Effects of real-time analytics-based personalized scaffolds on students’ self-regulated learning

Lyn Lim a,⁎, Maria Bannert a, Joep van der Graaf b, Shaveen Singh d, Yizhou Fan c, Surya Surendrannair a, Mladen Rakovic d, Inge Molenaar b, Johanna Moore c, Dragan Gašević c,d

a Chair for Teaching and Learning with Digital Media, TUM School of Social Sciences and Technology, Technical University of Munich, Arcisstraße 21, 80333 München, Germany
b Behavioural Science Institute, Radboud University, Postbus 91046500 HE Nijmegen, the Netherlands
c School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, Scotland, UK
d Centre for Learning Analytics, Faculty of Information Technology, Monash University, Clayton CampusBuilding 6, 29 Ancora Imparo Way, VIC 3800, Melbourne, Australia

⁎ Corresponding author.
E-mail addresses: lyn.lim@tum.de (L. Lim), maria.bannert@tum.de (M. Bannert), joep.vandergraaf@ru.nl (J. van der Graaf), Shaveen.Singh@monash.edu (S. Singh), yizhou.fan@ed.ac.uk (Y. Fan), Surya.Surendrannair@monash.edu (S. Surendrannair), mladen.rakovic@monash.edu (M. Rakovic), inge.molenaar@ru.nl (I. Molenaar), J.Moore@ed.ac.uk (J. Moore), dragan.gasevic@monash.edu (D. Gašević).

https://doi.org/10.1016/j.chb.2022.107547
Received 30 March 2022; Received in revised form 3 July 2022; Accepted 23 October 2022
Available online 27 October 2022
© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Effects of real-time analytics-based personalized scaffolds on students’ self-regulated learning

Self-Regulated Learning (SRL) is related to increased learning performance. Scaffolding learners in their SRL activities in a computer-based learning environment can help to improve learning outcomes, because students do not always regulate their learning spontaneously. Based on theoretical assumptions, scaffolds should be continuously adaptive and personalized to students’ ongoing learning progress in order to promote SRL. The present study aimed to investigate the effects of analytics-based personalized scaffolds, facilitated by a rule-based artificial intelligence (AI) system, on students’ learning process and outcomes by real-time measurement and support of SRL using trace data. Using a pre-post experimental design, students received personalized scaffolds (n = 36), generalized scaffolds (n = 32), or no scaffolds (n = 30) during learning. Findings indicated that personalized scaffolds induced more SRL activities, but no effects were found on learning outcomes. Process models indicated large similarities in the temporal structure of learning activities between groups which may explain why no group differences in learning performance were observed. In conclusion, analytics-based personalized scaffolds informed by students’ real-time SRL measured and supported with AI are a first step towards adaptive SRL supports incorporating artificial intelligence that has to be further developed in future research.

SRL and improved learning performance has been well documented in prior research (Dent & Koenka, 2016; Schunk & Greene, 2017). However, students most often do not spontaneously regulate their learning (Bjork, Dunlosky, & Kornell, 2013; Flavell, Beach, & Chinsky, 1966, pp. 283–299; Veenman, Kok, & Blote, 2005) and fail to regulate their learning successfully in digital and online learning settings (e.g., Azevedo & Feyzi-Behnagh, 2011). Recently, the need to support SRL in digital learning settings has become particularly urgent due to an abrupt increase in remote learning (i.e., digital and online), mainly as a result of the current pandemic (EDUCAUSE, 2021, p. 50).
Past research has shown that scaffolding different SRL activities led to better learning outcomes (Bannert, Sonnenberg, Mengelkamp, & Pieger, 2015; Guo, 2022; Zheng, 2016). Traditionally, scaffolds have been defined as instructional tools and strategies facilitated by a tutor (e.g., a teacher) with the aim of supporting learners to achieve what they are unable to do without support (Reiser & Tabak, 2014). Scaffolds have gradually been extended to prompts and hints used in tools and resources embedded in digital learning settings to support learning (Puntambekar & Hubscher, 2005). The overarching goal of scaffolding is to support students so they internalize the required skills and perform them independently (Wood, Bruner, & Ross, 1976). There has been several studies so far that developed and implemented computer-generated scaffolds to support SRL in digital learning settings with varied success (e.g., Azevedo, Johnson, Chauncey, & Burkett, 2010; Bannert, Sonnenberg, et al., 2015; Daumiller & Dresel, 2019; Molenaar, Roda, van Bokel, & Sleegers, 2012; Munshi & Biswas, 2022; Schumacher & Henthaler, 2021; Siadaty, Gasevic, & Hatala, 2016). However, there have been documented challenges related to past SRL interventions based on scaffolds (e.g., standardized prompts, Muer, Zumbach, & Deibl, 2017; self-directed prompts, Pieger & Bannert, 2018). For instance, the scaffolds developed in prior research insufficiently account for dynamics of SRL processes; they lack adequate adaptability and personalization to cater to learners’ ongoing learning progress, tailor support appropriately to individual learners based on the ongoing diagnosis of their learning, and then fade support when it is no longer needed. Low scaffold adaptability and personalization may have potentially led to poor compliance and hence, missing benefits on learning outcomes (Bannert & Mengelkamp, 2013; Lallé, Conati, Azevedo, Mudrick, & Taub, 2017).

Scaffolds need to be (inherently) adaptive and also personalized to support learning in an optimal manner. Scaffolds adapt by means of three components: 1) providing ongoing diagnosis of students’ learning progress and, 2) calibrating support accordingly (i.e., whether support is needed and hence, 3) fading support when students perform activities independently) (Molenaar et al., 2012; Puntambekar & Hubscher, 2005). Personalized support involves customizing the content of support provided according to the needs of students (Lim, Gentili, et al., 2021; Pardo, Jovanovic, Dawson, Gasevic, & Mirriahi, 2019). Adaptive scaffolds have been proposed to be more advantageous for learning (Azevedo & Hadwin, 2005). Yet, more work needs to be done to develop and test adaptive scaffolds (Guo, 2022; Zheng, 2016). Prior studies have embedded adaptive scaffolds in computer-based learning environments which detected students' learning progress and consequently adapted scaffolds to their SRL processes (Metatutor, Azevedo et al., 2010; Agtentic, Molenaar et al., 2012; Bettys Brain, Munshi and Biswas, 2022). However, such adaptive scaffolding systems are still uncommon and the solutions to date do not include personalized SRL support. Analytics-based approaches using trace data allows the measurement and support of students’ SRL in an unobstrusive manner (Lim, Gentili et al., 2021; Pardo et al., 2019; Siadaty, Gasevic, & Hatala, 2016). Furthermore, understanding how scaffold interventions affect students learning through process analysis methods (e.g., process mining) using event-based data, like trace data, visualizes the SRL process and increases explanatory power (Reimann, 2009). To that end, we developed a rule-based artificial intelligence (AI) system and investigated how this system can be used to govern personalized support on students’ SRL processing. Rule-based AI systems rely upon a set of rules that are often defined by domain experts. The rules are of a conditional statement, “IF [condition] THEN [act]” (Planinski, 2016). Conditions in our system were modeled from trace data, i.e., SRL processes detected in trace data—IF [condition]—signaled whether an SRL scaffold was needed at a particular point in a learning session—THEN [act]. Based on a theory- and data-driven SRL measurement approach, raw trace data were first labeled as learning actions, then sequences of learning actions were mapped to SRL activities. Acts were implemented as real-time personalized scaffolds, i.e., depending on the current condition of students’ SRL processing, appropriate personalized scaffold was provided as support. In this way, our rule-based system tracked students’ SRL in real-time and provided personalized support, when needed. The rules in our system were developed by learning scientists in our research group who consulted both SRL theory as well as analyzed students’ SRL in prior lab studies using the same learning context. Therefore, through a two-pronged approach of supporting students through the integration of an analytics-based SRL measurement protocol and real-time personalized scaffolding on the basis of trace data (i.e., analytics-based personalized scaffolds) via a rule-based AI approach, the study presented investigated and tested personalized scaffolds which adapted to students’ learning progress as well as customized support content to individual students.

In the present study, we tested and examined the effects of analytics-based personalized scaffolds on learning activities, learning performance, and temporal structure of learning activities by comparing students supported by personalized scaffolds with students supported by generalized scaffolds, which were standardized for all students, and students who were not supported by scaffolds. The aim of the paper is to investigate and discuss how analytics-based personalized scaffolds affect learning as the first steps to further improve and extend such interventions with other appropriate artificial intelligence methods (e.g., incorporating automatic coding of essays through natural language processing as further input for personalized scaffolding).

