Klar *Smart Learning Environments* (2025) 12:32
https://doi.org/10.1186/s40561-025-00385-2

Smart Learning Environments

RESEARCH

Open Access

Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots

Maria Klar[1*]

*Correspondence:
maria.klar@uni-due.de

[1] University of Duisburg-Essen, Universitätsstr. 2, 45141 Essen, Germany

Abstract

Generative AI (genAI) chatbots can be a powerful tool for learning, especially self-regulated learning, but they can also add more complexity to the learning process because of their numerous capabilities. Learners might not make use of these capabilities and fail to profit from them. One core genAI affordance is to adapt any learning material to the learner's individual needs. To help learners make use of these adaptation capabilities, instructional prompts can help, but they can also have negative effects on cognitive load. This study investigated K-12 learners' perceptions and interaction patterns with a chatbot, focusing on their use of content adaptation capabilities. In this experimental study with 106 secondary school students, a group receiving adaptation guidance (including brief instruction, prompt suggestions, and adjustable chatbot response length and language level) was compared to a control group without adaptation guidance. Results show that learners perceived chatbots as easy to use despite limited prior experience. Without guidance, they underutilized the chatbot's adaptation capabilities. The experimental group used twice as many adaptation prompts ($M = 6.0$) compared to the control group ($M = 2.5$), without experiencing increased cognitive load. Interestingly, both groups showed similar knowledge gains and reported high satisfaction levels. These findings suggest that easily implementable interventions can enhance students' use of genAI chatbot capabilities, potentially improving their self-regulated learning experiences. Future research should explore the long-term effects of adaptation guidance on learning outcomes and self-regulated learning skills.

Keywords: Generative AI chatbot, K-12, Perception, Interaction pattern, Adaptation, Support, Cognitive load

Introduction

Generative AI (genAI) chatbots are gaining prominence in educational settings, introducing new potentials but adding complexity to self-regulated learning. They offer numerous capabilities for various stages of the learning process, from initial research to content adaptations and self-assessment (Chiu, 2024). However, the very versatility that makes genAI chatbots useful learning aids could also present challenges for learners in terms of effective utilization and cognitive load management. There is a gap in understanding how K-12 students interact with tools, particularly in adapting chatbot responses to individual needs. This gap is significant, as the effective integration of genAI chatbots into K-12 education requires a nuanced understanding of students' usage patterns, perceptions, and the impact of supportive interventions.

This study aims to address this knowledge gap by investigating K-12 learners' perceptions and interaction patterns with genAI chatbots—hereinafter referred to as "chatbots", for simplicity—as well as the effects of adaptation guidance on prompt use and learning outcomes, and potential moderating factors such as prior knowledge and self-regulated learning skills. To explore these questions, an experimental study was conducted with 106 secondary school students (grades 9–10) in Germany. Both the control and the experimental group used a chatbot for research. The experimental group received three types of adaptation guidance: brief instruction on adaptation prompts (e.g., "Make it easier", "Give a summary"), follow-up prompt suggestions, and options to modify response length and language level.

By examining how K-12 students engage with chatbots and the impact of targeted support, this study contributes to an understanding of effective integration into secondary education. The findings have implications for educators and chatbot designers in addressing challenges related to self-regulated learning and cognitive load.

Theoretical background

Current educational settings often require significant self-regulation during learning (de Bruin et al., 2020), as students face complex problems that require choosing between numerous next steps, demanding metacognitive strategies (Azevedo et al., 2012; Winne, 2013). Acquiring self-regulated learning strategies can be a challenge and even when acquired, it can impose a higher cognitive load (de Bruin et al., 2020). Cognitive load increases with the number of mental processes performed simultaneously (Sweller, 2010). Tasks requiring the learner to search among multiple solutions induce *extraneous load*, i.e., load induced by processing information that is irrelevant to the task (Sweller, 2011). *Intrinsic cognitive load* stems from information processing demanded by the task and is a function of task difficulty and prior knowledge. *Germane load* refers to the resources attributed to deal with intrinsic information processing (Sweller, 2010). From this perspective, self-regulated learning can cause cognitive load due to constant decision-making. Conversely, cognitive load can also cause self-regulation, such as asking for help or breaking down tasks when they are perceived as difficult (Wang & Lajoie, 2023).

