How to design and evaluate personalized scaffolds for self-regulated learning

Joep van der Graaf[1] · Mladen Raković[2] · Yizhou Fan[3] · Lyn Lim[4] · Shaveen Singh[2] · Maria Bannert[4] · Dragan Gašević[2] · Inge Molenaar[1]

Abstract

Self-regulation is an essential skill for lifelong learning. Research has shown that self-regulated learning (SRL) leads to greater academic achievement and sustainable education, but students often struggle with SRL. Scaffolds are widely reported as an effective and efficient support method for SRL. To further improve digital scaffolds' effectiveness, real-time detection of learning behavior can be used to personalize scaffolds. Therefore, the present study aimed to inform the field of scaffolding SRL by reporting on the design and evaluation of digital scaffolds. We present decisions made during the design process of personalized scaffolds to inform future scaffold designs. We evaluated how scaffolds were personalized based on real-time detection SRL, how university students respond to the scaffolds (i.e., compliance), and how this response is related to learning outcomes (i.e., quality of an essay). The research design was a pre-posttest with three conditions (no, generalized (same for all), or personalized scaffolds). A 45-minute reading and writing task was used, during which SRL processes were recorded in real-time. Findings revealed that different real-time SRL processes could be used to personalize scaffolds, meaning that we were able to personalize the content of scaffold based on students' actual learning behavior. In addition, students in the personalized condition complied more with the scaffolds than students in the other conditions. This compliance with the scaffolds was generally associated with better learning outcomes. To conclude, our approach showed how design decisions could be evaluated and provided insight into the personalization of scaffolds.

Keywords Self-Regulated Learning · Personalized scaffolds · Design-based research · Real-time learning analytics · Educational technology

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

1 Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands

2 Centre for Learning Analytics, Monash University, Melbourne, Australia

3 Graduate School of Education, Peking University, Beijing, China

4 Technical University of Munich, Munich, Germany

Introduction

Lifelong learning has become more vital, given the expected acceleration in technology uptake. Self-regulated learning (SRL) is essential for lifelong learning (Schunk & Greene, 2018). SRL involves cognitive processes, e.g., reading and elaboration, and metacognitive processes, e.g., planning, monitoring, and controlling learning to make it more effective (Winne, 2018). Students often struggle with applying SRL during learning (Miller & Bernacki, 2019). Various approaches have been developed to support students' SRL, such as prompts, scaffolds, and virtual agents (van Merriënboer & de Bruin, 2019). Scaffolding a student means providing them with assistance on an as-needed basis (Wood et al., 1976) to support students in learning tasks they cannot accomplish by themselves (Hmelo-Silver & Azevedo, 2006; Sharma & Hannafin, 2007). Researchers have pursued so-called personalized scaffolds in an attempt to mimic human support, which has stressed the importance of measuring SRL during learning, via so-called learning analytics, to inform the digital scaffolding system (Crompton et al., 2020). To build such a digital scaffolding system, it would be helpful to know how scaffolds can change based on learners' SRL (i.e., personalization of scaffolds), how learners respond to scaffolds (i.e., compliance), and how compliance relates to learning outcomes. This way, it is possible to capture how personalization of scaffolds is associated with changes in SRL and learning outcomes (Schunk & Greene, 2018).

The feasibility of using real-time detection of SRL during learning has recently been shown (Siadaty et al., 2016). However, little is known about the design and evaluation of personalized scaffolds based on real-time SRL data. Therefore, we describe an infrastructure to personalize scaffolds based on real-time detection of SRL in the present study. The two aims of this paper were (1) to advance personalized scaffolding of SRL by reporting on the scaffold design process and (2) to evaluate scaffolding effects of generalized and personalized scaffolds on learners' SRL. Regarding the evaluation, we extend a previous study on the effects of generalized (same for all) and personalized scaffolds (personalized based on learners' learning process) on SRL and learning outcomes (Lim et al., 2023). We investigated the amount of personalization in the scaffolds, and the effect of personalized scaffolds on compliance and learning outcomes.

Self-Regulated Learning (SRL)

Multiple overlapping frameworks exist that conceptualize SRL (Panadero, 2017). Our conceptualization of SRL follows the definition by Winne and Hadwin (Winne, 1997, 2018; Winne & Hadwin, 1998): SRL is a process in which a learner, based on the task context, sets goals, executes strategies to reach set goals, and monitors and adapts strategies. The first phase is called task definition. A learner identifies the task's context and uses it to activate prior knowledge, beliefs, and/or strategies. The second phase is called goals and plans. A learner identifies and/or sets goals inherent to the task. The third phase is called tactics and strategies. A learner enacts tactics and strategies, such as reading or repeating information. The result of enacting tactics and strategies is called a product, which can be knowledge acquisition or a written essay depending on the goals. The final phase is adaptation, where the learner identifies what and how to adapt their learning. During all phases, a learner monitors their progress toward the set goals and controls their learning to stay on track toward achieving them (cf. Nelson & Narens, 1990).

It has been shown that students, who regulate their learning effectively, learn better (Richardson et al., 2012). The need for effective self-regulation has been further stressed by the widespread use of digital learning environments (Wong et al., 2019). SRL is especially relevant in digital learning environments because they are often more open-ended and, thus, require more regulation (Azevedo, 2005). For example, Lim et al., (2021) compared successful to less successful university students when reading informative texts and writing an essay in a digital learning environment. The groups were operationalized based on their transfer test performance. Results showed that successful students generally showed more metacognitive activities, specifically more monitoring, and that their metacognitive activities were better integrated with the cognitive activities (Lim et al., 2021). Although SRL is shown to be effective in promoting learning outcomes, students often do not spontaneously regulate their learning (Azevedo, 2005), stressing the need for scaffolds. For instance, students need help ignoring irrelevant information and monitoring their comprehension accurately (Jaeger & Wiley, 2014). Students struggle with allocating their study time effectively (Tekin, 2022). Thus, these behaviors can be used to develop scaffolds that foster relevant SRL activities.

Scaffolding SRL

Such scaffolds of SRL have been designed and studied to help students enact SRL and improve their learning performance (Azevedo & Hadwin, 2005). For example, providing metacognitive prompts increases students' regulation activities and improves their transfer test performance (Bannert & Reimann, 2012). It has been repeatedly shown that scaffolds can improve learning performance (Zheng, 2016). Meanwhile, the effectiveness of scaffolds depends on several other factors, including learner and task characteristics (Wong et al., 2019). This is something that humans are assumed to take into account when scaffolding (Azevedo et al., 2005). In contrast, digital scaffolding tends to rely on scaffolds that are fixed for all learners (so-called generalized scaffolds), which thus might not be the most effective.

Personalized scaffolds

There is a history of personalized scaffolds in human tutoring of SRL (e.g., Azevedo et al., 2008). It has been more difficult to personalize scaffolds in digital learning environment due to technological challenges, but progress has been made in measuring and coding SRL during learning with digital environments (Azevedo & Gašević, 2019). This progress has made to personalized feedback possible(Pardo et al., 2019) and inspired a framework for personalized SRL scaffolds (Munshi & Biswas, 2019).

Pardo and colleagues (2019) aimed to improve students' academic achievement by providing personalized feedback messages via e-mail. The message informed students about their learning progress in an online course, on which the feedback was based, and suggested what to do next. In order to provide such feedback, they developed a system to track learners engagement with videos and their performance scores on two types of exercises. This information was use to personalize the content of a feedback message, which was automatically sent to the students. Positive effects were reported on student satisfaction and midterm scores. The engagement and accuracy scores were transformed into quartile scores, which resulted in four subgroups of students per score. This means that the system calculated these scores at fixed times after completion of the activity, which is an effective way to personalize the feedback. However, such quartile grouping cannot be done when

using real-time process data. In addition, it might be that students who complied with the feedback message and showed the suggested behavior performed better on the midterm than students who did not comply (similar to Bannert et al., 2015).

Another study introduced a framework to personalize scaffolds of SRL in a digital learning environment called Betty's Brain (Munshi & Biswas, 2019). The proposal is to collect multimodal data (i.e., data from multiple sources) to capture SRL. For example, metacognitive information might be inferred from cognitive-affective inflection points. These are points where learners show a change in their cognitive and/or affective behavior. To our knowledge the authors have yet to report on experiments with this framework. It, therefore, remains to be investigated to what extent scaffolds would be personalized based on the SRL data. This framework, however, helps to consider the relationship between SRL processes and personalization by specifying which SRL processes might be used to personalize scaffolds, such as the use of cognitive-affective inflection points to trigger metacognitive scaffolds.