1. Supporting SRL with scaffolds

Self-regulated learning (SRL) is essential in successful learning, especially in computer-based learning environments (Azevedo, 2007). SRL is an active and constructive process, whereby students set goals and pursue them by regulating their learning (Pintrich, 2000). Theoretical models conceptualize SRL as a cyclical process consisting of the preparatory, performance, and appraisal phases (Panadero, 2017; Puustinen & Pulkkinen, 2001). At the preparatory stage of learning, students engage in metacognitive activities through analyzing the task to get an orientation, setting goals, and creating a plan (Zimmerman, 2000). The subsequent phase encompasses execution of the task through cognitive strategies and during SRL, students monitor and control the processing of learning content and the operations they apply to the processing of content (Winne, 2019). Cognitive strategies include both low (also known as surface or superficial) and high cognitive (also known as deep processing or learning) activities. Students first familiarize themselves with surface knowledge by constructing a knowledge base (Frey, Fisher, & Hattie, 2017), such as by reading the text. They then engage in other low cognitive activities such as rehearsal, and high cognitive activities, such as elaboration and organization (Weinstein & Mayer, 1983). Hence, cognitive activities, which include both low and high cognition, play different but interconnected roles in the performance phase. According to Nelson and Narens (1994), cognition is structured into two interdependent levels: “meta-level” and “object-level.” The meta-level is a mental model built from the object-level (i.e., one’s cognition). For instance, monitoring (e.g., judgments of learning) occurs at the meta-level representation and can alter cognition at the object-level through control strategies (e.g., elaboration). Lastly, at the appraisal phase, students evaluate and reflect on their learning by comparing their current progress with a previously set goal or standards (Zimmerman, 2000). Therefore, SRL comprises dynamic processes to actively regulate one’s own learning by means of monitoring and controlling SRL processes in the pursuit of one’s goals.

SRL has been found to be related to learning performance, especially in the contexts where students apply knowledge and skills to a new situation or problem (Schunk & Greene, 2017). Previous research has shown that metacognitive activities promote deeper understanding of learning content (Bannert, Hildebrand, & Mengelkamp, 2009), and that monitoring activities were associated with the increased use of deep learning strategies (Deekens, Greene, & Lobczowski, 2018). In two
experiments conducted by Roelle, Nowitzki, and Berthold (2017), they consistently found that students prompted to engage in metacognitive processes subsequently had a higher quality of organization processes. This so-called stage-setting effect occurs when metacognitive processes, such as monitoring, lead to correction of discrepancies at the object-level of cognition, allowing subsequent cognitive processes to begin from an improved “starting point” (i.e., knowledge base). Moreover, successful learning in previous studies was characterized by more strategic SRL behavior, especially in terms of the temporal characteristics of SRL as indicated by process analyses, as theorized by SRL models (e.g., Zimmerman, 2002), such as first analyzing the task, monitoring while performing various cognitive activities, and evaluating learning (Bannert, Reimann, & Sonnenberg, 2014; Lim et al., 2021; Paans, Molenaar, Segers, & Verhoeven, 2019). Yet, students most often do not spontaneously regulate their learning and experience difficulties adequately regulating their learning (production deficit, Flavell et al., 1966; Veerman et al., 2005; dysregulated learning, Azevedo & Feyzi-Behnagh, 2011), which is exacerbated by challenges in monitoring and controlling learning in computer-based learning environments (Azevedo, 2005; Broadbent & Poon, 2015).

To address this gap, scaffolds can be used to support students’ SRL and consequently, improve their learning outcomes. Scaffolding provides support to students on an as-needed basis, with the external support fading as students’ competencies increase (Wood et al., 1976). Different forms of scaffolds have been utilized in prior research to support SRL, e.g., prompts, tools, pedagogical agents or intelligent tutors, and feedback (Harley, Taub, Azevedo, & Bouchet, 2018; Molenaar & Chiu, 2014; Roll et al., 2006; Zheng, 2016). Prompts are short-term interventions that tackle students’ difficulties in regulation during learning by stimulating execution of existing skills and knowledge (Bannert, 2009). The underlying assumption behind prompting is that learners have production deficiency; this means that students possess the necessary regulatory skills and knowledge which typically develop to a sophisticated level, especially at tertiary education, through years of formal education, but are unable to spontaneously execute them (Veerman, 2016; Veerman, Van Hout-Wolters, & Aflerbach, 2006).

Learning outcomes, particularly scores on transfer tests, can be improved by scaffolding learners with respect to different SRL activities, including in computer-based learning environments (Bannert & Mengelkamp, 2013; Bannert, Sonnenberg, et al., 2015; Lim & Leeman, 1999; Müller & Seufert, 2018; Zheng, 2016). As previous research has shown that metacognitive activities stimulate deeper processing of learning materials through high cognitive activities, particularly relevant for deep knowledge tasks, such as transfer tasks (Bannert et al., 2009; Molenaar & Chiu, 2017; in press van der Graaf et al., 2022), scaffolding metacognitive activities can potentially lead to increasing high cognitive activities, such as elaboration and organization activities. However, scaffolds fostering SRL do not always have the desired effects on learning outcomes though they may influence learning behavior (Engelmann & Bannert, 2019; Pieger & Bannert, 2018). To optimally support SRL, scaffolds should be fundamentally adaptive in order to diagnose ongoing learning, calibrate support accordingly, and fade support when no longer needed (Puntambekar & Hubers, 2005). Effective scaffolds should also support different SRL activities throughout SRL phases (Zheng, 2016).

Recent meta-analyses have highlighted the need for more development and implementation of adaptive scaffolding to foster SRL (Guo, 2022; Zheng, 2016). Scaffolds used in previous studies were mainly generalized, meaning they provided standardized support to all students (e.g., Bannert & Mengelkamp, 2013; Daumiller & Dresel, 2015; Moser et al., 2017; Müller & Seufert, 2018; Schumacher & Ifenthaler, 2021), with the exception of self-directed prompts designed by students (Bannert, Sonnenberg, et al., 2015; Pieger & Bannert, 2018). Yet, adaptability is one of the significant moderators of SRL activities (Guo, 2022). However, studies using adaptive scaffolds to promote SRL are still limited (Zheng, 2016). Azevedo et al. (2010) adopted adaptive scaffolding features in the MetaTutor system, an intelligent hypermedia learning environment fostering SRL in the context of science learning. One of the functions of the system was that pedagogical agents prompted learners to engage in various SRL strategies depending on their progress on the task as measured by trace data. Molenaar et al. (2012) implemented a scaffolding and learning system, AgentSchool, which adjusted scaffolds according to students’ progress on the task. They introduced dynamic scaffolding to foster socially regulated learning with fifth grade students who worked in dyads. The system recorded and tracked students’ learning actions (or absence of actions, i.e., idle) and used students’ attention focus as a basis for triggering scaffolds. In both MetaTutor and AgentSchool systems, scaffolds can be considered adaptive in the sense that students’ progress on a task was taken into consideration when providing support, but the scaffolds lacked both continuous assessment of students’ learning and a calibration and fading mechanism. Munshi and Biews (2022) took adaptive scaffolding one step further from past studies with their recent development of an adaptive scaffolding framework embedded in an open-ended learning system, Betty’s Brain. They introduced ongoing detection of students’ learning behavior to trigger scaffolds and used key transition points to provide contextualized scaffolds. They then studied the use and effect of each adaptive scaffold on high and low performing students. They found that the adaptive scaffolds had a larger effect on learning gain for high performers. Their closer inspections of individual adaptive scaffolds revealed differential effects on both learning behavior and performance between the high and low performing groups. Although development of adaptive scaffolds in computer-based learning environments have advanced, in order to match students’ learning needs and progress, there is not only a need to diagnose students’ actual and prior SRL processes, but to also personalize these scaffolds to align with individual ongoing student learning behavior and include fading mechanisms.

1.2. Measuring and supporting SRL processes using an analytics-based approach

Research on SRL has shifted from self-reported SRL activities (i.e., variable-centered measures like the Motivated Strategies for Learning Questionnaire, Pintrich, Smith, Garcia, & McKeachie, 1993) to collecting event-based data to reflect how SRL processes dynamically unfold over time (Reimann, Markauskaite, & Bannert, 2014; Winne & Perry, 2000). Past studies have measured online SRL processes by using think aloud protocols (e.g., Bannert, 2007; Johnson, Azevedo, & D’Mello, 2011), micro-analyses (e.g., Cleary & Callan, 2017), and progressively trace data (e.g., Schumacher & Ifenthaler, 2021). Trace data (also termed as peripheral data, e.g., Hornmann & Bannert, 2016) are logs which contain digital traces of learners’ interactions within the learning environment used to infer SRL processes and include navigational logs, keystrokes, mouse movement, and eye gaze points (Bernacki, 2018). An advantage of using trace data is that they measure student activities in an unobtrusive manner (Winne, 2010). In Hornmann and Bannert’s (2016) work, they exemplified how typing behavior data was related to motivation and task performance and expressed the potential for machine learning algorithms to automatically label peripheral data in real-time. In the examples of scaffold-studies we provided in the previous section (i.e., Bannert, Pieger, & Sonnenberg, 2015; Müller & Seufert, 2018; Pieger & Bannert, 2018), the researchers collected and deduced learning behavior from a commonly used form of trace data—navigation logs. However, navigation logs alone do not reliably measure students’ ongoing learning processes as they are not sufficiently fine-grained (Jarvela & Bannert, 2021). Hence, analytics-based protocols have been developed to harness different forms of trace data (e.g., navigation logs, keystrokes, and mouse movement) for measuring and understanding SRL processes with finer granularity (Fan et al., 2022; Greene & Azevedo, 2009; Siedaty, Gasević, & Hatala, 2016). Analytics-based protocols refer to measurement protocols which utilize trace data and analytic methods for extracting SRL processes (e.g., Gasević, Jovanovic, Pardo, & Dawson,
computers were most effective for fostering SRL in the context of a workplace by using an analytics-based protocol to determine associations of micro-level SRL processes (e.g., task analysis, goal setting, and evaluation) and intervention use and how the interventions influenced participants’ micro-level SRL processes.

