lehmann, Hähnlein, & Ifenthaler, 2014; Nückles, Hübner, & Renkl, 2009). Nevertheless, the average learning performance, especially in the knowledge transfer tests but also in declarative knowledge, was rather low, indicating that the learning period was too short or participants did not take the task seriously. Thus, future studies should investigate prompts in authentic learning scenarios.

With regard to the knowledge transfer test results at t_2 and following the argumentation of Prieger and Bannert (2018) the fixed prompts used in this study might have been contrary to participants' needs or even interruptive. Potentially, the participants received too many or too great a variety of prompts in a relatively short learning period, leading to higher mental efforts or distraction (Bannert, 2007). However, besides the suggestion that learners should receive the support they need at the time they need it without increasing cognitive load (Sweller, 2011; Thillmann et al., 2009), no further recommendation of how many prompts are effective was found. Referring to the information available from studies using prompts the amount of prompts varied across studies and across conditions within the studies. For example, Backhaus et al. (2017) showed learners one to three of five possible prompts within e-modules of circa ten minutes, only one type of the prompts was related to increased test performance. In a study using reflective prompts learners had 35 min of learning time in a hypermedia learning environment and were prompted for metacognition as they should reflect and express each navigation decision within the environment leading to higher knowledge transfer performance compared to the control group but not to significant effects for performance in recall and knowledge (Bannert, 2006). In a study investigating short- and long-term effects of prompts, learners were designing their own metacognitive prompts before the learning occurred and should choose eight moments in time for receiving them during a 40 min learning period in a hypermedia learning environment. Results indicated differences between the prompted group and the control group with regard to their navigation patterns and their knowledge transfer test performance directly after the learning period and in a subsequent learning period without prompts, but no effects on recall and comprehension were found (Bannert et al., 2015). Moos and Bonde (2016) presented prompts related to planning (3 prompts), monitoring (4 prompts) and reflection (5 prompts) asking learners to verbalize in a learning session of approx. 45 min resulting in more self-regulated learning activities and higher test performance compared to the control group. Müller and Seufert (2018) showed their participants six prompts (3 cognitive and 3 metacognitive) within each of the two learning periods of thirty minutes resulting in increased performance in the first transfer test compared to the control group. Hence, the amount of prompts (see Section 2.3.2 for further details) participants received within this study was comparable with the amount in other studies. Hence, future research might further investigate the effects of the number of prompts and performance for example by using an experimental setting with different amounts of prompts, a think-aloud approach to gain insights into students' perceptions or by additionally measuring cognitive load.

In addition, for not being interruptive Molenaar and Roda (2008) argue that prompts should be in line with the learner's current goals and activities, whereas Wirth (2009) argues that prompts provide only limited information to the learner, thereby only insignificantly interrupting learning processes. To not interrupt learning processes machine learning approaches using trace data could be applied for predicting a moment in which learners are more likely to be in need of or to react and engage with the content prompted by using variables such as learners' current goals, emotional states, their past behavior or behavior indicating that learners are struggling but also demographic data (Pielot et al., 2017).

Further reasons why the varying prompting conditions might have had no considerable effect on learning performance might be due to the limited learning period, which did not allow the application of many different strategies, or may be due to the fact that learners who have already studied for an average of 4.86 (SD = 2.91) semesters have already established a relatively rigid learning behavior which will not be affected by temporary prompts. Furthermore, Prieger and Bannert (2018) found that learners with higher skills did not benefit from metacognitive prompts regarding learning performance, compared to those with lower skills. In this study, learners stated relatively high metacognitive awareness (knowledge about cognition M = .76; SD = .15; regulation of cognition M = .70; SD = .15; with a possible maximum of 1.0), hence, participants might already have known when to apply which strategy and might have felt distracted by the prompts.