Personalized scaffolds take into account the self-regulated learning *process*, and not only the learning *progress*, and, therefore, take scaffolding a step further than so-called dynamic scaffolds (Molenaar et al., 2012) or adaptive scaffolds (Munshi et al., 2022). Both studies detect learning actions that are related to students' progress in the task. Based on these actions, students are scaffolded. Molenaar and colleagues (2012) triggered cognitive scaffolds when students asked for it by clicking a specific button and they triggered metacognitive scaffolds at specific timepoints in line with the Zimmerman's (2002) model. An orientation scaffold, for example, was triggered when students progressed to executing a sub-task. In Munshi and colleagues (2022) study, students were shown a scaffold when making progress to encourage them suggesting to take a quiz to confirm success, or when they did not show progress they received a strategic hint. Therefore, these scaffolds can be considered based on learning progress, which is progress in a task-dependent and task-specific learning trajectory. These scaffolds have been found to be effective in learning outcomes (e.g., Molenaar et al., 2012; Munshi et al., 2022). SRL scaffolds can be made more effective when they are personalized in the sense that the content of the scaffold is personalized to the self-regulated learning processes that are in fact executed by the student. We proposed detecting SRL processes in a fine-grained manner during learning to personalize scaffolds (Fan et al., 2022). An example is that a monitoring scaffold that suggests a full range of specific actions to execute to trigger monitoring, when we do not detect monitoring processes before the scaffold is fired. In case monitoring processes were detected, these f suggestions were reduced.

There are different ways to approach personalization. For example, Maier and Klotz (2022) categorize personalized feedback based on (1) which feedback to assign to whom (rule-based and/or artificial intelligence), (2) which learner characteristics to personalize to (individual goals, current knowledge, progress measures, learning behavior, and/or emotional/motivational state), and (3) what parts of the feedback are personalized (evaluative parts, informative part, or both parts).

To address the design implications from these categories, we propose an integrated approach: who gets which scaffold is based on their SRL process (categories 1 and 2). Based on real-time detection of SRL processes, we change the options to enact control processes in the scaffold (category 3). For real-time processing, unobtrusive measurement of log data is needed (Azevedo & Gašević, 2019). Thanks to studies on interpretation of log data based on SRL theory (Siadaty et al., 2016) and studies on validation of log data with think-aloud (Fan et al., 2022), it has become possible to process log data real-time. Automatization of this process has already been done for SRL feedback (Pardo et al., 2019), but has yet to be done for SRL scaffolds.

Personalizing content of the scaffolds (category 3) based on SRL processes is promising. Personalized scaffolds help learners adapt via scaffolding the detection of discrepancies between standards and experiences (Winne, 2019). For example, by taking into account a learner's experience, such as reading the instruction (which is an indication of orientation), it can be assumed that the standard of orienting on the task might be met. In this example, the personalized scaffold would not suggest the learner to orient because it was already done. In contrast, when a learner has not oriented based on their SRL process, then the learner might choose to adapt. Adaptation can be done by replacing a previously applied operation, such as reading informative texts, with a new one, such as reading the instruction, or modifying a previously applied cognitive routine by reconsidering conditions, operations, or both (Winne, 2019). Personalized scaffolds can suggest adaptations relevant to the learner based on their SRL process, such as reading the instruction.

The current study

The aims of the current study were (1) to report on the process of designing personalized scaffolds and thereafter, (2) analyzing the effects of personalized scaffolds. The design process was informed by literature and our previous studies. The following research questions were formulated:

1. What is the amount of personalization of personalized scaffolds?
2. What is the difference in the level of compliance between students provided with generalized scaffolds and those with personalized scaffolds?
3. What is the effect of being able to select suggested actions in the scaffolds on compliance
4. What is the effect of compliance on essay scores?

With regards to the first research question, we personalized scaffolds based on learners' learning process, which means that the scaffolds differed between students. We analyzed the degree of personalization of the scaffolds in the first research question. The degree of personalization is operationalized as the proportion of participants for which a specific scaffold option was not displayed. The degree of personalization was calculated for each option of each scaffold. Our five scaffolds had four options each that suggested execution of a SRL process. Whenever processes were detected in the period before the scaffold was triggered, the scaffold content was personalized by not displaying the option associated with the already executed learning process. No hypotheses could be formulated here, because this is an exploration related to design effects.

Second, compliance was operationalized as executing the action suggested in the scaffold (cf. Bannert et al., 2015). We compared generalized and personalized scaffolds with a control group without scaffolds. It was hypothesized that compliance would be highest in the personalized condition and higher in the generalized than the control condition.

In research questions 3, we investigated the potential effect of a specific feature of our scaffolds on compliance. Our scaffolds offered the opportunity to create a checklist in which suggested options could be listed by selecting them. No hypotheses could be formulated here, because this is an exploration related to design effects.

Finally, we assessed the effect of compliance on a learning outcomes, which was the essay score. Students were tasked to write an essay in a learning environment with informative texts in a time of 45 min. It was hypothesized that the essay score would be highest in the personalized condition and higher in the generalized than the control condition.

Context of the current study

In this paper, we report on a study with personalized scaffolds that was part of a bigger project. Each study in this project informed the next. In each study, we used the same learning task and tested the same population, which was university students. We developed a digital scaffolding system to capture SRL in real-time and to decide what and how to scaffold whom in this project. The first step in the project was creating a baseline for SRL in a specific context (Lim et al., 2021). Students were tasked to read texts and write an essay in 45 min. The learning goal was to write a vision on the future of education with a role for scaffolding, differentiated instruction, and artificial intelligence. The texts were about these topics and texts were added that were not directly related to the learning goal to make the learning environment more open and trigger SRL. This meant that students' most frequent activities were reading and writing. In the learning environment, which main functionalities remained similar but was developed throughout the project, students could navigate via a menu and access several tools, see Fig. 1. These tools included a search tool to search for information in the texts, an annotation tool to make highlights and take notes, a scaffold button which showed previous scaffolds, a timer which showed the time left, and a planner where students could plan their activities. Students' think-aloud was recorded and coded as SRL processes. Log data was also captured, which included mouse clicks and keyboard strokes. The results showed that successful students, as indicated by a higher transfer test performance, showed more metacognitive processes during learning, especially monitoring, and more high cognition compared to the less successful students. Furthermore, the successful students showed better-integrated metacognition.

Then, we developed an algorithm to label log data as SRL processes (Fan et al., 2022). Raw logs were first labeled as actions, such as relevant reading, and actions were subsequently coded into patterns that reflect SRL. For example, the subsequent actions of irrelevant reading, general instruction, and relevant reading were coded as monitoring. Think-aloud was used as a reference, and the pattern labels were improved using a data-driven approach. The development of the algorithm resulted in a match rate of about 55% between log data and think-aloud. Based on this result, we decided to continue with development of real-time processing of log data to be able to personalize scaffolds.

In the next study, we evaluated how the SRL processes were related to writing a good essay (van der Graaf et al., 2022). SRL processes related to essay quality were high

Fig. 1 The digital learning environment

cognition (i.e., elaboration and organization) and processing (a low-level cognitive process related to rereading or editing own products). Furthermore, an indirect effect of prior meta-cognitive knowledge was found: those with higher metacognitive knowledge, read less and wrote better essays than those with lower metacognitive knowledge. This result suggested that metacognitive knowledge helps to deploy tactics to meet the standards, i.e., a good essay.

Designing scaffolds

Upon these previous studies and existing infrastructure, we designed scaffolds (aim 1). The decisions in our design were specifications of the theoretical framework outlined in the introduction. We added five scaffolds to our existing learning task of 45 min. Each scaffold had a generalized and personalized version. The learning goal of the task remained the same: to write an essay about the future of education. We did not collect think-aloud, but log data was recorded for real-time detection of SRL processes. The design of scaffolds was based on our previous studies. We re-analyzed parts of the data. However, mostly our decisions were informed by SRL theory (Winne, 2019) and supplemented with Multimedia Learning theory (Clark & Mayer, 2016). We made the following decisions in designing our scaffolds:

1. Support was provided utilizing scaffolds.
2. The scaffolds took over monitoring from students.
3. The scaffolds suggested concrete actions to enact control.
4. Each scaffold had a specific purpose.
5. Five scaffolds were delivered.
6. The scaffolds had specific timings.
7. Scaffolds displayed four options for actions.
8. The amount of text in a scaffold was limited.
9. The scaffolds were personalized based on the actual SRL process.
10. We triggered scaffolds in breaks during learning based on a breakpoint analysis.
11. A notification that a scaffold was ready to be triggered appeared before the scaffold.
12. We offered the opportunity to create a checklist of selected actions.
13. Previous scaffolds and checklists could be revisited.

We provide support to students via scaffolds (*Decision 1*). We categorize our support as scaffolds because we provide assistance on an as needed-based (Wood et al., 1976). Furthermore, our scaffolds took over part of SRL, namely monitoring, and triggered the other part, namely enacting control (Nelson & Narens, 1990). Monitoring was taken over by explaining which SRL process was relevant at that time, such as stating the need to start reading in our second scaffold (*Decision 2*). In addition to an explanation, our scaffolds displayed four suggestions for students to enact control, such as selecting what to read (*Decision 3*). Our scaffolds took over monitoring because students struggle with SRL (e.g., Tekin, 2022) and suggested concrete actions to problems with enacting control (Winne, 1997).