In addition to assessing learning behavior, the analytics-based approach offers a way to personalize support for learners. Learner trace data hence can be dynamically analyzed to provide at scale support tailored to learners’ immediate needs. Pardo et al. (2019) provided personalized feedback messages in a university course using algorithms to customize feedback content. They reported a significant positive impact of personalized feedback on students’ exam scores and higher levels of satisfaction with the feedback. Lim, Gentili, et al. (2021) used a learning analytics-based system to personalize feedback through emails sent to students in an undergraduate course. They predetermined rules to configure feedback content based on trace data recorded from the learning management system. They found that personalized feedback led to more regular studying and higher course grades. The studies from Pardo et al. (2019) and Lim, Gentili, et al. (2021) demonstrated the viability of implementing an analytics-based approach to personalize support leading to better learning outcomes. Additionally, findings of higher student satisfaction are noteworthy since the use of tools (e.g., feedback tools) is influenced by how students experience these tools (Gasevic, et al., 2017). Although past studies demonstrated innovative methods to personalize support using meaningful information about students’ learning processes obtained by trace data, the challenge remains to combine the analytics-based approach for the measurement of students’ actual ongoing SRL processes (i.e., SRL processes in real-time) with key features of (adaptive) scaffolding which also offer personalized support content. By that we mean, how can scaffolds be gradually calibrated during the learning process, both in terms of level of support and content of support, and faded when students are able to regulate their learning independently?

1.3. Understanding SRL processes with process mining

Analyzing the sequences of actions learners take while learning helps us understand SRL processes in addition to learning outcomes (Roll & Winne, 2015). Molenaar and Järvelä (2014) highlighted the need to empirically study the temporal characteristics of SRL to reveal how SRL dynamically unfolds over time. This is linked to the move of SRL research from a variable-centered approach to an event-based approach which assume SRL to occur as a sequence of events in temporal space (Molenaar, 2014; Reimann, 2009). In slight contrast to the learning analytics-based focus on the measurement SRL learning behavior, educational data mining approaches, such as process discovery modeling, focus on learning patterns (Romero & Ventura, 2015). Previous studies have modeled SRL patterns to gain deeper understanding of the temporal structures of SRL activities in relation to theoretical models and assumptions (Bannert et al., 2014; Cerezo, Bogarin, Esteban, & Romero, 2020; Huang & Lajoie, 2021; Maldonado-Mahauad, Pérez-Sanagustin, Kizilcec, Morales, & Munoz-Gama, 2018; Wong, Khalil, Baars, de Koning, & Paas, 2019). In order to understand why and how scaffolds worked or did not work, studies have further investigated the temporal and sequential structure of students’ SRL activities to comprehend the effects of scaffolds on the learning process, Sonnenberg and Bannert (2015) coded and analyzed think aloud protocols and found a mediating effect of monitoring activities on students’ transfer performance. Through process mining, they found better integration of preparatory activities (i.e., orientation, planning, and goal specification) and more links between cognitive and metacognitive activities in the scaffolded group. Likewise, Engelmann and Bannert (2019) were able to gain an insight as to why transfer performance did not improve when students were supported by metacognitive prompts, despite increase in frequency of metacognitive activities. Using coded think aloud protocols, they modeled SRL processes of students with and without metacognitive prompt support. They found little differences in how SRL activities were arranged in the process models between groups; they also found inconsistencies with theoretical SRL models, such as weak integration of monitoring activities. Nonetheless, they were able to detect finer differences like better integration of evaluation activities among students supported with prompts. Hsu, Wang, and Zhang (2017) investigated the effects of cognitive and metacognitive prompts on learning actions and outcomes using screen logs from ninth-grade students working on inquiry-based learning tasks. They found that successful students had higher scores and exhibited more metacognitive actions when prompted. They further examined how prompts led to differences in learning outcomes by looking at sequential patterns of metacognitive actions. They reported more sequential links between prompts and metacognitive actions in the successful group, such as more recurrence of evaluation and monitoring. These methods enable us to take a closer look into the learning process as well as the effects of scaffolds on the learning process as a means to optimize future scaffold interventions. Hence, the study presented focused on both optimizing SRL support within the learning session (i.e., analytics-based personalized scaffolds) as well as improving SRL support in future studies through insights obtained from process mining models.

1.4. The present study

As introduced, engagement in SRL can boost learning outcomes (Schunk & Greene, 2017). However, students need support to improve regulation of learning, especially in computer-based learning environments (Azeevedo, 2005; Broadbent & Poon, 2015). Using scaffolds to foster SRL has shown a potential to benefit students’ learning (Guo, 2022; Zheng, 2016). To optimally support students’ SRL, scaffolds such as prompts need to be inherently adaptive (i.e., they take into account students’ progress on a task and adapt accordingly) and personalized (i.e., customizing scaffold content to learners’ needs) to students’ learning progress. Using analytics-based approaches which combine different forms of trace data (i.e., navigation logs, keystrokes, mouse movements) offer a way to track students’ learning activities and progress on-the-fly, and, based on that, dynamically provide personalized scaffolds adapted to learners’ needs. Analytics-based measurement and support of SRL—which in this study focused on real-time detection and coding of learning activities in order to identify SRL gaps to offer personalized scaffolds (i.e., when and what to scaffold)—aimed to provide more optimal support within the learning session. Process mining, on the other hand, provided the possibility to understand the effects of personalized scaffolds on learning patterns post hoc in order to improve future scaffold interventions and adaptations of the study. On the whole, the study investigated the effects of the rule-based AI real-time measurement and support of micro-level SRL activities via analytics-based personalized scaffolds using an experimental research design (see Fig. 1).

To investigate the effects of personalized scaffolds on learning, we addressed the following research questions and hypotheses in the present study:

1) What are the effects of personalized scaffolding on students’ learning activities?

H1. Based on previous findings reporting increase in frequencies of SRL activities with scaffold support (Engelmann & Bannert, 2019; Sonnenberg & Bannert, 2015), we hypothesized that students supported by personalized scaffolds engage in higher frequencies of metacognitive learning activities than students supported by generalized scaffolds and students who did not receive scaffolds (H1a). Additionally, we hypothesized that students who received personalized and generalized scaffolds...
have higher frequencies of metacognitive learning activities than students who did not receive scaffolds (H1b). Furthermore, we examined the effects of (the type of) scaffolds on individual learning activities; we had no specific hypotheses regarding each learning activity but expected to find some differences between the groups because of the scaffold intervention. Since metacognitive activities promote deeper processing (Bannert et al., 2009; Mollenár & Chiu, 2015; in press van der Graaf et al., 2022 ), we expected that scaffolded groups, especially with personalized scaffolds, engage in higher frequencies of high cognitive activities than the group with no scaffolds.

2) What are the effects of personalized scaffolding on students’ learning performance?

H2. Prior research have largely pointed to SRL scaffolds improving learning outcomes (Guo, 2022; Zheng, 2016), especially in transfer tasks (Bannert & Mengelkamp, 2013; Bannert, Sonnenberg, et al., 2015; Lin & Lehman, 1999; Müller & Seufert, 2018). Additionally, we designed the scaffolds to support students’ learning within our learning task where the main goal was to write an essay. Hence, we hypothesized that personalized scaffolds have a positive effect on learning performance in comparison to generalized and no scaffolds (H2a), and besides, personalized and also generalized scaffolds may have positive effects on learning performance in comparison to no scaffolds (H2b).

3) What are the effects of personalized scaffolding on the temporal structure of self-regulated learning activities?

H3. Analysis methods such as process mining has been increasingly used in SRL research to complement statistical analyses in order to gain insights into the temporal and sequential nature of SRL beyond frequencies (Saint, Fan, Gasević, & Pardo, 2022). We considered the third research question to be more exploratory, hence we did not have specific hypotheses but rather, explored and compared the temporal structures of the groups by means of process discovery. Nevertheless, we state our general expectations. In general, we expected students to engage in substantial reading and writing activities (i.e., applying and integrating what they have learned) due to demands of the learning task. We also expected that monitoring plays a prominent role in groups that received scaffolds due to the support they received.

2. Method

2.1. Participants

A total of 104 university students from German universities participated voluntarily in our study. The participation criteria required students to have German as their first language and to be studying in university. As a result of technical errors (e.g., scaffolds triggered incorrectly; learning tools not working), we removed six observations from our dataset and thus obtained usable data for 98 participants (Mage = 23.45 years, SDage = 3.88 years, 70.4% female). We performed a preliminary screening of participants’ scores on the tests and questionnaires and further excluded four extreme outliers (i.e., one from the control group, two from the generalized group, and one from the personalized group)—participants with scores three or more three standard deviations from the mean. The final sample consisted of 94 participants. There were 49 students from Bachelor degree programs and 34 from Master degree programs. The remaining 11 students were enrolled in programs that did not fit in either category (e.g., medical program). Students came from more than 50 different degree majors such as architecture, business administration, education, engineering, health sciences, informatics, law, medicine, and philosophy. All participants gave active informed consent prior to data collection and received 20 euros for their participation.