4.2. Findings on perceptions with regard to the prompts

Besides the impact that the prompts might have on learning performance it is also relevant to investigate students' perceptions of such interventions. As outlined earlier, prompts might infer with learning processes. For example, in the study of Bannert and Reimann (2012) students reported in interviews after learning with prompts that they felt disturbed and interrupted by the prompts. Hence, prompts might evoke negative perceptions or diminish learners' perceived responsibility for their learning. Based on the findings of this study prompts were perceived differently with regard to learning support, and negative perceptions but not with regard to usefulness. The *MP group* significantly perceived more learning support than the *AP group*. That the *MP group* perceived the highest learning support through the prompts is in line with their learning performance in the declarative knowledge test and when only considering the experimental groups also in the transfer test. Furthermore, the *AP group* had the highest negative perceptions with regard to the prompts and significantly higher than the *CP group*. Thus, on a descriptive level the *AP group* had the highest negative perceptions, perceived the lowest learning support and lowest usefulness. As this group received potentially the highest amount of prompts and the most diverse prompts they might have felt distracted by the prompts. This assumption was further supported through additional analyses on participants' perception of having received too many prompts on item basis. Results showed, that *AP group* perceived more than all other groups having received too many prompts $M = 2.00$ ($SD = .903$), *MP group* $M = 1.55$ ($SD = 1.27$), and *CP group* $M = 1.07$ ($SD = 1.08$). One-way ANOVA revealed significant differences between the groups $F(2,86) = 5.247$, $p = .007$, $\eta^2 = .109$ with the *AP group* perceiving significantly more prompts than the *CP group*, $p = .005$. However, participants' perception of having received too many prompts is still relatively low on a scale from 1 to 5 with 5 indicating strong agreement on having received too many prompts. Hence, additional studies might use think-aloud methods to investigate students' perceptions of receiving prompts in more detail.

4.3. Findings on prompting and learning behavior indicated by trace data

To gain further insights if different prompts affect learners' behavior in a typical digital learning environment (handout, additional learning material, video, overall interaction), MANOVA was used.

Based on the trace data available for the learning unit findings indicate significant differences between the groups in *viewing the handout* and the *additional material*. However, no significant differences were found for *video views* and the *overall interaction* with the learning unit. With regard to the *views of the handout* the *control group* viewed the handout significantly more often than the *AP group*. Regarding the *views of the additional material* the *control group* viewed the material significantly more frequently than all experimental groups. Based on the descriptive results the participants mainly interacted with the video compared to the handout and additional material. Most participants started to watch the video at least once (77.3%) or twice (15.5%), only three participants did not start watching it. In contrast 35.5% of the participants did not view the handout, and 54.5% did not click on the additional material.

With focus on *number and length of notes taken* in the learning unit descriptive statistics show that the prompt asking participants to take notes did impact their behavior as the *control group* did not take notes in contrast to all prompting conditions. In addition, when comparing the experimental groups, the *AP group* significantly took more notes than the *MP group*. However, the prompted note taking did not seem to have a positive effect on learning performance, as indicated by previous studies (Nye, Crooks, Powley, & Tripp, 1984; Peverly, Brobst, Graham, & Shaw, 2003). As taking many notes might also indicate that students did not select which information is relevant (Zimmerman, 2008). Hence, it is necessary to further investigate the content and quality of the notes taken. Furthermore, even though solely the *AP group* received a prompt with the information, that there is additional learning material available, this group accessed the additional material least of all. One reason might be that they received too many prompts and were busy in taking notes as indicated by the descriptive statistics. However, the *control group* viewed the handout and additional material the most but had the lowest overall interaction with the learning unit.

In summary, the findings on trace data are ambiguous, as prompts to take notes seemed to have an effect, but the *AP group* did not follow the prompt pointing to the additional material. Potentially, the *control group* applied their own strategies and browsed more efficiently trough the learning unit as indicated by the overall interaction, whereas the prompted groups relied on the guidance offered through the prompts possibly inferring their own learning strategies. The results from this quasi-experimental study make it difficult to infer how to design prompts as guidance for learner by not impairing their self-directedness or established strategies. Especially, when considering that interventions in higher education should support processes of self-regulated learning and learning performance. In addition, further evidence is required with regard to what successful navigation patterns are. For example, linearity of navigation behavior, as the prompts in the study of Prieger and Bannert (2018) resulted in different learning behavior with regard to linearity but these behavioral differences did not impact learning performance. In the study of Müller and Seufert (2018) the groups did not differ significantly with regard to navigation behavior and only non-linear navigation behavior was negatively related to performance in one transfer test. Whereas, in the study of Bannert et al. (2015) results indicated that self-directed metacognitive prompts increased the visit of relevant pages which had an impact on transfer test performance.