Five scaffolds were offered to the students (*Decision 4*). In a previous study, eight scaffolds were provided in 40 min (Bannert et al., 2015). The results showed a beneficial effect of scaffolds when students complied. The authors suggest that the low compliance might be because some students felt disturbed by the scaffolds. In an earlier study (Bannert &

Reimann, 2012), three scaffolds were delivered in either 45 min (Experiment 1) or 35 min (Experiment 2). The scaffolds aimed to foster (1) orientation, (2) monitoring, and (3) evaluation. Scaffolds were associated with a change in SRL (Experiments 1 and 2) and a higher transfer test score (Experiment 2). These findings, in combination with the 45 min in our task, suggested using four or five scaffolds. To proceed with five scaffolds was based on the scaffolds' purpose (*Decision 5*) and the timing (*Decision 6*).

We have developed five scaffolds, each with a specific purpose (*Decision 5*): (1) orientation, (2) reading, (3) monitoring of reading, (4) writing, and (5) monitoring of writing. These align with the cycle of SRL (Winne, 1997) and with the learning goal of the present task, which was to write an essay based on the texts. The cycle of SRL also has implications for the timing of scaffolds based on their purpose. Regarding the timing of the scaffolds (*Decision 6*), we analyzed data from our previous study (van der Graaf et al., 2022). The analyses identified when specific SRL processes should be executed by investigating three groups of students those with a poor essay, average essay, and good essay. Then, we analyzed the accumulative duration of SRL processes over time based on think-aloud. An example can be found in Fig. 2. The figure shows that students with good essays orientated early and stopped orienting earlier than those with average or poor essays. The analyses showed that those with a good essay (1) orientated early and stopped orienting earlier, (2) started reading earlier, (3) monitored more during reading, (4) started writing earlier, and (5) monitored more during writing. Based on how the good essay group learned compared to the others, we determined the timings: orientation at 2 min; reading at 7 min; monitoring of reading at 16 min; writing at 21 min; and monitoring of writing at 35 min. We did not see clear differences between the groups on other SRL processes.

Previous studies about self-regulated learning and recommendation/feedback systems emphasized the importance of actionable insight: learners need to be supported to enhance the quality of their work or learning strategies by conducting specific learning actions (Du & Hew, 2022; Matcha et al., 2020). Each of our scaffolds displayed four suggested options based on our lab study (*Decision 7*). For example, in the first scaffold, which was related to the SRL process of orientation, learners who used certain learning tools (e.g., use the
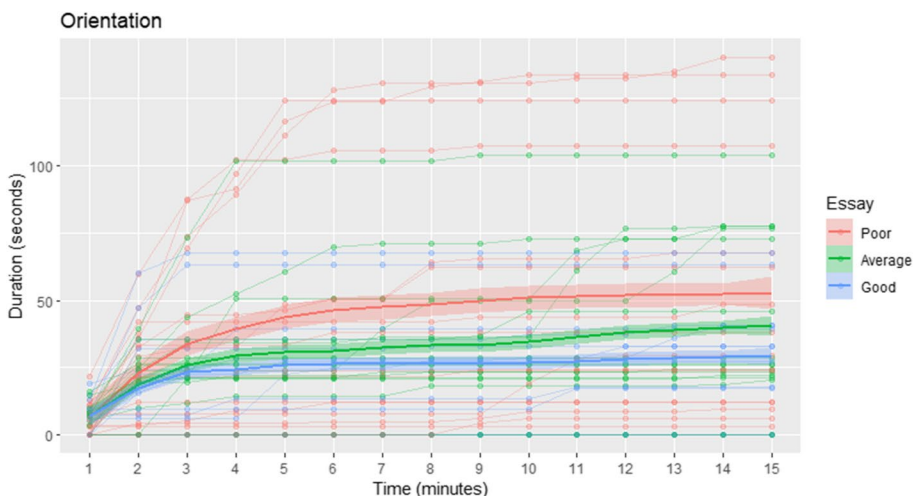

Fig. 2 Duration of orientation over time for three groups of essay scores

annotation tool to take notes about learning goals in the beginning) or performed specific learning actions (e.g., have a quick overview of the reading materials using the navigation zone) outperformed the other students. This decision process resulted in four suggested options in each of the five scaffolds, see Table 1.

Following the recommendations in (Clark & Mayer, 2016), we limited the amount of extraneous text provided in scaffolds (*Decision 8*) to control for working memory demands and cognitive load (Paas et al., 2004) imposed on learners at the moment when they receive scaffolds. Moreover, controlling for the learner's cognitive load may be particularly critical in cognitively demanding tasks, such as writing based on multiple source texts. In each scaffold, we thus provided a concise summary (no more than 11 words) of what a learner was advised to do and a concise instruction (no more than 12 words) suggesting appropriate learning actions. The number of learning actions suggested in a prompt was also limited, i.e., in each scaffold, we suggested up to four actions to learners, see Fig. 3. Figure 3 has four options and this is what all students in the generalized condition saw. Students in the personalized condition saw this scaffold, if they did not already perform one of these options, see *Decision 9*. It was possible to identify four relevant actions for each purpose and, thus, for each scaffold. Each action represented a pattern in the log data (see Fan et al., 2022, for details).

We decided to create a so-called personalized condition, in which students received scaffolds personalized in real-time based on their SRL process (*Decision 9*). The reasons to create personalized scaffolds were to prevent scaffolds from being experienced as disturbing (Bannert et al., 2015) and to make them more directive and thus less cognitively demanding (Paas et al., 2004). We personalized the scaffolds by removing any suggested option that was already performed. For example, in the orientation scaffolds, the option of checking the learning goal and instruction was not displayed when students spent at least five seconds on the instruction page. For Fig. 3, this would mean that the first (top-left) option was removed. We created rules for each option of each scaffold. Importantly, we decided that only actions executed after the previous scaffold (or after that start for the first scaffold) were taken into account. These actions were assumed to be executed to attain specific goals at specific times, in line with Decisions 5 and 6 about the purpose and timing of the scaffolds. Students in the personalized condition could see zero to four options in their scaffolds; when zero options remained, the scaffold was not displayed.

A large number of studies have shown that although scaffolds can play a positive role in learners' learning, poorly designed scaffolds often interrupt or interfere with learners' SRL process, which in turn causes learners' dislike and poor user experience (Álvarez et al., 2022; Munshi et al., 2022; Shih et al., 2010). Therefore, the scaffold window did not pop up directly at the triggering time in our design. Instead, we first displayed an unread envelope button (*Decision 10*) in the lower right corner of the learning interface to get learners' attention at the triggering time (e.g., minute 2) and remind them that there is an unread scaffolding suggestion. Then, we found a suitable time to pop up this scaffold window in the next minute using breakpoint analysis (*Decision 11*). This breakpoint analysis (Molenaar & Roda, 2008) found natural breakpoints in the learning process (e.g., closing a learning tool, saving a note or a highlight, or finishing writing one sentence and saving the essay) and when scaffolds would be provided at these times, students were expected to experience the lowest feeling of being interrupted. If we still could not find such a breakpoint within one minute, the scaffold window was presented when the one minute was up. At this time, the students had already had time to mentally prepare for the scaffolds (from seeing the unread envelope button). In this way, we triggered the scaffolding on time but minimized interruption.

**Table 1** Suggested options per scaffold

Scaffold	Option 1	Option 2	Option 3	Option 4
1. Orientation	Check the learning goals and instructions	Check the essay rubric	Use the menu to get an overview and skim the text	Process information by taking notes
2. Reading	Select what to read	Search for (specific) information	Note down important information	Check the time left
3. Monitoring of reading	Review annotations to check learning so far	Check the essay rubric	Check essay to determine what to read next	Review the learning goals and instructions
4. Writing	Draft essay by transferring learning to main points	Review the essay rubric	Check the remaining time	Write the essay with help from notes
5. Monitoring of writing	Review the essay rubric	Check the timer to manage your time	Edit your essay	Check the learning goals and instructions
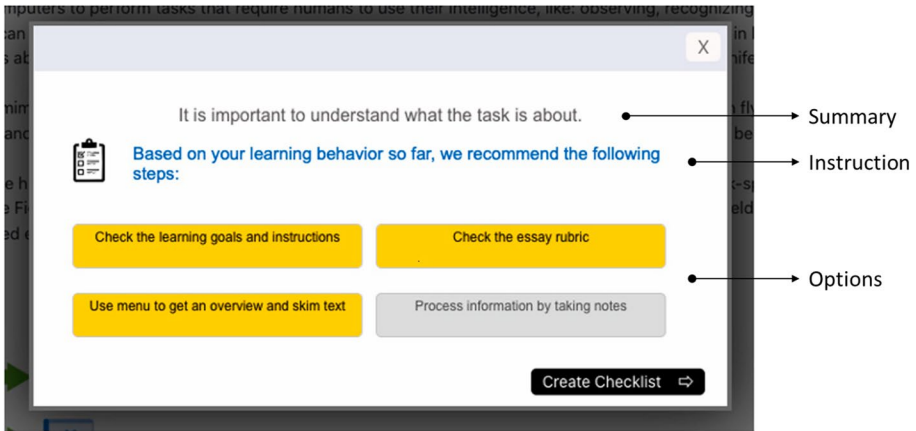
Fig. 3 The summary, instruction, and options of the orientation scaffold; yellow indicates options that were selected