2.2. Research design

In a pre-posttest between-subject design (see Fig. 1), students learned in one of three experimental conditions. In the personalized scaffolding condition (EG1, n = 35), students received personalized scaffolds during learning which adapted scaffold options based on their real-time learning processes. Students in experimental group 2 (EG2, n = 30) were prompted with generalized scaffolds during learning which contained standardized messages and options. Students of the control group (CG, n = 29) learned without scaffolds.

We conducted data collection by starting with control condition participants (CG), followed by generalized scaffold condition participants (EG2), and finally personalized scaffold condition (EG1) participants over a span of one year with multiple breaks in-between due to local COVID-19 restrictions. The main reason for collecting the data in this sequence of conditions was to develop, test, and refine the personalized scaffolding model. To elaborate, we started with the control group with no scaffolds as a so-called technical pilot. We subsequently performed post hoc simulations on both the control and generalized scaffold...
groups (i.e., CG and EG2) to check if scaffolds would have been triggered correctly during real-time processing of learning behavior for the personalized scaffold group (EG1), which had the most technically complex setup; the actual experimental sessions for CG and EG2 were not affected by this procedure as checks were solely related to technical aspects and performed either after the session or on the backend of the system. This procedure is referred to as adaptivity in the design and step loop, whereby new scaffold interventions are informed by empirical findings before design and implementation, and also during the learning task (Aleven, McLaughlin, Glenn, & Koedinger, 2016).

2.3. Learning environment and materials

All the participants learned using an online learning environment (see Fig. 2) with embedded learning tools and texts about three topics: artificial intelligence, differentiation in a classroom, and scaffolding in education. Their task was to write a 300-400-word essay about the future of education with only the help of the texts provided. We also included some pages of texts for each topic which were neither directly relevant for the learning goals nor the essay task (e.g., the history of artificial intelligence). Therefore, students had to learn strategically. The learning environment consisted of 16 pages of learning content, one instruction page, and an essay Rubric page. Students could navigate to each page via the navigation menu, read the texts using the reading area, and use the following learning tools: a) Annotation tool for highlighting, note-taking, and searching through previous annotations, b) Planner, c) Countdown timer, d) Search tool, e) Essay-writing tool which automatically saved the essay text throughout the session, and f) for EG1 and EG2, a prompt tool where students could access and interact with current and previous prompts and prompt checklists. The participants were free to navigate to any pages and use the learning tools in any desired manner. Interactions (e.g., page visits, mouse clicks, and keystrokes) within the learning environment were logged and additionally, for EG1, analyzed in real-time as input to the personalized scaffolding model.

2.4. Real-time analytics-based personalized scaffolds

2.4.1. Coding protocol of SRL activities using trace data

In our prior studies (Fan, Lim, et al., 2022; Fan, van der Graaf, et al., 2022; Lim et al., 2021; van der Graaf et al., 2022), we developed a theory- and data-driven SRL process library based on enriched trace data (i.e., based on multiple trace data sources) to parse raw trace data into patterns of learning actions which are then interpreted as SRL activities (see supplementary material for comprehensive coding protocol with lists of learning actions, Appendices A and B). To elaborate, in the two prior lab studies, we recorded multi-channel data (i.e., think aloud protocols and trace data which included navigational logs, peripheral data such as keystrokes and mouse movement, and eye gaze data). We segmented and coded the think aloud protocols using an adapted coding scheme from prior research (Bannert, 2007; Molenaa, van Bokxel, & Sleegers, 2011). According to their theoretical frameworks, SRL in computer-based and hypermedia environments can be categorized into Metacognition, Cognition, and Motivation. They further sub-divided the SRL activities into micro-level SRL activities (e.g., orientation, rereading etc.). The same theoretical framework was used to develop the SRL process library for trace data (i.e., the theory-driven aspect) with the exclusion of the Motivation category due to its high reliance on verbal data for coding. Using think aloud as a “reference point”, we then aligned the multi-channel data consisting of navigational logs enhanced with peripheral data (e.g., keystrokes and mouse movement) and eye-tracking data to test and compare the proportion of agreement of SRL processes as well as to redefine the processes in the process library (i.e., the data-driven aspect). The raw trace data logs were first categorized into meaningful learning actions (e.g., GENERAL_INSTRUCTION). Then, we detected SRL processes based on sequences of actions. An example of an Orientation activity coded from the action sequence was GENERAL_INSTRUCTION/RUBRIC to NAVIGATION to RELEVANT_READING. In this example, students performed orientation by first analyzing the task requirements in the general instruction page where information was provided about the essay task and also the learning goals. They then used the navigation panel to a learning text page which was relevant to the task and goals. This is in line with the preparatory stage of SRL. Fig. 3 shows an example of the process taken to code SRL activities from raw trace data. We also refer the reader to our prior work (Fan et al., 2022) on the development and validation process of our SRL process library.

Table 1 presents the coding categories and general descriptions of each SRL activity mapped in the categories. The coding protocol contained three major categories, Metacognition, Low Cognition, and High Cognition, with specific SRL activities defined in each category. The last category, Other, included actions which were not possible to map to an SRL activity, for example, isolated navigational actions (not part of a pattern) and interactions with scaffolds (e.g., clicking on scaffold options).

2.4.2. SRL scaffolds developed

We presented the SRL scaffolds (see Fig. 4) to EG1 and EG2 up to five times during learning as a pop-up box with a message and up to four suggested learning activity options. The content of the scaffolds was developed based on past literature and our analyses of two previous lab...
studies (Lim et al., 2021; van der Graaf et al., 2022) with the same tasks and content, which used a rule-based artificial intelligence approach to inform the development of the personalized scaffolds. Based on our studies, we determined key time intervals to present a scaffold. Specifically, we investigated the critical time points in the learning paths of students who were successful in the learning task. The learning paths were obtained by analyzing micro-level SRL activities (e.g., planning, orientation, and monitoring etc.) based on students’ trace data measured via an analytic-based measurement protocol. For example, although orientation is optimal in the start of the initial learning phase, students who were successful in the learning task additionally performed orientation activities in a relatively efficient manner (i.e., in the first 2 min) before switching to (first) reading activities (i.e., between minute two to seven). The scaffolds supported the following activities: a) Orientation (minute 2), b) Reading (minute 7), c) Monitoring of reading (minute 16), d) Writing (minute 21), and e) Monitoring of writing (minute 35). Each scaffold contained a message and up to four options which correspond to a suggested learning activity. In the personalized scaffolding condition (EG1), scaffold options were adapted to their learning progress; if students had demonstrated the engagement of specific learning activities (i.e., a sequence of learning actions like proceeding to reading the learning text after reading the instructions was coded as an orientation activity) matching the suggested learning activity, the option would be hidden; when all suggested learning activities were already performed, the scaffold would be

![Fig. 3. An example of how trace data was parsed to code SRL activities and a scaffold rule.](image)
skipped for the student. There were three categories of rules which formed the set of pre-determined rules used in the rule-based AI approach. The first two categories were implemented only for the personalized scaffold condition. First, we determined if gaps existed in students’ micro-level SRL. For instance, when students did not perform any orientation activities in the first 2 min of learning, the condition would be met to trigger a personalized scaffold. Second, we determined which specific learning activities were performed and which needed further support. For example, in the first 2 min of learning, when students checked the learning goals and instruction but not the essay rubric, which would be important for the learning task, the condition would be met to present the personalized scaffold option of “Check the essay rubric”). Thirdly, for both personalized and generalized groups, we determined breakpoints to present the scaffolds based on students’ ongoing learning activities as captured by the real-time analyses of trace data. We considered breakpoints to be a break in a continuous activity (e.g., typing continuously in the essay text field). The breakpoint analysis was included so the scaffolds would not unnecessarily interrupt students in the middle of an activity; the intention was to promote students’ scaffold use. In addition, all generalized and personalized scaffolds first appeared as a small notification icon, which students have the option to click on to open the scaffolds on their own. The scaffolds were subsequently presented in the middle of the screen whenever there was a breakpoint in the learning activities, or if 1 min has lapsed. For example, even though the orientation scaffold should appear at minute two, if a student is in the middle of typing a note, the scaffold would “wait” until the student switches activity (e.g., closing the note tool and navigating). In order to present the scaffolds within the window of time the support was needed, scaffolds would be queued for a maximum of 1 min if there were no breakpoints before automatically forced-displayed. Students in the generalized scaffolding condition (EG2) received all four options regardless of their learning progress. For both personalized and generalized scaffolds, a selection of a minimum of one and a maximum of three options were required for each scaffold. After each scaffold was displayed and the options selected, the scaffold was converted to a checklist in the prompt tool. There was also a small “x” button on each scaffold which students could click to close the scaffold prompt without any selection. In such cases, no checklist would be created and previous scaffolds cannot be revisited.