4.4. Findings on trace data informing learning performance

One major aim of learning analytics approaches is to use learners' behavior in digital learning environments for predicting learning performance (El-Rady, Mohamed, & El Fakharany, 2017). However, in this study only participants' number of views of the handout was a significant predictor of their learning performance in the transfer test. More relevant for predicting learning performance were academic characteristics of the participants such as the semesters studied or their perceived confidence. However, when facing the issue that many current learning analytics systems do not even refer to additional information about learners (Vieira et al., 2018), this might significantly reduce their validity.

In summary, in this quasi-experimental study, trace data did not, as expected, provide explanation for learning performance. However, it needs to be kept in mind the limitedness of the trace data in this study as it was only possible to track participants' initial click on the resources which opened in another window of the browser not allowing to track further interaction such as scrolling the material.

Given these findings on trace data, and considering that learning analytics approaches have only limited data available (e.g., not all indicators can be captured through the system or the system has no access to learner characteristics) and furthermore, that the data available are affected by learning processes outside the digital learning environment, learning analytics might only offer very limited insights into students' learning processes (e.g., Ferguson, 2012; Wilson et al., 2017; Winne, 2017b). Digital learning environments in higher education to date are limited, as they only allow data to be captured on online behavior, for example, accessing a folder or downloading the slides. At the most, they can track the length of time students spend watching videos or passing self-assessments, and they might be aggregated to behavioral patterns. But as the systems do not offer sufficient learning opportunities, any actual learning tends to take place outside the system, leading to biases in the indicators and predictions. In addition, by predominantly relying on quantifications of behavioral data the implicit assumption would be that numbers of trackable actions are related to the quality of learning and performance (Gašević et al., 2015; Wise & Shaffer, 2015). Furthermore, the data cannot be analyzed without relating them to the context they were collected in (Gašević et al., 2015; Macfadyen & Dawson, 2012) by considering the instructional design, intent and tasks plus the functionalities of the digital learning environment. Consequently, inferring from these data on learning requires further empirical evidence (Ferguson & Clow, 2017).

4.5. Limitations and further research

This study shows several limitations which need to be addressed when interpreting the findings and preparing future research. For example, the experimental setting allowed only a limited learning time and limited learning material, resulting in restricted possibilities for tracking learning behavior or validly inferring strategy use from trace data. Even though the university's common digital learning environment was used, and participants were told that they have to pass the knowledge tests to receive the credit points, the learning scenario is not comparable to authentic learning settings. For example, motivational dispositions such as achievement goals are relevant factors for successful learning processes (Schunk, Pintrich, & Meece, 2008). In this quasi-experimental approach, learners' motivation and goals may differ from those they have in authentic learning situations.

Prompts in this study were not adaptive or personalized based on learners' current behavior or their characteristics, raising the issue that prompts that are supportive for one learner might not be helpful for another, due to different prerequisites or characteristics (Backhaus et al., 2017; Lin & Lehman, 1999; Prieger & Bannert, 2018). Further analyses might investigate if relations of prompts, learner characteristics and learning performance exist. In addition, participants only received prompts during t_2, which, in total, took about 45 min. Hence, they were not familiar with receiving prompts, used them only for a short time and received many prompts. However, presenting prompts to students over a longer period might reduce the possible benefits of prompts related to learning performance which Nückles, Hübner, and Renkl (2008) refer to as over-prompting. In addition, it needs to be further analyzed how many prompts the learners were actually facing during their individual learning paths.