Learners were afforded the opportunity to select some or all of the learning actions suggested by the scaffold (the yellow options in Fig. 3 were selected), rank them by priority, and assemble a learning plan in the form of a checklist, see Fig. 4 *(Decision 12)*. By doing so, we aimed at encouraging learners to interact with prompts and engage in planning and goal setting, a group of self-regulatory learning processes that have been widely documented to benefit motivation and learning performance (Bowman et al., 2020; Schippers et al., 2015; Schunk & Rice, 1991). Learners could also revisit their checklists at any time, re-arrange the actions in the list to change the priority, and cross off the actions that have been completed *(Decision 13)*. In this way, learners could engage in metacognitive monitoring of their progress towards goals and plans, another set of self-regulatory learning processes that have been shown to boost motivation and learning performance (Schunk,

Fig. 4 The checklist with one option crossed off
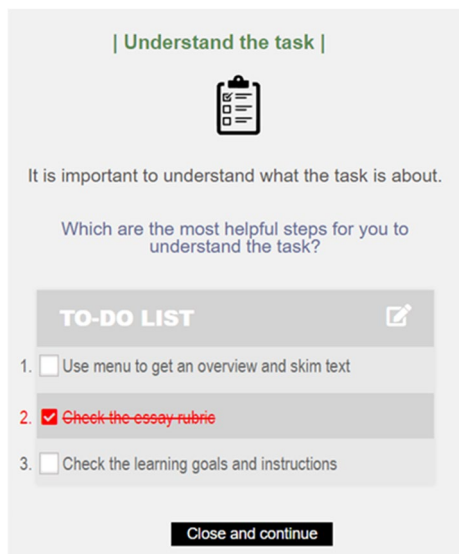
2003). A crossed-off action appeared as red, strikethrough text with a ticked box to make clear that the action was completed.

Methods

The method is the same as a previous study (Lim et al., 2023), because we re-analyzed the data to address the four research questions. The sample size is slightly smaller, due to technical errors. Note that the sample used to report current findings changes depending on the specific question. Sample sizes are reported accompanying the respective results.

Participants

The participants were 94 students (aged 23.45, *SD* age 3.88, 70% female) from German universities. The criteria for participation required students to use German as their first language and to study at the university. There were 49 Bachelor's students and 34 Master's students. The remaining 11 students enrolled in programs that were not suitable for any category (e.g., medical programs).Students came from more than 50 different fields, including business management, and philosophy. Participants actively consented and received 20 euros for participation. Due to technical errors leading to data loss, we continued analyses with 81 students.

Design of the study

In a pre-post-test design, the students learned under one of the three conditions. All students performed the same task, the essay writing task of 45 min. In the control condition, no scaffolds were presented. In the generalized condition, students received a scaffold that was the same for everyone, see the previous section. In the personalized condition, the options in the scaffold were adapted based on learners' learning process, see *Decision 9*. In the pretest and posttest, students filled in questionnaires regarding demographics, prior domain knowledge, and metacognitive knowledge. These were not used in the current analyses.

Learning task

Students were instructed to write an essay about the future of education in 45 min. They were given texts about Artificial Intelligence (AI), differentiation in education, and scaffolding of learning. They could navigate through the texts via a menu and had access to a list of tools, including an essay box, see Fig. 1. In addition to the texts, there was a page with detailed instructions and a page with a grading rubric.

SRL processes

During the task, keyboard strokes and mouse clicks were recorded. This raw data was labelled as learning actions, for example *General Instruction* when it was a click to navigate to the instruction page. Then, actions were combined into patterns that were the SRL processes used in the analyses. Patterns could indicate for example, *Orientation*, if it was

the following sequence of actions: *General Instruction*, *Navigation*, *Reading*. Please refer to our previous work (Fan et al., 2022) for more information on the labelling of SRL processes.

Essay score

The essays were rated by five grade components: (1) Themes: Quality of the explanation and application of each text's theme, (2) Connections: Quality of the connection of the themes to the future of education, (3) Ideas: Quality of the suggested applications of each theme in future education, (4) Originality: Extent of how original the essay was in comparison to the provided texts, (5) Words: Extent to which the essay length complied with the requirements (300–400 words). All components were rated between zero and three points, except for Themes with a maximum score of three points per theme and a total of nine points for this component. The total number of points obtained by the test was 21 points. Two trained coders graded the essays. The inter-rater reliability (weighted $\kappa = 0.88$) represents an excellent agreement (Fleiss et al., 2003).

Analyses

We conducted four analyses to test our scaffold decisions. First, the amount of personalization of the scaffolds was analyzed. Second, we analyzed compliance (i.e., whether students executed the suggested action of particular options in the scaffolds). Students in the control condition could not comply because they did not receive scaffolds. We analyzed their behavior in the same time windows as we would expect a compliant action in the other conditions. We can, therefore, not refer to compliance in the control condition, but rather use this measurement as an indicator of what happens in the specific time windows regardless of the scaffolds. Students in the personalized condition could receive less than four options based on their behavior. Therefore, compliance was analyzed in the personalized condition only when the particular scaffold option was displayed. In this condition, we chose to stick to compliance as executing the suggested action. This means that we did focus on actions that were suggested and therefore, disregarding the actions that were not suggested due to the personalization of the scaffold. Third, we analyzed to what extent compliance was related to selecting or not selecting a particular option in the scaffold. Fourth, we compared essay scores between students who did with those who did not comply.

Results

Before providing the additional analyses, the previous analyses (Lim et al., 2023) are summarized: Learning outcomes did not differ, but frequencies of two SRL processes did. The personalized condition showed more high cognition and monitoring than the control condition, but there was no effect on the temporal structure of the overall SRL process. Students in all conditions seemed to integrate monitoring well in their SRL process, especially the cognitive activities. This result might explain why all conditions performed well on the knowledge tests.

We conducted four additional analyses to test our scaffold decisions (aim 2): three descriptive analyses and one exploratory statistical analysis. The first analysis was a frequency count of the personalization of the scaffolds in the personalized condition. Options

of the scaffolds were not displayed when the learner had already executed the specific action preceding the scaffold. It was counted per option and scaffold how often an option was displayed or not displayed. This analysis was done with data from 18 participants in the personalized condition who had complete scaffold usage data. There were large differences in how often options were displayed between the scaffolds and options, see Fig. 5. Scaffold 3 was personalized the least. Options 2, 3, and 4 were displayed to all participants in the personalized condition. This showed that none of the participants executed the learning process that was suggested in options 2, 3, and 4 before scaffold 3. In other words, these options were not personalized, because they were displayed to all participants. On the other hand, scaffolds 2 and 5 had two options that were not displayed for almost all participants in the personalized condition. This showed that most of the participants executed the learning process that was suggested in these options before the scaffolds were presented. In other words, these options were personalized, because they were not displayed to most participants. A pattern of variation in the amount of personalization between these two

Fig. 5 Personalization of the scaffolds per scaffold and per option indicating how frequent options were not displayed as part of the personalization in the personalized condition

extremes emerged, suggesting that there is a variety in the execution of learning processes that were used to personalize the content of the scaffold.