2.5. Procedure

We carried out the experiment sessions with participants individually in the laboratory with an experimenter present throughout. Each session lasted approximately two to two-and-a-half hours. We presented all study stimuli, including tests and the learning environment, on a 23.8-inch monitor using a web browser with participants having access to a keyboard and mouse at all times. All participants started with the pre-learning phase where they completed a demographic questionnaire and the pretest. Afterwards, they went through a training module to familiarize themselves with the learning environment, learning tools, and scaffolding mechanism (only for EG1 and EG2). The students were provided an example of a scaffold to illustrate how the scaffolding mechanism worked. The example contained generic text (i.e., “This is a prompt”) and not the actual scaffold content students received during the learning session. The students in the control group had the same training content with the exclusion of content related to the scaffold tool. If participants skipped any part of the training topics, the experimenter instructed them to return to complete those topics. The training took about 10–15 min. Prior to the start of the learning phase, the experimenter instructed the participants that they should work efficiently and decide what they read because they would not have

<table>
<thead>
<tr>
<th>Main categories</th>
<th>SRL activities</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metacognition</td>
<td>Orientation</td>
<td>Orientation on the learning-related activities; reading of general instruction and essay rubric.</td>
</tr>
<tr>
<td></td>
<td>Planning</td>
<td>Planning of the reading and writing process by arranging activities and determining strategies. Proceeding to the next topic.</td>
</tr>
<tr>
<td></td>
<td>Evaluation</td>
<td>Monitoring and checking the reading and writing process; checking of progress according to instruction or plan.</td>
</tr>
<tr>
<td></td>
<td>Monitoring</td>
<td>Evaluation of the learning process; checking of content-wise correctness (e.g., the essay content) of learning activities.</td>
</tr>
<tr>
<td>Low Cognition</td>
<td>First Time</td>
<td>Reading information from the materials and pictorial representations for the first time.</td>
</tr>
<tr>
<td></td>
<td>Reading</td>
<td>Rereading information in the text or figures.</td>
</tr>
<tr>
<td>High Cognition</td>
<td>Elaboration/</td>
<td>Elaborate by connecting content-related comments and concepts when writing the essay; organizing content by creating an overview; writing down information point by point in notes or essay window; summarizing; adding information generated by oneself; and editing information by rephrasing or integrating information with prior knowledge.</td>
</tr>
<tr>
<td></td>
<td>Organization</td>
<td>Elaboration/organization.</td>
</tr>
<tr>
<td>Others</td>
<td>Not applicable</td>
<td>Uncategorizable actions (e.g., short navigations) and scaffold interactions.</td>
</tr>
</tbody>
</table>

Fig. 4. An example of a personalized scaffold and checklist.
sufficient time to read all the texts and write the essay in 45 min. During the 45-min learning phase, participants were free to use any learning tools (timer, search tool, planner, annotation tool, essay writing tool) and navigate to any pages. When 45 min had passed, a pop-up dialog box prompted the participants that time was up and they were not allowed further interaction with the learning environment. Finally, in the post-learning phase, participants completed the posttest, transfer test, and metacognitive strategy knowledge questionnaire (MESH) respectively. Based on past studies using the same instruments and prior pilot tests, we fixed the maximum time limit for pre- and post-domain knowledge tests at 20 min and the transfer test at 30 min.

2.6. Instruments

2.6.1. Domain knowledge test

The domain knowledge test consisted of 30 mandatory multiple-choice items with four response options each ($\alpha = 0.61, \lambda_2 = 0.64, \omega = 0.62$). The reliability was acceptable (Kline, 2000). One point was given for each correct answer with a maximum of 30 points. The domain knowledge test addressed comprehension of the texts which were relevant to the learning goals. An example of an item was: “How can an algorithm work better?” with options, (A) “By making the series longer,” (B) “By building in more supervision,” (C) “By analyzing more data” (correct answer), and (D) “By simulating more human behavior.” Pre and posttest items were identical but the items were presented in a different order in the posttest.

2.6.2. Transfer test

We measured transfer knowledge with 26 mandatory multiple-choice items with four response options each. Students were asked to apply their knowledge to questions addressing artificial intelligence in the medical field, differentiation in the workplace, and scaffolding in sports. An example of a transfer test item was: “Which of the following describes how artificial intelligence has been used by the healthcare industry?” with options, (A): “Using augmented reality architecture systems to develop quicker and more efficient paths for transporting patients at the emergency department,” (B): “Using natural language processing to analyze thousands of medical papers for better informed treatment plans” (correct answer), (C): “Automatic transfer of patient information whenever another hospital requests for it,” and (D): “Using robots to prepare meals that meet patients’ treatment and dietary needs as indicated in the patient file.” Each correct answer was awarded one point. Three items which correlated negatively to the rest of the items were removed which led to the total score being 23 points. The test reliability was $\alpha = 0.52, \lambda_2 = 0.56, \omega = 0.53$, indicating a moderate reliability (Hinton, McMurray, & Brownlow, 2014).

2.6.3. Metacognitive strategy knowledge questionnaire (MESH)

We assessed metacognitive strategy knowledge using the Meta-cognitive Strategy Inventory for Hypermedia Learning (MESH) questionnaire (Bannert, Pieger, & Sonnenberg, 2015). Reliability was found to be good (Kline, 2000), $\alpha = 0.89, \lambda_2 = 0.89, \omega = 0.90$. The MESH questionnaire consisted of seven learning scenarios with five to six learning strategies; students gave a rating from one (very suitable) to six (not suitable) for each strategy taking the learning scenario into account. Scores were calculated for each learning scenario by comparing how students rated the suitability of strategies within each learning scenario. These comparisons were cross-checked with ratings from experts. When the ratings were in line with the expert rating, a maximum of two points was given—two points for an exact match with the expert rating, one point for a semi-match (e.g., one rates two specific strategies as being equally suitable). The lowest possible MESH score is zero and the highest possible MESH score is one. A high MESH score indicates a high level of metacognitive strategy knowledge.

2.6.4. Essay

The students’ main learning task was to integrate what they have learned from all three text topics into an essay where they envision and suggest how learning in schools would look like in the year 2035. They wrote the essay by typing their text directly in the essay tool. The essays were graded by five grading components: 1) Topic—how well each text topic was explained and applied in the essay, 2) Connection—how well the topics connected to the future of education, 3) Idea—suggestions of how each topic can be applied in future education, 4) Originality—how original the essay was (i.e., not plagiarized from learning material), 5) Word Count—how close the word count adheres to the requirement. All components were graded on a score between zero and three points, with the exception of topic score which had a maximum of three points per topic, with a total of nine points for the component. The total possible essay score was 21 points. Two trained coders graded the essays and inter-rater reliability (weighted $\kappa = 0.88$) was calculated by randomly selecting 25 essays. The weighted kappa value we have obtained represented excellent agreement between the coders (Fleiss, Levin, & Paik, 2003).

2.7. Data analysis

For our first research question about the effects on students’ learning activities, we conducted descriptive analysis on the counts of activities coded from the trace data during the learning session and tested for group differences in metacognitive activities using analysis of variance (ANOVA). For individual SRL activities, we conducted a multivariate analysis of variance (MANOVA). We corrected the significance level for subsequent separate univariate analyses on each SRL activity to .007. To examine the effect of personalized scaffolding on students’ learning performance for our second research question, we evaluated the differences between groups in posttest, transfer, and essay scores. For learning gain from pre to posttest, we conducted a mixed ANOVA with Time (pre/posttest) as within-subject-factor, and Condition (EG1/EG2/CG) as between-subject-factor. Then, we conducted one-way ANOVAs with transfer and essay scores for the testing of group differences. Although previous research indicated that strategic learning skills are important factors influencing learning achievement and the use of SRL and MESH scores would be included as a covariate, the MESH scores showed varying correlation with learning performance variables in each group and was deemed unsuitable to be included as a covariate in our analyses. For our third research question, we explored the effects of personalized scaffolds on the temporal structure of students’ learning activities by using the process mining analytical approach, specifically the process discovery approach (van der Aalst et al., 2012). We further compared the process model of the personalized scaffold group (EG1) with the generalized (EG2) and no scaffold group (CG). We first extracted the event logs of the SRL activities each student engaged in within the learning session. Each event log contained the case ID (i.e., student ID), activities (e.g., SRL activity such as Orientation), and the related timestamps of the start and end times of the activities. To create the process models, we performed stochastic process mining using the pMiner package (Gatta et al., 2017). The pMiner algorithm has the advantage of displaying micro-level process such as the SRL learning activities in this study in the form of probability metrics that can be interpreted with ease (Saint, Pan, Singh, Gasevic, & Pardo, 2021). Specifically, we trained and generated the first order Markov models (FOMMs) by executing the three-step procedure (load data → train model → plot model) within the pMiner algorithm using students’ SRL activities measured with trace data. Comparison models were visualized using the same procedure with the additional step of mapping the FOMM model of one group (i.e., EG1) to the other. The resulting process maps display nodes representing an SRL activity, directional arcs connecting nodes, and corresponding transition probabilities (TPs). We set up an inclusion threshold of 0.05, i.e., TPs below this threshold were excluded from the process map.
3. Results

For all statistical analyses, the alpha level was set to 5% except the Bonferroni-corrected multiple comparisons. We first checked if the assignment of participants resulted in the different conditions having similar learner characteristics. We found that the participants in the assignment of participants to the three conditions did not differ significantly on their prior knowledge, $F(2, 91) = 0.029$, $p = .971$, and metacognitive strategy knowledge (i.e., MESH scores), $F(2, 85) = 1.885$, $p = .158$. Hence, we determined that the assignment of participants to the three conditions did not result in unbalanced subsamples.