According to Seufert (2020), self-regulation can be conceptualized as a U-shaped function of task difficulty on the one hand and learners' resources and imposed cognitive load on the other: When tasks are easy, resources are high, and no regulation is needed.

When tasks are difficult, and resources are low, regulation is impossible. With AI chatbots entering educational settings, the question is how they influence the learners' resources and cognitive load. Attention has been given to shortcuts taken by students via outsourcing tasks to generative AI (e.g., Playfoot et al., 2024). This (vastly) reduces the task difficulty and the need for self-regulation. However, using generative AI as a tool for complex tasks without shortcuts might add a further layer of complexity and thereby cognitive load.

Chatbots as "easy to use"

GenAI chatbots have an intuitive interface that is similar to a search bar or a chat window in a messenger app. Therefore, it is no surprise that students rate chatbots such as ChatGPT as positive (Stöhr et al., 2024) and as easy to use (Ngo, 2023; Shoufan, 2023). However, chatbots come with complexities beneath the surface. First, there are hallucinations (Tlili et al., 2023; Zhang et al., 2023) and risks of biased outputs (Kotek et al., 2023). Second, the interface implies affordances like back-and-forth conversations. However, these chatbots provide more affordances than that. Antonenko et al. (2017) suggest distinguishing *direct* and *indirect affordances*: direct affordances are those affordances that were intended by the design of the object, while indirect affordances are those abilities of an object perceived by the user that were not intended by design. With genAI chatbots, there is the (possibly rare) situation that products are launched with a wide range of indirect affordances (Pozdniakov et al., 2024). For example, in educational settings, chatbots can generate feedback, adapt learning material, role-play, and generate self-tests (Kasneci et al., 2023), adding a wide range of options.

The fact that learners perceive chatbots as easy to use is an indication that they underutilize these beneficial capabilities and possibly also invest too little mental effort. In a seminal paper aptly titled "Television is 'easy' and print is 'tough': The differential investment of mental effort in learning as a function of perceptions and attributions", Salomon (1984) made the case that learners learn less from the same narrative material when it is presented as a silent TV film than when it is presented as text because they invest less mental effort. The TV is "familiar, overlearned and lifelike" (p. 650), so the children participating in the study perceive it as easy compared to text and thus only engage in shallow processing. As chatting with a chatbot can similarly be perceived as familiar, overlearned, and lifelike, a similar phenomenon might occur: Learners might not perceive the manifold indirect affordances of the chatbot, underutilize them, and invest too little effort, as suggested in a recent study by Stadler et al. (2024).

To investigate this, there is more research needed on how learners interact with chatbots. General research shows university students use AI for tasks like research, translations, problem-solving, and text processing (von Garrel & Mayer, 2023). However, little research exists on specific interaction patterns and their interactions with learner characteristics. Abdelhalim (2024) found that students with high metacognitive awareness used ChatGPT more skillfully for research than those with lower awareness, suggesting students need tailored support to effectively use chatbots for self-regulated learning.

Supporting learning with chatbots

As there are numerous ways to utilize chatbots in self-regulated learning, there are equally numerous ways of support for learners. Frequent use cases of chatbots among students are research and literature studies (von Garrel & Mayer, 2023). In this context, one core capability is adapting content to the learner's needs (Hadi Mogavi et al., 2024). However, chatbots are not tutoring systems and lack a learner model, requiring students to manage their learning through metacognitive strategies. For example, if they perceive a task as too difficult, they could ask the chatbot for step-by-step explanations (a worked example, as suggested by Cognitive Load Theory), simpler language, scaffolding, or alternate formats, like bullet points, flash card questions, summaries, tables, and more. Deciding among these adaptation options can induce cognitive load. Therefore, there are aspects of chatbots that can increase cognitive load, such as the numerous indirect affordances and lack of SRL support, while others, like the intuitive interface, may reduce it.