Second, compliance in the control ($n=30$), generalized ($n=33$), and personalized ($n=18$) conditions was analyzed. In Fig. 6, the proportion of students that complied per scaffold and option can be found. Please note that in the control condition, the analyses addressed regular studying behavior, as students in this condition did not receive scaffolds. In the personalized condition, compliance is specified as compliance with the personalized scaffolds, taking only the actions that were suggested into account in the analyses. Compliance was measured by recording students' actions after the scaffold. Compliance was analyzed as not compliant (red color), late compliance any time after the next scaffold (green color), and immediate compliance after the current and before the next scaffold (blue color). Immediate compliance is what is generally desired, as it indicates performing the suggested action after encountering the scaffold and before seeing a new one. The results indicated that the conditions showed similar compliance overall, with differences in
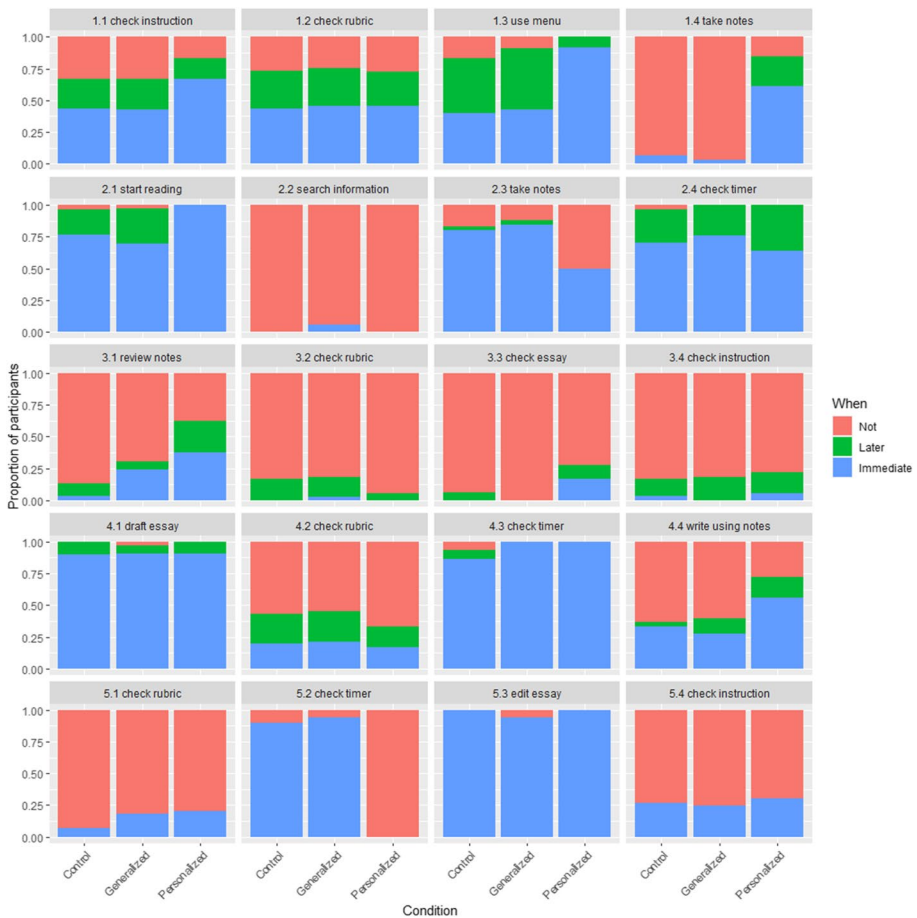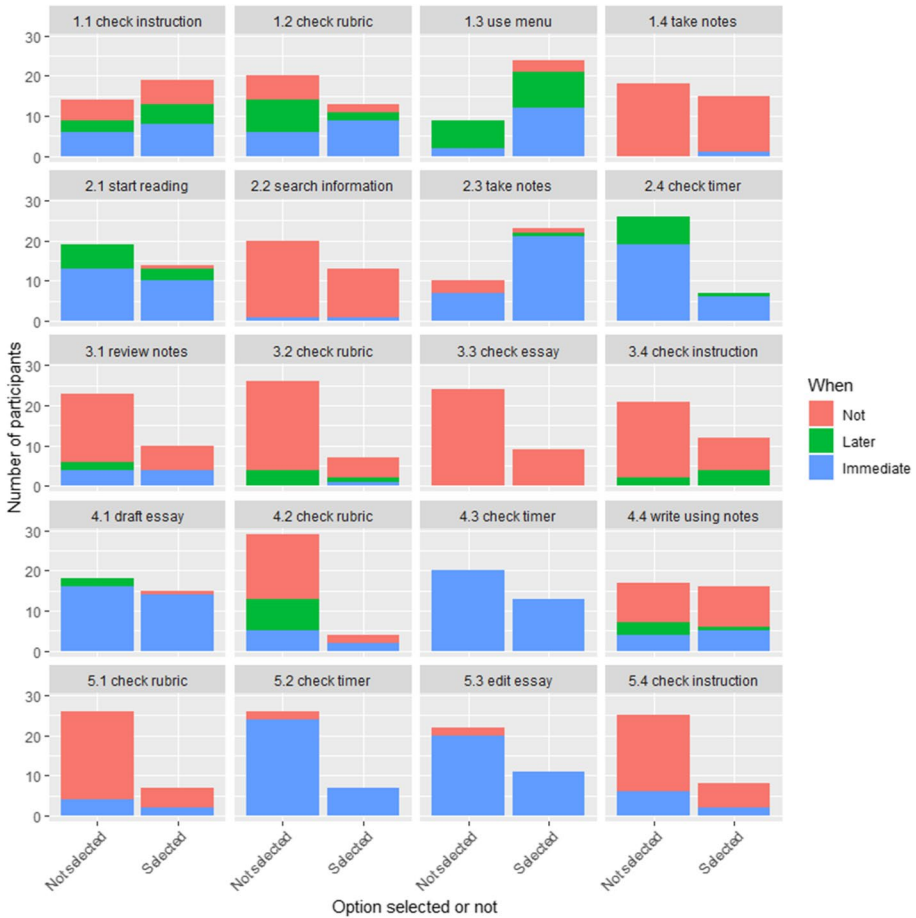

Fig. 6 Compliance in the control, generalized, and personalized condition for each scaffold (row) and option (column), as divided into no compliance (not), compliance after the next scaffold (later), and compliance before the next scaffold (immediate)

scaffold options related to notetaking. Students in the personalized condition appeared to take notes earlier, as indicated by high compliance for scaffold 1 option 4 and lower compliance for scaffold 2 option 3 compared to the control and generalized condition. Students in the generalized condition showed higher compliance to scaffold 3 option 1, and students in the personalized condition showed the highest compliance. These findings mean that students in the personalized condition reviewed their notes. Students in the personalized condition also complied more with using their notes to write the essay (scaffold 4 option 4). Another large difference was scaffold 5 option 2: Check the remaining time. Students in the personalized condition did not comply with this option, while compliance was high in the other conditions. Several small differences were observed. There was maximum compliance in the scaffold condition for scaffold 4 option 3: Checking the remaining time. Compliance was also higher in the scaffold conditions for scaffold 5 option 1: Check the essay rubric.

The third analysis was conducted to investigate the effect of selecting one of the suggested options in the scaffolds, see Fig. 7. We analyzed the generalized condition only due to the personalization in the personalized condition (which caused options not to be displayed, making it hard to compare options and scaffolds) and scaffolds not being present in the control condition. Visual inspection of Fig. 7 indicated that the overall pattern of compliance was the same for selecting or not selecting an option, with four exceptions indicating higher compliance when selecting the option: (1) Scaffold 1 option 2: Check the essay rubric. Students who selected this option executed the action more often than those who did not. (2) Scaffold 2 option 3: Note down important information. Students who selected this option executed the action more often than those who did not. (3) Scaffold 4 option 2: Review the essay rubric. Students who selected this option executed the action directly, while those who did not executed the action later (after the next scaffold). (4) Scaffold 5 option 3: Edit your essay. Students who selected this option executed the action more often than those who did not.

The fourth analysis was conducted on compliance and essay scores, see Table 2. Only the generalized condition was analyzed because their scaffolding was the same, i.e., they received all scaffolds and all options. Essay scores were compared between compliers (i.e., the group that did execute a suggested option of a scaffold) with non-compliers (i.e., the group that did not execute a suggested option of a scaffold). The sizes of these groups differed per comparison based on the compliance and the timing of compliance, see Fig. 7. Therefore, separate t-tests were conducted. Since we had not fully reached a sample size supporting the power needed for the multiple comparisons that follow, we increased the alpha error-level to 0.150 with the purpose of not overlooking compliance effects and exploring our design decisions in more depth. We found six (near) significant effects (see Table 3, 4, and 5 in the Appendix for all tests). (1) Scaffold 3 option 1: Review annotations to check learning so far. Students who executed this action after the next scaffold showed higher essay scores than those who did not, $t(6.98) = 2.82$, $p = .026$, $d = 1.220$. (2) Scaffold 4 option 1: Draft essay by transferring learning to main points. Students who executed this action before the scaffold showed a trend towards higher essay scores than those who did not, $t(29.95) = 1.48$, $p = .148$, $d = 0.519$. (3) Scaffold 4 option 2. Review the essay rubric. Students who executed this action before the scaffold showed a trend towards higher essay scores than those who did not, $t(12.82) = 2.07$, $p = .059$, $d = 0.801$. (4) Scaffold 4 option 4: Write the essay with help from notes. Students who executed this action after the next scaffold showed lower essay scores than those who did not, $t(11.60) = 2.45$, $p = .031$,

Fig. 7 Compliance in the generalized condition for each scaffold (row) and option (column), as divided into no compliance (not), compliance after the next scaffold (later), and compliance before the next scaffold (immediate), taking into account whether an option was selected or not

Table 2 Descriptive statistics for total as a function of condition	Condition	*M*	*SD*
	Control	11.27	3.33
	Generalized	10.67	2.88
	Personalized	11.89	3.63

$d = 1.000$. (5) Scaffold 5 option 1: Review the essay rubric. Students who executed this action before the scaffold showed a trend towards higher essay scores than those who did not, $t(30.08) = 1.79$, $p = .084$, $d = 0.623$. (6) Scaffold 5 option 4: Check the learning goals and instructions. Students who executed this action before the scaffold showed a trend towards higher essay scores than those who did not, $t(30.08) = 1.79$, $p = .084$,

$d = 0.623$. We see beneficial effects of executing the suggested action overall with one exception: notetaking after the last scaffold.

Discussion

The present study aimed to design and evaluate personalized SRL scaffolds. Using a design-based research approach (McKenney & Reeves, 2012), we analyzed the problem in the introduction, drafted a solution, and tested the solution. In this discussion, we first evaluate our design by discussing the effect of design decisions and listing principles for the future design of SRL support (aim 1). Next, we interpret the results in light of the research questions (aim 2).