Nevertheless, the experimental setting allowed to control for possible external learning behavior of the participants, making the results on the insufficient predictive power of online trace data even more significant. The digital learning environment used for this experiment is a well-known system among European universities, however, current tracking is limited to track clicks in the system. Many digital learning environments used in higher education, however, are comparable or even less complex than the learning scenario used in this experimental setting, and learning analytics approaches are applied in such systems for identifying students at risk or predicting learning performance (Zhang & Almeroth, 2010). Consequently, when designing new digital learning environments they should offer authentic learning opportunities, and possibilities to capture learners' behavior need to be considered and implemented at the very beginning.

As participants of this quasi-experimental study were only from one university, findings cannot be generalized, but might be investigated further including students from more universities.

Hence, future research should investigate prompts in real learning settings, which will be the next step in this research project. Particularly, research on learners' reaction to prompts should be in focus using trace data. As no adaptive or personalized prompts were used in this study, referring to Backhaus et al. (2017), the system will offer adaptive prompts to the learners based on their (self-reported) learning characteristics and their learning behavior in the online system. For example, when students only download the lecture slides, they will receive a prompt referring to the lecture recordings, the self-assessments, or further readings. Furthermore, upcoming analyses might investigate individual navigation patterns related to the learning performance or the prompts received.

5. Conclusion

The purpose of this quasi-experimental study was to investigate the effect of prompts on the learning performance in declarative knowledge and transfer tests, as well as on learning behavior. Furthermore, participants' perceptions of the prompts were investigated and whether and how trace data can inform predicting learning performance.

Findings indicated that prompting only limitedly affected declarative knowledge and only might have had an impact on knowledge transfer in a subsequent test. With regard to the experimental groups they evaluated the prompts differently in terms of perceived learning support and negative perceptions. In addition, differences in learning behavior were found between the control group and the experimental groups. Furthermore, the power of trace data to inform predictions on learning performance was rather limited. In this study, learning performance with regard to knowledge transfer was only predicted by the frequency of views of the related handout.

Ferguson and Clow (2017) claim that learning analytics still lack empirical evidence, this study, in using a quasi-experimental approach and by reporting 'negative' findings, highlights the potential limitations of learning analytics approaches especially when facing small data sets, and aims to encourage upcoming studies to use experimental study designs.

Prompts in this research might have not been efficient, as they were not related to students' characteristics or behavior, resulting in inappropriate support. However, information based on trace data might be helpful in generating effective instructional means and should be investigated further.

Research on prompts using learning analytics approaches needs to further investigate how trace data can be used for technology-driven interventions, how they may support learning and how students perceive such algorithm-based recommendations and feedback.

Declaration of Competing Interest

None.

References

Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovations, and social justice in education. *TechTrends, 62*, 37–45. https://doi.org/10.1007/s11528-017-0226-9.

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.

Backhaus, J., Jeske, D., Pointstingl, H., & Koenig, S. (2017). Assessing efficiency of prompts based on learner characteristics. *Computers, 6*(1), 7. https://doi.org/10.3390/computers6010007.

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

Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia. Erfassung, Beschreibung und Vermittlung wirksamer metakognitiver Lernstrategien und Regulationstechniken.* Münster: Waxmann.

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie, 23*(2), 139–145.

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.

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*, 193–211. https://doi.org/10.1007/s11251-011-9167-4.

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Prieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior, 52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038.

Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology, 41*(4), 377–397. https://doi.org/10.1111/j.1464-0597.1992.tb00713.x.

Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research, 31*(6), 445–457.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies and academic achievement in online higher education learning environments: A systematic review. *Internet and Higher Education, 27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007.

Cassidy, S. (2011). Self-regulated learning in higher education: Identifying key component processes. *Studies in Higher Education, 36*(8), 989–1000.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653.

Corrin, L., & de Barba, P. (2014). Exploring students' interpretation of feedback delivered through learning analytics dashboards. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rethoric and reality: Critical perspectives on educational technology. Proceedings of the Ascilite conference* (pp. 629–633) (Dunedin, NZ).

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior, 73*, 247–256. https://doi.org/10.1016/j.chb.2017.01.047.

Daumiller, M., & Dresel, M. (2018). Supporting self-regulated learning with digital media using motivational regulation and metacognitive prompts. *The Journal of Experimental Education*. https://doi.org/10.1080/00220973.2018.1448744.