There are two main ways to support the learners in making effective use of the affordance of adaptation: instruction and design changes to the chatbots themselves. Instruction could focus on prompting tips alone or integrate them with cognitive and metacognitive strategies related to the task at hand (Klar, 2024). Domain-specific training offers better transfer (Schuster et al., 2020), but domain-general training has also been shown to be effective (Klaykaew & Prasittichok, 2024; Stebner et al., 2019).

There are several approaches to support the more effective use of chatbots by changing their design (Molenaar, 2022; Pozdniakov et al., 2024). One approach is to make the chatbot's capabilities clearer. Many providers of chatbots now present suggestions for possible prompts to make their capabilities more transparent. Another option is to make these capabilities more easily accessible. For example, students can find the chatbot answers too long and formal (Theophilou et al., 2023) and it would be tiresome to repeatedly ask the chatbot to generate shorter answers. To alleviate this, there could be general adjustment options. Furthermore, instructional prompts could be included in the chatbot interface, reminding the learners to engage in certain self-regulation or metacognitive activities (Azevedo et al., 2022). These changes should be evaluated for their impact on cognitive load, avoiding overload (Gentner & Seufert, 2020).

In summary, genAI chatbots in education present both opportunities and challenges for self-regulated learning and cognitive load management. While their intuitive interface may reduce cognitive demands, the numerous indirect affordances of these chatbots can increase cognitive load by introducing additional complexities and decision-making processes for learners. Students' perception of chatbots as easy may lead to underutilization of useful features. To address these issues, two main approaches are suggested: providing instruction that integrates prompting tips with cognitive and metacognitive strategies and implementing design changes to make chatbot capabilities more transparent and accessible. These interventions aim to support learners in effectively leveraging chatbots for self-regulated learning while managing cognitive load.

Research gap and questions

Current research on students' adoption of genAI chatbots often focuses on what students use chatbots for and how often (e.g., Stöhr et al., 2024; von Garrel & Mayer, 2023). While these studies provide valuable insights, in a next step the question is how exactly students interact with these chatbots within individual chat sessions (e.g., Abdelhalim, 2024). Insights into these interaction patterns help understand whether students underuse affordances like adaptability. There is also emerging research on how to design chatbots in such a way that they support learning (e.g., Wiboolyasarin et al., 2024) rather than merely implementing chatbots off-the-shelf. For example, Abdelhalim (2024) found interface features like prompt suggestions can be of benefit. Therefore, chatbot designs to facilitate learning merit further investigation.

In view of these research gaps, to investigate perceptions and interaction patterns of chatbots in K-12 students and to evaluate whether instruction and chatbot interface features can help students adapt chatbot responses to their needs, this study addresses the following research questions. Where applicable, the assumed hypotheses are stated.

- RQ1: How do learners perceive the generative AI chatbot with and without adaptation guidance?
- RQ2: How do learners use the chatbot during research?
- RQ3: Does adaptation guidance affect the number of adaptation prompts used, extraneous cognitive load, and knowledge increase?

  - H3.1: Adaptation guidance leads to more adaptation prompts.
  - H3.2: Adaptation guidance leads to reduced extraneous load.
  - H3.3: Adaptation guidance leads to a larger knowledge increase.

- RQ4: Do prior knowledge, chatbot experience, or self-regulated learning skills moderate the effect of adaptation guidance on the number of adaptation prompts, extraneous load, or knowledge increase?

  - H4.1: The effects assumed in H3.1 to 3.3 are stronger for students with lower prior knowledge than students with higher prior knowledge.
  - H4.2: The effects assumed in H3.1 to 3.3 are stronger for students with lower chatbot experience than students with higher chatbot experience.
  - H4.3: The effects assumed in H3.1 to 3.3 are stronger for students with lower SRL skills than students with higher SRL skills.

The way in which adaptation guidance was provided will be explained below.

Method and instruments

This study follows an experimental, value-added research design. It combines a quantitative analysis of variables on perception, satisfaction, cognitive load, and knowledge gain with a qualitative analysis of interaction patterns. Participants were recruited from five secondary schools. 9th and 10th graders were invited to participate in the study that took place after class within the respective school. The students were informed about the study and provided parental consent. Up to seven students could take part in one session lasting 50 min. A compensation of 10€ (about $11) was given.