Design decisions

Decision 1 to use scaffolds seems warranted because an effect on SRL processes was found, especially in the case of personalized scaffolds. More extensive support might be needed to find an effect on learning outcomes. However, this would imply that the system takes over even more of the regulation. We, therefore, recommend finding a balance between the need for scaffolding and students' ability to regulate (Molenaar, 2022). Support should be based on what students need to perform well in a specific context, and this need might differ across students and contexts (Roll & Winne, 2015). We created a baseline in a preceding study, and we recommend future studies to do so as well to identify some of the student and context factors.

Decision 2 to take over monitoring seemed to increase the frequency of monitoring overall (Lim et al., 2023). An increased frequency is associated with higher essay scores (van der Graaf et al., 2022). Therefore, scaffolding monitoring seems appropriate to increase its frequency and foster learning outcomes. Additionally, it is important that monitoring accuracy is high. This is not only the case when a learner is monitoring but also when the learning system is monitoring. For this purpose, we developed an algorithm to capture a learner's SRL (Fan et al., 2022), which was used to develop our scaffolds. Future research on monitoring scaffolds could use a similar detection of a learner's SRL to take over monitoring of one's SRL.

Decision 3 to suggest concrete actions to enact control might not be effective because the scaffold conditions did not show higher essay scores than the control condition. We provided concrete actions, which were presented as options with few words to prevent cognitive overload (Paas et al., 2004). Still, students might have neglected the scaffolds due to limited cognitive capacity, possibly caused by the time pressure (Barrouillet et al., 2007). A recommendation is to make the scaffolds even more directive as intended with Decisions 7 and 8. Our additional analyses showed that for some options, we see beneficial effects of executing them on the essay scores. These beneficial effects imply that the suggestion of concrete actions can be effective in some, but not all cases. To further support this point, we found a detrimental effect of one suggested action when it was executed after the next scaffold. To conclude, the effect of suggesting concrete actions seems to depend on the action, the timing of the scaffold, and, thus, the task itself. It is therefore recommended to identify which actions can be beneficial at what time to inform the design of scaffolds.

Decision 4 to deliver five scaffolds is hard to evaluate. On the one hand, based on the observations and logs during test, we had the impression that a number of students found the scaffolds disturbing, suggesting that five is too much (similar to reports of eight scaffolds in 40 min; Bannert et al., 2015). On the other hand, scaffolds did not affect learning outcomes suggesting that five is too little. The analysis of the personalization of scaffolds showed that no scaffold was completely dropped for all students. This result indicates that the scaffolds display options that students have not yet thought of and/or implemented, suggesting a need for all five scaffolds. The personalization also showed that some students did not need a scaffold to execute relevant SRL processes. Based on these findings and the effect on the SRL processes, also see Decision 5, five scaffolds was assumed to be an adequate number in a 45-minute learning session for most students, but more research is needed. The number of scaffolds can be manipulated to investigate its effect on SRL processes and outcomes, as there can be different explanations for five scaffolds not being sufficient to enhance learning performance.

Decision 5 to have a specific purpose for each scaffold seemed to have affected the frequencies of the intended behaviors, mainly monitoring and high cognition. It might be concluded that the scaffolds worked as intended. We, therefore, recommend having one purpose per scaffold. Furthermore, the scaffold options were used to different extents. These results suggest multiple ways to attain a specific goal (in line with theory: Winne, 1997). These ways, in turn, can be used to personalize the scaffolds.

Decision 6 to have specific timings of the scaffolds was based on a re-analysis of data from a previous study (Van der Graaf et al., 2022) with the same setup, see Section 2: "Designing Scaffolds", and, therefore, seemed adequate. The additional analyses regarding compliance showed that for each scaffold, there is at least one suggested action that was executed by few to no students. This indicates that the suggestion seemed adequately timed because it was in line with students' spontaneous SRL. Nevertheless, we saw individual differences in compliance: some did not execute the action, and others executed it immediately or later. There were also individual differences during learning. Some students in the personalized condition received fewer scaffolds, while others received all options of all scaffolds. These findings suggest making the scaffolds even more personal, but that would require a more sophisticated way to deal with the individual differences in the temporal aspect of SRL.

Decisions 7 and 8 to make the scaffold directive by having four options and a limited amount of text might not have worked as intended, see the discussion of Decision 3. In an intensive task like ours, it might be recommended to be even more directive. This might be done by having more personalized scaffolds or proposing fewer options. It can also be debated whether a scaffold with one specific goal should display multiple options to attain that goal. Dealing with multiple ways to attain a goal is a process that can also be personalized by determining the priority of each option and only displaying the option with the highest priority while also taking into account which options were already executed.

Decision 9 to personalize the scaffolds appeared meaningful because scaffolds were personalized based on the individual student's SRL process. Personalization was also effective in changing learning behavior. Students in the personalized condition generally showed the most monitoring and high cognition, while they received fewer suggestions than the generalized condition. We, therefore, recommend personalizing scaffolds based on learners' SRL processes.

Decisions 10 and 11 to have breakpoints and a notification preceding the scaffold were hard to evaluate because their effect on learning was unclear. The intention was to minimize the potentially disturbing effects of the scaffolds, which seemed to have worked for some students, who liked the scaffolds, but not others, who felt disturbed in their learning. It might be worthwhile to identify how often breakpoints occur. If these occur rarely, then a scaffold disturbs the student. If these occur every five minutes, scaffolds might be triggered at those moments.

Decisions 12 and 13 to have a checklist, which can be revisited, along with the scaffold, might be more appropriate in longer learning sessions. Observations indicated that these functionalities were barely used. Nevertheless, additional analyses would be required to evaluate the effects of the checklist and revisiting on SRL processes and outcomes. We did find an effect of selecting a specific option to be in the checklist. Selecting an option was generally associated with higher compliance. This does not imply causality in that selecting an option increased compliance. We might have captured the intention to comply by measuring the selection.

Research findings

In relation to aim 2, we found that the scaffolds were personalized to different extent, with scaffold 2 and 5 showing most personalization and scaffold 3 the least. Compliance was overall similar across conditions, with the exception of suggested actions related to note-taking, where compliance was greater in the personalized condition. Being able to select a suggested action in the scaffold resulted in higher compliance for four suggested actions (out of 20 in total). This suggests that students could have the intention to comply without revealing it by selecting the option or that their intention to comply was not present at the time of the scaffold, but rather emerged later. Finally, essay scores were generally higher when students complied with actions that suggested to read the instructions and rubric, and to draft the essay. In contrast, note-taking after the last scaffold was associated with poorer essay scores.

The first finding was that all scaffolds were personalized to some extent. This means that our digital scaffolding system was able to use real-time detection of SRL processes to personalize the content of the scaffolds. We used a binary rule for the presence or absence of specific SRL processes to determine whether to display a specific suggestion or not. This is different from the quartile grouping at fixed times (Pardo et al., 2019). Our binary rule seems much more specific, which also means there is a more specific personalization of the SRL support. This specificity seemed to worked well in the current context, because all scaffolds were personalized. As SRL is proposed to be highly contextual, such as the learning task (Ben-Eliyahu & Bernacki, 2015), we have built on previous work in the same task to built our scaffolding system. It can therefore be recommended to first get a thorough understanding of actual SRL in a specific context before developing a support system.

Part of this finding is that scaffold 2 and 5 were personalized most and scaffold 3 least. An explanation could be that scaffold 3 was offered at a stage during which learners tended to focus on cognitive activities, mainly reading. It has been shown that reading generally is followed by reading (Lim et al., 2023), suggesting that reading is a more isolated process compared to the other processes. Therefore, chances are lower to find other processes, including the one that could trigger a rule to personalize a scaffold. In contrast, the timings

of scaffold 2 (near the start) and 5 (near the end) better align with preparatory or reflective processes (Zimmerman, 2000).

The second finding was that compliance was similar across the generalized and personalized condition, with the exception of note-taking. Students' compliance in the personalized condition suggested earlier note-taking than the other conditions. It might be that personalized scaffolds were more directive and easy to process as generally fewer options were provided, which could reduce cognitive load (Clark & Mayer, 2016). Note-taking can be considered as one of the more cognitively loading suggested actions in the scaffolds and therefore, the effect of personalizing the scaffold might be largest for note-taking.

The third finding was that selecting a suggestions to act upon later was associated with higher compliance for four out of 20 scaffold options. Note that this was analyzed in the generalized condition only, because options were not comparable across students in the personalized condition and the control condition did not include scaffolds. Interestingly, these four suggestions included different actions, namely reading the rubric, taking notes, and writing the essay. This could be in line with the notion that every learning action is a result of a decision (Winne, 1996) and that we were able to capture this process. Future studies might take investigate the effect of personalization of the scaffolds on the selection of the options, and subsequently on the compliance with the scaffold. This can be done by manipulating the number of options in the scaffold, which was done in the current study in a flexible manner based on students' previous learning actions. In addition, the specific actions suggested can be manipulated investigating whether the action suggested affects the intention to comply. Finally, previous scaffold interactions can be investigated in relation to the intention to comply and compliance. A larger sample size would be needed to consider subgroups of students that have similar previous scaffold interactions.