In the span of 45 min of learning, a total of 4254 activities were coded for EG1, 2976 activities for EG2, and 2598 patterns for CG. Table 2 presents the descriptive statistics for all activities coded in each group. As the category Other was not indicative of an SRL activity, all events coded as Other were not considered for further analyses. On average, EG1 students showed the greatest number of SRL activities ($M_{EG1} = 93$, $SD_{EG1} = 29.23$), followed by EG2 ($M_{EG2} = 79.07$, $SD_{EG2} = 23.68$), and the least number of SRL activities were shown by CG ($M_{CG} = 69.52$, $SD_{CG} = 23.52$).

3.1. Effects of personalized scaffolds on students’ learning activities

To test our first hypothesis, we investigated the effects of personalized scaffolding on students’ learning activities. We first focused on the aggregated frequency of metacognitive activities, then the individual SRL activities. As reported in Table 2, the descriptive differences between the groups in terms of frequency of metacognitive activities are in accordance with our expectation. EG1 had the highest mean frequency of metacognitive activities ($M_{EG1} = 26.11$, $SD_{EG1} = 10.35$). The mean frequency of metacognitive activities was lower in EG2 ($M_{EG2} = 23.07$, $SD_{EG2} = 8.73$) as compared to that of EG1. Both scaffold groups engaged in more metacognitive activities than CG ($M_{CG} = 20.45$, $SD_{CG} = 10.41$). However, a one-way ANOVA showed no statistically significant difference between groups in terms of metacognitive activities, $F(2, 91) = 2.630$, $p = .078$, $\eta^2 = 0.055$. Post hoc comparisons with Bonferroni correction revealed that although the test between EG1 and CG exceeded the threshold for statistical significance, there was a medium effect size ($\eta^2 = 0.074$, $d = 0.55$).

Next, we zoomed in on the individual SRL activities. The highest mean frequency of metacognitive activities in all groups was Monitoring. Orientation activities had similar frequencies between groups. All groups showed low frequency of Planning activities, and the lowest frequency of metacognitive activities for all groups was Evaluation. All groups showed low frequency of Planning activities, and the lowest frequency of metacognitive activities in all groups was Monitoring. The subsequent highest frequency was seen in High Cognition activities.

For our second research question, we tested the effects of personalized scaffolding on students’ learning performance. Table 3 shows the descriptive statistics of all learning measures. Despite some group differences detected in frequency of SRL activities as we reported above, we observed similar scores between groups across all learning measures. Students in all groups showed an increase in learning (i.e., posttest scores higher than pretest scores). EG1 and EG2 students showed slightly higher mean transfer scores than students of CG. Mean essay scores were similar for EG1 and CG and lowest for EG2 students. However, the differences observed were marginal between groups, and did not correspond to the differences we hypothesized (EG1 > EG2 > CG). A mixed ANOVA further supported our observations of the descriptive results that although there was a significant main effect of time (i.e., pre and posttest scores), $F(1, 91) = 164.53, p < .001$, $\eta^2 = 0.644$, there was no lowest, followed the same pattern within each group; EG1 had the highest frequencies in First Time Reading, then High Cognition, and Rereading, succeeded by EG2 in the same order, and lastly, CG. Despite the descriptive differences we reported, a MANOVA using Pillai’s trace showed no significant effect of (the type of) scaffolds on SRL activities, $V = 0.209$, $F(14, 172) = 1.436$, $p = .141$, $\eta^2_p = 0.105$. Separate univariate ANOVAs with Bonferroni correction (adjusted $p$-value = .007) revealed that scaffolds had a significant effect on High Cognition, $F(2, 91) = 6.848$, $p = .002$, $\eta^2 = 0.131$, and a marginally significant effect on Monitoring, $F(2, 91) = 5.124$, $p = .008$, $\eta^2 = 0.101$. Post hoc tests revealed significant differences between EG1 and CG in High Cognition ($MEG1 = 27.11$, $SDEG1 = 11.93$; $MCG = 18.52$, $SDCG = 6.29$) and Monitoring ($MEG1 = 15.74$, $SDEG1 = 7.25$; $MCG = 10.59$, $SDCG = 6.50$). Further investigation revealed a significant positive correlation ($r = 0.44$, $p = .008$) between monitoring and high cognition activities in EG1.

In summary, our first hypothesis was partially confirmed on the basis of the statistical test results. In addition, we note that supporting students’ learning with personalized scaffolds led to an increase in metacognitive activities, particularly Monitoring activities, but the differences between all groups were only detected descriptively. A significant difference in Monitoring and High Cognition was detected between students who received personalized scaffolds and no scaffolds.

3.2. Effects of personalized scaffolds on students’ learning performance

Table 2

<table>
<thead>
<tr>
<th>Activity</th>
<th>EG1: Personalized (n = 35)</th>
<th>EG2: Generalized (n = 30)</th>
<th>CG: No scaffolds (n = 29)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Metacognition</td>
<td>26.11</td>
<td>10.35</td>
<td>4</td>
</tr>
<tr>
<td>Orientation</td>
<td>8.23</td>
<td>4.28</td>
<td>1</td>
</tr>
<tr>
<td>Planning</td>
<td>1.49</td>
<td>2.13</td>
<td>0</td>
</tr>
<tr>
<td>Monitoring</td>
<td>15.74</td>
<td>7.25</td>
<td>0</td>
</tr>
<tr>
<td>Evaluation</td>
<td>0.66</td>
<td>1.08</td>
<td>0</td>
</tr>
<tr>
<td>High cognition</td>
<td>27.11</td>
<td>11.93</td>
<td>11</td>
</tr>
<tr>
<td>Low cognition</td>
<td>39.77</td>
<td>15.39</td>
<td>11</td>
</tr>
<tr>
<td>First reading</td>
<td>29.34</td>
<td>11.26</td>
<td>11</td>
</tr>
<tr>
<td>Rereading</td>
<td>10.43</td>
<td>9.32</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>28.54</td>
<td>10.51</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. a,b, ab points. a,b, ab points. a,b, ab points.
Fig. 5. Markov models (FOMMs) for each group based on students’ SRL activities.

Note. Panel A: FOMM of EG1. Panel B: FOMM of EG2. Panel C: FOMM of CG. Elliptical nodes represent SRL activities and annotated arcs represent transition paths and the corresponding transition probabilities. The blue box marks the four-node Markov chain of Monitoring and all cognitive activities.
significant interaction effect for group and time, \( F(2, 91) = 0.374, p = 0.689, \eta^2_p = 0.008 \). Due to minimal descriptive differences, ANOVA test results further showed no significant differences between groups for transfer knowledge, \( F(2, 91) = 0.920, p = 0.402, \eta^2_p = 0.02 \), and essay scores, \( F(2, 91) = 0.096, p = .909, \eta^2_p = 0.002 \). Hence, our second hypothesis was not supported by our findings.

3.3. Effects of personalized scaffolds on temporal structure of learning activities

For our third research question, we explored the effects of personalized scaffolds on the temporal structure of students’ learning activities. We present the FOMM models for each group in Fig. 5. The transition probability matrices are available in the supplementary document (Appendix C).

Students supported by personalized scaffolds started exclusively (i.e., \( TP = 1 \)) with Orientation activities. Similarly, the students in EG2 (\( TP = 1 \)) and CG (\( TP = 0.93 \)) appeared to have started the learning session with Orientation activities. Following Orientation, students in EG1 were most likely (\( TP = 0.21 \)) to move on to Monitoring. For example, the students may have read the general instruction and rubric pages and then proceed to check the timer. Fig. 5a illustrates that Monitoring is at the heart of the process map and can be seen as a transition step which preceded transitions to cognitive activities. There is a clear four-node Markov chain of Monitoring and all cognitive activities (see marked area in Fig. 5a). The probability of any of the cognitive activities following Monitoring were in general higher, in comparison to the rest of the transitions in the process map. An example of a scenario would be: Students read the text, monitored their reading, reread parts of the text after monitoring, and proceeded to write their essay or edit/label their notes. The same four-node Markov chain can be similarly observed in the process map of EG2, as we expected, but also in the process map of CG. There are no clear end activities in all the groups as indicated by the isolated “End” node. This means students showed highly varied end activities, unlike start activity, Orientation. In terms of self-loops (i.e., activities which are followed by the same activities), there are three activities with the highest self-loop probabilities in EG1: First Time Reading (\( TP = 0.55 \)), Orientation (\( TP = 0.43 \)), and High Cognition (\( TP = 0.4 \)). As we expected, similar to EG1, First Time Reading and High Cognition also appeared as prominent activities in the process maps of EG2 and CG due to substantial reading and writing in the learning task. Regarding Orientation, students may have switched back and forth from the general instructions page to the essay rubric page, as established by the high self-loop probability. Among the SRL activities observed in EG1, Planning and Evaluation stand out because there are no transitions to these activities, only transitions from these activities. This is likewise observed in the process maps of EG2 and CG. Students seemed to have engaged in these activities more spontaneously than they did with other activities which were more systematically grouped together. Additionally, we note that for Planning, the activities were labeled by students’ use of either the planner or search tool. However, not all students used these tools and the presence of this activity in the process maps reflects only the students who did use them.