Drachsler, H., & Kirschner, P. A. (2012). Learner characteristics. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 1743–1745). Boston, MA: Springer US.

El-Rady, A. A., Mohamed, S., & El Fakharany, E. (2017). Predicting learner performance using data-mining techniques and onthology. In A. E. Hassanien, K. Shaalan, T. Gaber, A. T. Azar, & M. F. Tolba (Eds.)*, vol. 533. Proceedings of the international conference on advanced intelligent systems and informatics. AISI 2016* (pp. 660–669). Cham: Springer.

Engelmann, K., & Bannert, M. (2019). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.05.002.

Eradze, M., Väljataga, T., & Laanpere, M. (2014). Observing the use of e-textbooks in the classroom: Towards "offline" learning analytics. In Y. Cao, T. Väljataga, J. K. T. Tang, H. Leung, & M. Laanpere (Eds.), *New horizons in web based learning. ICWL 2014 international workshops* (pp. 254–263). Cham: Springer.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317.

Ferguson, R., & Clow, D. (2017). Where is the evidence? A call to action for learning analytics. In *Proceedings of the international learning analytics and knowledge conference* (pp. 56–65). New York, NY: ACM.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technolgy & Society, 15*(3), 42–57.

Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Journal of Educational Technology & Society, 15*(1), 38–52.

Ifenthaler, D. (2015). In J. M. Spector (Ed.), *The Sage Encyclopedia of Educational Technology, (Vol. 2,* pp. 447–451). Los Angeles, California: Sage Publications.

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y.

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240. https://doi.org/10.1007/s10758-014-9226-4.

Kauffman, D. F. (2004). Self-regulated learning in web-based environments: Instructional tools designed to facilitate cognitive strategy use, metacognitive processing, and motivational beliefs. *Journal of Educational Computing Research, 30*(1 & 2), 139–161.

Kramarski, B., & Kohen, Z. (2017). Promoting preservice teachers' dual self-regulation roles as learners and as teachers: Effects of generic vs. specific prompts. *Metacognition and Learning, 12*, 157–191. https://doi.org/10.1007/s11409-016-9164-8.

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. https://doi.org/10.1016/j.chb.2013.07.051.

Lin, X., & Lehman, J. D. (1999). Supporting learning of variable control in computer-based biology environment: Effects of prompting college students to reflect on their own thinking. *Journal of Research in Science Teaching, 36*(7), 837–858.

Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technolgy & Society, 15*(3), 149–163.

Mah, D.-K. (2016). Learning analytics and digital badges: Potential impact on student retention in higher education. *Technology, Knowledge and Learning, 21*(3), 285–305.

McLoughlin, C., & Lee, M. J. W. (2010). Personalized and self regulated learning in the web 2.0 era: International exemplars of innovative pedagogy using social software. *Australasian Journal of Educational Technology, 26*(1), 28–43.

Molenaar, I., & Roda, C. (2008). Attention management of dynamic and adaptive scaffolding. *Pragmatics and Cognition, 16*(2), 224–271. https://doi.org/10.1075/p&c.16.2.04mol.

Moos, D. C., & Bonde, C. (2016). Flipping the classroom: Embedding self-regulated learning prompts in videos. *Technology, Knowledge and Learning, 2016*, 225–242. https://doi.org/10.1007/s10758-015-9269-1.

Müller, N. M., & Seufert, T. (2018). Effects of self-regulation prompts in hypermedia learning on learning performance and self-efficacy. *Learning and Instruction, 58*, 1–11. https://doi.org/10.1016/j.learninstruc.2018.04.011.

Nadasen, D., & List, A. (2017). Predicting four-year student success from two-year student data. In B. Key Daniel (Ed.), *Big data and learning analytics in higher education* (pp. 221–236). New York: Springer.

Nückles, M., Hübner, S., & Renkl, A. (2008). Short-term versus long-term effects of cognitive and metacognitive prompts in writing-to-learn. In *, vol. 2. Proceedings of the eighth international conference for the learning sciences – ICLS 2008* (pp. 124–131). International Society of the Learning Sciences.

Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*, 259–271. https://doi.org/10.1016/j.learninstruc.2008.05.002.

Nussbaumer, A., Dahn, I., Kroop, S., Mikroyannidis, A., & Albert, D. (2015). Supporting self-regulated learning. In S. Kroop, A. Mikroyannidis, & M. Wolpers (Eds.), *Responsive open learning environments: Outcomes of research from the ROLE project* (pp. 17–48). Cham: Springer International Publishing.

Nye, P. A., Crooks, T. J., Powley, M., & Tripp, G. (1984). Student note-taking related to university examination performance. *Higher Education, 13*(1), 85–97. https://doi.org/10.1007/BF00136532.

Peverly, S. T., Brobst, K. E., Graham, M., & Shaw, R. (2003). College students are not good at self-regulation: A study on the relationship of self-regulation, note taking, and test taking. *Journal of Educational Psychology, 95*(2), 335–346. https://doi.org/10.1037/0022-0663.95.2.335.

Pielot, M., Cardoso, B., Katevas, K., Serrà, J., Matic, A., & Oliver, N. (2017). Beyond interruptibility: Predicting opportune moments to engage mobile phone uses. In *, vol. 1. Proceedings ACM on interactive, mobile, wearable and ubiquitous technologies*. Article 91.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Prieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. *Computers in Human Behavior, 86*, 165–173. https://doi.org/10.1016/j.chb.2018.04.022.

Schmitz, B. (2001). Self-monitoring zur Unterstützung des Transfers einer Schulung in Selbstregulation für Studierende. Eine prozessanalytische Untersuchung. [Self-monitoring to support the transfer of a self-regulation instruction for students: A process study]. *Zeitschrift für Pädagogische Psychologie, 15*(3/4), 181–197.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460–475.

Schunk, D. H. (2012). *Learning theories. An educational perspective*. Boston: Pearson.

Schunk, D. H., Pintrich, P. R., & Meece, J. L. (2008). *Motivation in education* (3 ed.). Upper Saddle River: Pearson/Merrill Prentice Hall.

Sonnenberg, C., & Bannert, M. (2016). Evaluating the impact of instructional support using data mining and process mining: A micro-level analysis of the effectiveness of metacognitive prompts. *Journal of Educational Data Mining, 8*(2), 51–83.

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre, & B. H. Ross (Eds.)*, vol. 55. Psychology of learning and motivation. Cognition in education*. San Diego, CA: Academic Press.

Thillmann, H., Künsting, J., Wirth, J., & Leutner, D. (2009). Is it merely a question of "what" to prompt or also "when" to prompt? The role of point of presentation time of prompts in self-regulated learning. *Zeitschrift für Pädagogische Psychologie, 23*(2), 105–115.

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers in Education, 122*, 119–135. https://doi.org/10.1016/j.compedu.2018.03.018.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 315–327). New York: Macmillan.

Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. *Teaching in Higher Education, 22*(8), 991–1007.

Winne, P. H. (2017a). Cognition and metacognition within self-regulated learning. In P. A. Alexander, D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-reguation of learning and performance* (pp. 36–48). New York, NY: Routledge.

Winne, P. H. (2017b). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (1 ed., pp. 241–249). Society for Learning Analytics Research.

Winne, P. H., & Baker, R. S. J. D. (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. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence.

Wirth, J. (2009). Promoting self-regulated learning through prompts. *Zeitschrift für Pädagogische Psychologie, 23*(2), 91–94.

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in age of big data. *Journal of Learning Analytics, 2*(2), 5–13. https://doi.org/10.18608/jla.2015.22.2.

Zhang, H., & Almeroth, K. (2010). Tracking student activity in online course management systems. *Journal of Interactive Learning Research, 21*(3), 407–429.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego, CA: Academic Press.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 1–35). New York: Routledge.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70.

Zimmerman, B. J. (2008). Investigating self-regulated learning and motivation: Historical background, methodological developments, and future prosepects. *American Educational Research Journal, 45*(1), 166–183.