The fourth finding was that compliance with a set of suggested actions was related to a learning outcome, namely the essay score. Roughly the same actions were involved as in the previous finding, namely reading the rubric, reviewing (instead of taking) notes, and writing the essay. This means that students who executed these suggested actions generally had higher essay scores. The rubric was used for grading and writing the essay led to actually having an essay, and therefore, from the perspective of this specific task it makes sense to find these associations, as students did what was requested. The rubric can be used for planning and monitoring, which both have been found to be more frequent in more successful students than less successful students (Engelmann & Bannert, 2021). Essay writing conceptualized in high cognitive behavior has previously been found to be positively related to essay scores (van der Graaf et al., 2022). If the content of the notes was in line with the rubric or with what the students intended to write in the essay, then this explanation also holds for reviewing notes. Reading notes, as part of so-called deep processing, has been previously found to be associated with learning outcomes (Deekens et al., 2018).

Limitations and suggestions

There is one major limitation to this study: we could not disentangle the effects of our 13 decisions. To further complicate things, there is a dependency between some of the decisions. It would be interesting to manipulate only one aspect and study its effects. However, such effects might be smaller than a combination of manipulations. Another

limitation is the sample size. Larger sample sizes would be needed to analyze subsets of learners, such as those whose scaffolds were not personalized. It can be expected that students whose scaffolds were personalized show different SRL and a different association of SRL processes with learning outcomes than students whose scaffolds were not personalized because there already was a difference in SRL before the scaffold. Another reason to expect a difference is that due to personalization one or more options were displayed which could affect the perceived usefulness of the remaining options that were displayed. Larger sample sizes can also help to analyze the details of the relationship between compliance and learning outcomes. We expect that more beneficial effects of compliance can be found with larger sample sizes. Our study was also limited because scaffolds were based only on process data. Therefore, it is unsurprising that the results showed scaffold effects on processes but not outcomes or products. Future research could also incorporate learners' products in the scaffolds, which has been done in personalizing feedback (Maier & Klotz, 2022). A final consideration is to incorporate relevant learner characteristics that might interact with task characteristics in affecting SRL (Seufert, 2018), for example working memory or prior experience with digital learning or scaffolds.

Implications

The present study showed how scaffolds for SRL can be designed, tested, and evaluated. We used a design-based approach with additional analyses to help evaluate our decisions. The preparatory studies were especially helpful in the design of our scaffolds. These studies provided relevant data about students' SRL processes, which informed decisions about the number of scaffolds, the timing of scaffolds, and the personalization of scaffolds. Regarding evaluating our scaffolds, it is useful to distinguish between the effects on SRL processes and learning outcomes. It might well be that scaffolds foster SRL without directly fostering learning outcomes (Molenaar et al., 2010, 2011). The compliance analysis also proved fruitful in the evaluation, in line with previous studies (e.g., Bannert & Reimann, 2012).

Conclusion

To conclude, our study ended a cycle of design-based research revealing the effectiveness of our decisions in designing SRL scaffolds. Our findings showed when scaffolds were effective in fostering which aspects of SRL for which learners. These findings provide guidelines for the design of generalized scaffolds and, more innovatively, for the design of personalized scaffolds. Furthermore, we offer an evaluative framework to identify the effects of personalization and scaffolding on SRL and learning outcomes.

Appendix

Table 3 Essay scores per compliance group in the generalized condition for early compliance

Scaffold and option	Non-compliance		Compliance				
	M (SD)	n	M (SD)	n	t	Df	p
1.1	11.00 (2.67)	8	10.56 (2.99)	25	0.39	13.11	0.700
1.2	11.40 (2.88)	10	10.35 (2.89)	23	0.97	17.26	0.348
1.3	10.33 (3.56)	15	10.94 (2.24)	18	0.58	22.70	0.570
1.4	10.48 (2.57)	25	11.25 (3.85)	8	0.53	9.09	0.609
2.1	11.33 (4.55)	6	10.52 (2.47)	27	0.43	5.67	0.686
2.2	10.67 (2.88)	33		0			
2.3	10.78 (2.91)	9	10.63 (2.93)	24	0.13	14.53	0.895
2.4	11.00 (2.52)	17	10.31 (3.26)	16	0.67	28.27	0.506
3.1	10.79 (2.74)	29	9.75 (4.11)	4	0.49	3.38	0.653
3.2	11.00 (2.43)	26	9.43 (4.16)	7	0.96	7.14	0.370
3.3	10.69 (2.92)	32	10.00 (-)	1			
3.4	11.00 (2.43)	26	9.43 (4.16)	7	0.96	7.14	0.370
4.1[††]	10.00 (2.83)	18	11.47 (2.83)	15	1.48	29.95	0.148
4.2[†]	10.23 (2.93)	26	12.29 (2.14)	7	2.07	12.82	0.058
4.3	10.75 (3.10)	4	10.66 (2.91)	29	0.06	3.77	0.957
4.4	10.43 (3.10)	23	11.20 (2.35)	10	0.78	22.54	0.445
5.1[†]	9.82 (2.63)	17	11.56 (2.94)	16	1.79	30.08	0.084
5.2		0	10.67 (2.88)	33			
5.3	8.00 (2.65)	3	10.93 (2.80)	30	1.82	2.47	0.186
5.4[†]	9.82 (2.63)	17	11.56 (2.94)	16	1.79	30.08	0.084

*$p < .05$. †$p < .10$. ††$p < .15$

Table 4 Essay scores per compliance group in the generalized condition for immediate compliance

Scaffold and option	Non-compliance		Compliance				
	M (SD)	n	M (SD)	n	t	Df	p
1.1	10.63 (2.27)	19	10.71 (3.65)	14	0.07	20.27	0.941
1.2	10.44 (2.18)	18	10.93 (3.61)	15	0.46	22.07	0.651
1.3	10.89 (2.90)	19	10.36 (2.92)	14	0.52	28.05	0.605
1.4	10.50 (2.76)	32	16.00 (-)	1			
2.1	10.50 (3.27)	10	10.74 (2.77)	23	0.20	14.87	0.843
2.2	10.74 (2.95)	31	9.50 (0.71)	2	1.70	4.34	0.158
2.3	11.40 (3.85)	5	10.54 (2.74)	28	0.48	4.75	0.652
2.4	10.88 (2.80)	8	10.60 (2.96)	25	0.24	12.43	0.815
3.1	10.48 (3.07)	25	11.25 (2.25)	8	0.77	16.14	0.455
3.2	10.66 (2.92)	32	11.00 (-)	1			
3.3	10.67 (2.88)	33		0			
3.4	10.67 (2.88)	33		0			
4.1	8.00 (2.65)	3	10.93 (2.80)	30	7.82	2.47	0.186
4.2	10.42 (2.98)	26	11.57 (2.44)	7	1.05	11.36	0.315
4.3		0	10.67 (2.88)	33			
4.4	10.50 (3.13)	24	11.11 (2.15)	9	0.64	21.19	0.531
5.1	10.59 (2.66)	27	11.00 (4.00)	6	0.24	6.02	0.820
5.2	9.50 (0.71)	2	10.74 (2.95)	31			
5.3	8.50 (0.71)	2	10.81 (2.91)	31			
5.4	11.04 (2.64)	25	9.50 (3.46)	8	1.15	9.74	0.276

*p < .05. †p < .10. ††p < .15

Table 5 Essay scores per compliance group in the generalized condition for late compliance

Scaffold and option	Non-compliance		Compliance				
	M (SD)	n	M (SD)	n	t	Df	p
1.1	10.70 (2.96)	20	10.62 (2.87)	13	0.08	26.33	0.936
1.2	10.17 (2.67)	17	11.19 (3.08)	16	1.00	29.78	0.324
1.3	9.75 (1.71)	4	10.79 (3.00)	29	1.02	5.99	0.346
1.4	10.66 (2.88)	33		0			
2.1	10.22 (2.33)	9	10.83 (3.09)	24	0.610	19.10	0.549
2.2	10.66 (2.88)	33		0			
2.3	11.33 (3.03)	12	10.29 (2.80)	21	0.98	21.53	0.336
2.4		0	10.66 (2.88)	33			
3.1*	10.21 (2.79)	28	13.20 (2.05)	5	2.82	6.98	0.026
3.2	10.44 (2.56)	27	11.67 (4.18)	6	0.69	5.86	0.518
3.3	10.66 (2.88)	33		0			
3.4	10.44 (2.56)	27	11.67 (4.18)	6	0.69	5.86	0.518
4.1	8.50 (0.71)	2	10.81 (2.91)	31			
4.2	11.30 (2.39)	20	9.69 (3.38)	13	1.49	19.75	0.152
4.3	9.50 (0.71)	2	10.74 (2.95)	31			
4.4*	11.32 (2.66)	25	8.63 (2.72)	8	2.45	11.60	0.031

*p < .05. †p < .10. ††p < .15. There could not be late compliance after scaffold 5, because it was the last scaffold before the task ended

Declarations

Informed consent Informed consent was acquired from the participants, and the study was approved by the Ethics Committee of the Faculty of Social Sciences of the Radboud University.