To summarize, we found that the observations we made regarding
the overall structure of students’ learning activities in EG1 were generally true for EG2 and CG as well. We further investigated the finer differences in the process maps between EG1 and the other groups by generating comparison diagrams (see Fig. 6).

In the four-node Markov chain of monitoring and all cognitive activities, the comparison diagrams revealed three areas where there were differences between EG1 and the other groups. First, as shown in Fig. 6a, EG1 showed higher self-loop probability than CG for Monitoring (TP = 0.16; TP = 0.16; TP = 0.16; TP = 0.11) and High Cognition (TP = 0.43; TP = 0.43; TP = 0.37). As shown in Fig. 6b, EG1 showed higher self-loop probability than EG2 only in Monitoring (TP = 0.16; TP = 0.11; and High Cognition (TP = 0.43; TP = 0.43; TP = 0.33). Second, students in EG1 were more likely than students in CG to engage in Rereading activities after High Cognition (TP = 0.22; TP = 0.13). Thirdly, students in EG1 were less likely than EG2 to monitor after High Cognition (TP = 0.15; TP = 0.2). We also observed more variation between EG1 and the other groups regarding TP of activities following Planning and Evaluation. For example, students in EG1 were more likely to perform another preparatory activity, Orientation, or start reading after planning, while students in CG were more likely to monitor or perform high cognitive activities such as writing the essay.

As we expected, monitoring was well-integrated in the process maps of both scaffolding groups but we observed that this was also true in the control group with no scaffolds. As expected, there was a large emphasis in all groups on reading the text and high cognitive activities, such as writing the essay. Though there were some differences we observed in the frequency of SRL activities in RQ1, the overall temporal structure of SRL activities revealed underlying similarities in how the activities are connected. Nevertheless, we also observed the influence of the (type of) scaffolds on the likelihood of transitioning from one activity to another, which these activities are, and which activities are closer linked to each other.

4. Discussion

In the study presented, we tested analytics-based personalized scaffolds based on a rule-based AI approach which measured and supported students’ real-time learning. We investigated the effects of personalized scaffolding on students’ learning activities, learning performance, and temporal structure of learning activities. We compared the respective effects with a generalized scaffolding and a no scaffolding condition within the study.

4.1. Findings on the effects of analytics-based personalized scaffolds on learning activities

With regard to our first research question, we investigated the effects of personalized scaffolding on students’ learning activities by examining the frequency of SRL activities and our hypothesis was partly confirmed. Our findings indicated trends in descriptive statistics showing higher frequency of metacognitive activities during learning when students received personalized scaffolds in comparison to generalized scaffolds and no scaffolds. Furthermore, our results indicated that students who received scaffolds (personalized or generalized) performed more metacognitive activities than when they did not have scaffolds. The finding is in line with past scaffolding studies (Engelmann & Bannert, 2019; Hsu et al., 2017; Moos & Bonde, 2016; Sonnenberg & Bannert, 2015) whereby scaffold interventions stimulated more metacognitive activities. Based on the results, the personalized scaffolds induced metacognitive activities as we expected, but not at a sufficient magnitude to indicate an overall main significant difference in the omnibus test, rather only marginally in the post hoc comparison between the personalized scaffold and control groups. When looking at the individual SRL activities, we found minimal differences in all but one metacognitive activity—monitoring. Generally, the scaffolds induced more monitoring activities and differences were sufficient to be detected between students who received personalized scaffolds and no scaffolds. In addition to previous scaffolding studies whereby scaffold interventions resulted in increased monitoring frequencies (Engelmann & Bannert, 2019; Sonnenberg & Bannert, 2015), our results indicated that effects can differ between a more adaptive and individualized scaffold as opposed to a standardized scaffold. Thus, with respect to metacognitive activities, personalized scaffolds were most effective in promoting monitoring activities. Secondary to this finding, we found differences pertaining to cognitive activities between groups. Noticeably, there was a large emphasis on reading the learning text (for the first time) in all the groups, as indicated by the high frequency of activity. This can be seen as a result of the learning task, where students were required to read the text adequately in order to write the essay. Knowledge acquisition by means of reading the learning material is not a low cognitive activity per se and is necessary for forming a foundational knowledge necessary for deeper learning (Frey et al., 2017). We observed a significant difference in high cognitive activities, and similar to metacognitive activities, this difference was identified between the personalized scaffold group and control group. Past research has suggested that metacognitive activities, especially monitoring, corresponds with deeper processing (i.e., high cognition), also potentially critical for higher essay quality (Bannert & Mengelkamp, 2013; Molenaar & Chiu, 2015, 2017). The higher frequency of high cognitive activities in the personalized scaffold group could potentially have been associated with substantial increased monitoring.

4.2. Findings on effects of analytics-based personalized scaffolds on learning performance

With regard to our second research question, we analyzed differences in learning performance measured by students’ test scores to examine the effects of personalized scaffolds and our hypothesis was not confirmed. We found that students in all groups had a learning gain, though the improvement in scores was not due to the intervention. We found no group differences in transfer and essay performance. In contrast to previous scaffolding studies (Bannert & Mengelkamp, 2013; Bannert, Sonnenberg, et al., 2015; Lin & Lehman, 1999; Müller & Seufert, 2018) and against our expectations, we found that students who received scaffolds did not perform better. Furthermore, even though our intervention aimed to scaffold students on their essay task, they did not perform better in their essays. A possible explanation is that students who received scaffolds experienced additional cognitive load as a result of the scaffolds, thereby leading to less cognitive capacity to both learn and use the scaffolds appropriately. Automated processing requires time and substantial practice but also leads to less conscious effort when well-practiced (Sweller, Van Merrienboer, & Paas, 1998). Given that students do not interact with scaffolds on a day-to-day basis in their learning and would therefore need substantial conscious effort, dealing with five scaffolds in a span of 45-min might have been overwhelming for some students. Additionally, the scaffolds used in this study provided more extensive usage possibilities such as creating checklists and reviewing and editing previous scaffolds, which most likely took away some learning time. Yet, we observed that despite the potential overload and lesser time, learning performance remained stable in scaffolded groups; furthermore, the personalized scaffolds affected students’ learning activities. In order for students to be adept at using tools (i.e., scaffolds in this study) and recognize its value, they must first develop proficiency (Gisic, et al., 2017). Thus, we suggest that next iterations of studies with more complex support tools could include follow-up learning sessions which would allow sufficient practice with scaffolds and potentially, with the extended exposure, effects would be detected in learning outcomes (e.g., higher essay scores).

In addition, how students used the personalized scaffolds could explain our overall findings. Although personalized scaffolds were designed to improve overall use as they adapt and align to ongoing learning behavior, past studies have also indicated that students...
Fig. 6. Comparison diagrams of EG1 and CG, and EG1 and EG2.
Note. Panel A: FOMM comparison of EG1 and CG. Panel B: FOMM comparison of EG1 and EG2. Black arcs represent similar TPs in the compared models. Red arcs represent higher TP in EG1, while green arcs represent lower TP in EG1.
sometimes use scaffolds in a suboptimal manner (Bannert & Mengelkamp, 2013; Bannert, Sonnenberg, et al., 2015; Engelmann, Bannert, & Melzner, 2021; Moser et al., 2017). These studies found discrepancies in effects of scaffolds based on students’ compliance. Bannert and Mengelkamp (2013) found that only about half of the students in the experimental group carried out reflection and metacognitive prompts in the intended manner, even with prior training. Students who complied with the prompts were found to have had better transfer performance in contrast to students who did not. Despite overall benefits reported by scaffolding students with self-directed metacognitive prompts, Bannert, Sonnenberg, Mengelkamp, and Piegler (2015) found that students who used the prompts exactly as instructed performed better in transfer tests after a first learning session. Moser et al. (2017) observed that benefits on learning outcomes corresponded to appropriate and regular prompt use. Students who did not fulfill these criteria showed no improvements in their learning outcomes. Further, Engelmann et al. (2021) found that students interpreted prompts differently, which led to a different long-term effect on transfer performance. They coded think aloud protocols and were able to differentiate two distinct uses of prompts—reflection request or call to action. Students who reflected upon receiving prompts performed better in the transfer test in a subsequent learning session than students who merely enacted the activities stated in the prompts. In the present study, we incorporated multiple scaffold trigger rules to increase scaffold use. For example, scaffolds first appear as a notification to alert students that a scaffold would be presented. Scaffolds were displayed when students were not actively engaged (e.g., typing in the essay field) on learning. Irrespective of these incorporated features, there is likelihood that some students were not able to reap the benefits of the personalized scaffolds provided due to poor compliance. We elaborate on this issue in the limitations section.