Conflict of interest We have no known conflict of interest to disclose. This article was (co)authored by guest editors of this special issue. The article was reviewed and revised following the journal's regular review process, without any involvement of the guest editor in the selection of reviewers or the decision making process.

References

Álvarez, R. P., Jivet, I., Pérez-Sanagustín, M., Scheffel, M., & Verbert, K. (2022). Tools designed to support self-regulated learning in Online Learning environments: A systematic review. *IEEE Transactions on Learning Technologies, 15*(4), 508–522. https://doi.org/10.1109/TLT.2022.3193271

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

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with Hypermedia. *Instructional Science, 33*(5), 381–412. https://doi.org/10.1007/s11251-005-1273-8

Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior, 96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – implications for the design of computer-based scaffolds. *Instructional Science, 33*(5/6), 367–379.

Azevedo, R., Moos, D. C., Greene, J. A., Winters, F. I., & Cromley, J. G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development, 56*(1), 45–72. https://doi.org/10.1007/s11423-007-9067-0

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science, 40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4

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

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning Memory and Cognition, 33*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

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

Bowman, N. A., Jang, N., Kivlighan, D. M., Schneider, N., & Ye, X. (2020). The impact of a goal-setting intervention for Engineering students on academic probation. *Research in Higher Education, 61*(1), 142–166. https://doi.org/10.1007/s11162-019-09555-x

Clark, R. C., & Mayer, R. E. (2016). *E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning* (4th ed.).Wiley https://www.wiley.com/en-us/e+Learn ing+and+the+Science+of+Instruction%3A+Proven+Guidelines+for+Consumers+and+Designers+ of+Multimedia+Learning%2C+4th+Edition-p-9781119158660. Accessed 19 Feb 2023.

Crompton, H., Bernacki, M., & Greene, J. A. (2020). Psychological foundations of emerging technologies for teaching and learning in higher education. *Current Opinion in Psychology, 36*, 101–105. https:// doi.org/10.1016/j.copsyc.2020.04.011

Deekens, V. M., Greene, J. A., & Lobczowski, N. G. (2018). Monitoring and depth of strategy use in computer-based learning environments for science and history. *British Journal of Educational Psychology, 88*(1), 63–79. https://doi.org/10.1111/bjep.12174

Du, J., & Hew, K. F. T. (2022). Using recommender systems to promote self-regulated learning in online education settings: Current knowledge gaps and suggestions for future research. *Journal of Research on Technology in Education, 54*(4), 557–580. https://doi.org/10.1080/15391523.2021.1897905

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Learning and Instruction, 72*, 101205. https://doi.org/10.1016/j. learninstruc.2019.05.002

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*. https://doi.org/10.1007/ s11409-022-09291-1

Fleiss, J.L., Levin, B., Paik, M.C. (2003). Statistical methods for rates and proportions (3rd ed.). John Wiley & Sons.

Hmelo-Silver, C. E., & Azevedo, R. (2006). Understanding complex systems: Some core challenges. *The Journal of the Learning Sciences, 15*(1), 53–61.

Jaeger, A. J., & Wiley, J. (2014). Do illustrations help or harm metacomprehension accuracy? *Learning and Instruction, 34*, 58–73. https://doi.org/10.1016/j.learninstruc.2014.08.002

Lim, L., Bannert, M., van der Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal Assessment of Self-regulated learning by Mining Students' think-Aloud protocols. *Frontiers in Psychology, 12*, 749749. https://doi.org/10.3389/fpsyg.2021.749749

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10. 1016/j.chb.2022.107547

Maier, U., & Klotz, C. (2022). Personalized feedback in digital learning environments: Classification framework and literature review. *Computers and Education: Artificial Intelligence, 3*, 100080. https://doi.org/10.1016/j.caeai.2022.100080

Matcha, W., Uzir, N. A., Gašević, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies, 13*(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

McKenney, S. E., & Reeves, T. C. (2012). *Conducting educational design research*. Routledge.

Miller, C. J., & Bernacki, M. L. (2019). Training preparatory mathematics students to be high ability self-regulators: Comparative and case-study analyses of impact on learning behavior and achievement. *High Ability Studies, 30*(1–2), 167–197. https://doi.org/10.1080/13598139.2019.1568829

Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence, 3*, 100070. https://doi.org/10.1016/j.caeai.2022.100070

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

Molenaar, I., Roda, C., van Boxtel, C., & Sleegers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning environment. *Computers & Education, 59*(2), 515–523. https://doi.org/10.1016/j.compedu.2011.12.006

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2010). The effects of scaffolding metacognitive activities in small groups. *Computers in Human Behavior, 26*(6), 1727–1738. https://doi.org/10. 1016/j.chb.2010.06.022

Molenaar, I., van Boxtel, C. A. M., & Sleegers, P. J. C. (2011). Metacognitive scaffolding in an innovative learning arrangement. *Instructional Science, 39*(6), 785–803. https://doi.org/10.1007/ s11251-010-9154-1

Munshi, A., & Biswas, G. (2019). Personalization in OELEs: Developing a data-driven framework to model and scaffold SRL Processes. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, &

R. Luckin (Eds.), *Artificial Intelligence in Education* (pp. 354–358). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_65

Munshi, A., Biswas, G., Baker, R., Ocumpaugh, J., Hutt, S., & Paquette, L. (2022). Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. *Journal of Computer Assisted Learning*, *n/a*(n/a), https://doi.org/10.1111/jcal.12761

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the Interaction between Information structures and Cognitive Architecture. *Instructional Science, 32*(1/2), 1–8.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for Research. *Frontiers in Psychology, 8*, 422. https://doi.org/10.3389/fpsyg.2017.00422

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128–138. https://doi.org/10.1111/bjet.12592

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138*(2), 353–387. https://doi.org/10.1037/a0026838

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

Schippers, M. C., Scheepers, A. W. A., & Peterson, J. B. (2015). A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap. *Palgrave Communications, 1*(1), 1. https://doi.org/10.1057/palcomms.2015.14

Schunk, D. H., & Greene, J. A. (2018). Historical, contemporary, and future perspectives on self-regulated learning and performance. *Handbook of self-regulation of learning and performance* (pp. 1–15). Routledge.

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly, 19*(2), 159–172. https://doi.org/10.1080/10573560308219

Schunk, D. H., & Rice, J. M. (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Reading Behavior, 23*(3), 351–364. https://doi.org/10.1080/10862969109547746

Seufert, T. (2018). The interplay between self-regulation in learning and cognitive load. *Educational Research Review, 24*, 116–129. https://doi.org/10.1016/j.edurev.2018.03.004

Sharma, P., & Hannafin, M. J. (2007). Scaffolding in technology-enhanced learning environments. *Interactive Learning Environments, 15*(1), 27–46. https://doi.org/10.1080/10494820600996972

Shih, K. P., Chen, H. C., Chang, C. Y., & Kao, T. C. (2010). The development and implementation of scaffolding-based self-regulated Learning System for e/m-Learning. *Journal of Educational Technology & Society, 13*(1), 80–93.

Siadaty, M., Gasevic, D., & Hatala, M. (2016). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, *3*(1), https://doi.org/10.18608/jla.2016.31.11

Tekin, E. (2022). Can learners allocate their study time effectively? It is complicated. *Educational Psychology Review, 34*(2), 717–748. https://doi.org/10.1007/s10648-021-09645-2

van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Gašević, D., Bannert, M., & Molenaar, I. (2022). The Dynamics Between Self-Regulated Learning and Learning Outcomes: An Exploratory Approach and Implications. Metacognition and Learning. https://doi.org/10.1007/s11409-022-09308-9

van Merriënboer, J. J. G., & de Bruin, A. B. H. (2019). Cue-based facilitation of self-regulated learning: A discussion of multidisciplinary innovations and technologies. *Computers in Human Behavior, 100*, 384–391. https://doi.org/10.1016/j.chb.2019.07.021

Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. *Handbook of self-regulation of learning and performance* (pp. 36–48). Routledge.

Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences, 8*(4), 327–353. https://doi.org/10.1016/S1041-6080(96)90022-9

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology, 89*(3), 397–410.

Winne, P. H. (2019). Paradigmatic dimensions of instrumentation and analytic methods in research on self-regulated learning. *Computers in Human Behavior, 96*, 285–289. https://doi.org/10.1016/j.chb.2019.03.026

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting self-regulated learning in Online Learning environments and MOOCs: A systematic review. *International Journal of Human–Computer Interaction, 35*(4–5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in Problem Solving*. *Journal of Child Psychology and Psychiatry, 17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Zheng, L. (2016). The effectiveness of self-regulated learning scaffolds on academic performance in computer-based learning environments: A meta-analysis. *Asia Pacific Education Review, 17*(2), 187–202. https://doi.org/10.1007/s12564-016-9426-9

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20*(1), 99–149. https://doi.org/10.1006/drev.1999.0497

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice, 41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102_2