4.3. Findings on the effects of analytics-based personalized scaffolds on temporal structure of learning activities

Our findings relating to the third research question provided a more comprehensive picture explaining the incongruence between increased frequency of learning activities and no effects on learning outcomes. Through the process maps generated, we found that while the frequencies of learning activities pointed to greater benefits of the personalized scaffolds, the overall temporal structures of learning activities for all groups were comparable. In Sonnenberg and Bannert’s (2015) work, they found differences between the sequential patterns of a group that received metacognitive prompts and the control group. They found less integration of preparatory activities and less regulation steps when students were not supported with metacognitive prompts. Conversely, our results showed that students with no support executed patterns of learning activities in a comparable manner as students who received support, which could explain why students in all groups had similar learning outcomes. In Engelmann and Bannert’s (2019) work, they found no effects of metacognitive prompts on transfer performance even though effects were seen in frequency of metacognitive events. When they inspected the process models of the intervention and control group, they found minimal differences, and additionally, poor integration of monitoring in the models. Our study partly corroborates their findings. We observed that higher frequency of SRL activities alone does not necessarily lead to superior learning outcomes. However, in our study, we found that all students, with or without support, integrated monitoring activities well in their learning process. Moreover, monitoring was at the core of all activities, especially cognitive activities. Roelle et al. (2017) found that students who engaged in metacognitive activities (e.g., comprehension monitoring) first prior to high cognitive activities (i.e., organization and elaboration) performed better in the posttest due to the enhanced quality of their organization activities, while quantity played an insignificant role. Although quality of activities in this study was not assessed, it is plausible that students in all groups performed similarly well due to metacognitive activities setting the stage for high cognitive activities, likely resulting in a higher quality of organization activities, regardless of the differences reported in frequencies between the groups. According to Bannert (2009), missing learning outcome effects could also be explained by control group students spontaneously deploying SRL strategies. This seems to be the case in the study presented. An explanation could lie in the design of the learning environment. Unlike completely open-ended environments, the learning environment used in this study had a semi-structured layout (e.g., clear navigational menu, limited content on each page) with specific learning tools (e.g., planner) added to support the learning process. Although students still had to make decisions on what and how to learn, with the help of the learning tools and environment, students could have been implicitly cued to regulate their learning. For example, the planner tool could have signaled to students to consider a plan. Taken together, we observed that while personalized scaffolds induced SRL activities, as we expected, they strengthened the activities that students were already performing. For example, students supported by personalized scaffolds had higher self-loop probabilities for Monitoring and High Cognition, indicating multiple iterations of these activities. Although it was not detrimental to learning, the iterations could have been unnecessary. Nevertheless, we continue to find similar trends with past research indicating that students engage in little evaluation and planning activities (Bannert et al., 2014; Engelmann & Bannert, 2019; Lim et al., 2021; Sonnenberg & Bannert, 2015). Evaluation and planning activities remain difficult to support, with low occurrences of these activities in our study. In spite of the fact that these activities remain rare, we observed transitions to multiple activities following these activities. Thus, the effects of personalized scaffolds on temporal structure of learning activities could be more distinct with more focused support of planning and evaluation activities in future studies.

4.4. Limitations and future research

In the study presented, we utilized an analytics-based approach using trace data to measure and support SRL with personalized scaffolds. Although there are advantages (e.g., unobtrusive to learning) to this approach, there are also limitations that need to be addressed and considered for future research. The present study had a limited sample size which needs to be taken into consideration in the interpretation of the results. However, the findings presented are meaningful due to the reasonable number of SRL activities (i.e., coded sequences of raw log activities, mouse clicks, and keystrokes) recorded per participant, and furthermore, the study utilized multiple performance assessment measures which were also used to compare with past studies. As suggested by past studies which investigated scaffold compliance, students who used scaffolds appropriately and reflectively gained from the support provided. In these studies, compliance was coded from think aloud protocols (e.g. (Engelmann et al., 2021), and written text (Moser et al., 2017), which permitted assessment of quality of scaffold use and how students interpreted the scaffolds (i.e., in a reflective way). Our study used trace data which offered only information about how students interacted with the learning environment and not how they reflected on the scaffolds. However, there have been efforts to evaluate compliance via trace data, such as by using scaffold interaction logs and eye tracking data (e.g., Lall et al., 2017). Lall et al. (2017) operationalized compliance via student responses to prompts and by mining eye gaze data. They found that learning was influenced by compliance to specific prompts. In the study we presented, compliance to prompts was not included as a criterion in the personalization of the scaffolds but this is currently under development for the next iteration of the study. Additionally, when students recognize the value and utility of the learning task, they are more likely to strive for it (Eccles & Wigfield, 2002). As our study was conducted in the laboratory, instead of an authentic setting where students fulfill curriculum requirements, students might have felt less reason and value to use the personalized scaffolds to improve their learning in the given task. Future studies could embed
personalized scaffolds within actual course tasks. Another limitation is that scaffold interactions are currently not considered (i.e., coded as Other) in the process library, though work is underway to include them. Taking scaffold interactions into account in the process library, the personalized scaffolding model could further cater to how students interacted with scaffolds and calibrate the support given accordingly.

The findings from the study have implications for future research and practice. Investigations into SRL behavior as well as personalized scaffold interventions also need to consider the learning context and task. As seen in the study presented, although scaffolds have been proposed to better support learning in general, having a more structured learning environment with learning tools could potentially already enhance SRL sufficiently within a limited duration, despite a more complex learning task. For example, Calisir and Gurel (2003) found that a structured learning environment could support learning, especially for students who lack prior knowledge. On the other hand, the more complex learning task (i.e., integrating information from multiple texts into a vision essay) and limited learning time could have driven students to tap on their SRL skills (i.e., control group students regulated in a similar manner to students who received scaffolds). Despite specific (personalized) scaffold options targeting planning and evaluation activities, we continue to see less effects on these activities. Whether these activities could be equally as and sufficiently supported with such interventions (e.g., prompts) as other activities, such as monitoring, needs to be investigated further. Furthermore, the analytics-based SRL measurement protocol requires additional fine-tuning when applied to different learning contexts (e.g., when using different learning tasks or materials). Regarding methodological implications and limitations, we developed a rule-based AI system which followed pre-determined rules. However, such systems are limited in the sense that they are highly dependent on context and rely on prior lab research as well as experts to shape the rules and are less capable to deal with unanticipated situations. One possible direction for future research is to capitalize on machine learning methods which adopt a bounded rationality approach (Simsek, 2020) by personalizing scaffolds to further adjust to unexpected situations (e.g., different learner attributes etc.). The findings reported also present some practical implications for future personalized scaffold design. To reduce potential overload and increase appropriate scaffold use, future personalized scaffolds can narrow down the suggested activities provided. This will likely reduce the time students need to process each scaffold in the limited learning time. For example, instead of offering a maximum of four suggested learning activity options, future personalized scaffolds can provide a maximum of two to three options. The more focused support could also increase planning and evaluation activities, which has been consistently shown to occur at low frequencies.

5. Conclusion

In conclusion, our study demonstrated how an analytics-based approach using a rule-based AI system measured and supported SRL in real-time. We investigated the effects of real-time analytics-based personalized scaffolds on learning activities, learning performance, and temporal structure of learning activities. We found that analytics-based personalized scaffolds induced SRL activities but had no effect on learning performance. By exploring the temporal structure of learning activities and comparing process models of students supported by personalized scaffolds with generalized scaffolds and a control group, we found similarities in the process models across the groups and that personalized scaffolds strengthened beneficial activities that students were already performing. Using trace data with an analytics-based real-time measurement and support approach is viable and sets the stage for future development of personalized scaffold models which incorporate machine learning techniques. Finally, the development and improvements to AI-based systems which integrate different fields of expertise (e.g., learning sciences, educational psychology, learning analytics, and artificial intelligence) require interdisciplinary collaboration for further advancement.

Credit author statement

Lyn Lim: Conceptualization of this study, Methodology, Development of instruments and materials, Data Collection and Analyses, Writing- Original Draft Maria Bannert: Supervision, conceptualization and design of project which the study is part of, acquisition of funding for project, Writing- Review and editing. Joep van der Graaf: Development of instruments, materials, and scaffolding model. Shaveen Singh: Development of learning environment and scaffolding model. Yizhou Fan: Development of process library and scaffolding model, Writing- Review and editing. Surya Surendranair: Development of learning environment. Mladen Rakovic: Writing- Review and editing. Inge Molenar: Conceptualization and design of project which the study is part of, acquisition of funding for project, Writing- Review and editing. Johanna Moore: Conceptualization and design of project which the study is part of, acquisition of funding for project, Writing- Review and editing. Domenyvan Galevicius: Conceptualization and design of project which the study is part of, acquisition of funding for project, Writing- Review and editing.

Acknowledgements

This work is a result of the FLORA research project funded by DFG (Germany), NWO (Netherlands), and ESRC (United Kingdom) as part of Open Research Area (ORA) B20144/10–1, NWO464.18.104, and ES/ S015701/1. We thank Linda Greiner and Laura Pflieger, the student research assistants of this project, for their valuable help.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2022.107547.

References


Roelle, J., Nowitzki, C., & Berthold, K. (2017). Do cognitive and metacognitive processes set the stage for each other? Learning and Instruction, 50, 54–64. https://doi.org/10.1016/j.learninstruc.2016.11.009