Research setup

Each group of up to seven participants was randomly assigned to one of the two conditions. On average, there were four students per group, so while randomization did not take place at the individual level but at the level of a small group, this study can be considered a true experimental study in contrast to quasi-experimental studies where whole classes are randomized. At the start, the participants were informed about the procedure. The students then replied to the first part of the survey and received the instructions for the research phase. As part of the adaptation guidance, the experimental group additionally received a 2.5-min introduction on how to use adaptation prompts such as "Make it easier," "Give me a summary of our chat so far," or "What aspects have we not covered yet?". The introduction included the metacognitive pointer that such adaptations help receiving answers that are neither too challenging nor too easy. In both conditions, the students then performed exploratory research on the topic of "cognitive biases" for 20 min. They were allowed to use the chatbot and take notes. The experimental group worked with a chatbot with supportive features, as described below. After the research phase, students in both conditions worked through the second part of the survey.

Both chatbots were based on ChatGPT3.5 with a generic system prompt like "You explain concepts to students." As part of the adaptation guidance, in the experimental group, the chatbot included several features that aimed to support the students in their chatbot interaction (Wiboolyasarin et al., 2024).

As shown in Fig. 1, the chatbot for the experimental group included prompt suggestions (1), and a feature to adapt the output length and language level (2). The control group did not have these features.

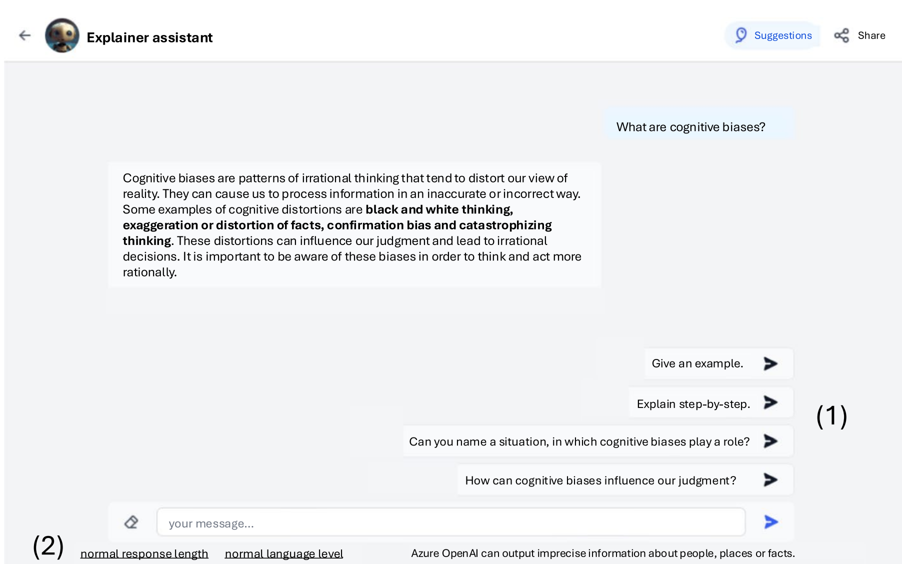

Fig. 1 Chatbot interface for the experimental group. Screenshot taken by the author. Translations were added

In summary, both groups conducted a 20-min research on the topic “cognitive biases”, using only a chatbot and taking notes if they wished to do so. The adaptation guidance provided to the experimental group consisted of three aspects: 1) a 2.5-min instruction on adaptation prompts like “Make it easier,” or “Give me an outline,” 2) suggestions on what to prompt next, including two adaptation prompts and two follow-up questions, 3) the options to change the length and language level of all subsequent chatbot responses.

Instruments

In the pre-survey, students answered questions on prior chatbot experience and self-efficacy, self-regulated learning skills, prior knowledge, interest in using a chatbot, and interest in the topic of “cognitive biases”. Chatbot experience was assessed with three items on the frequency of chatbot use at home and school, the scale adapted from Thompson (2013), with a Cronbach’s alpha of 0.68. SRL skills were assessed via the subscale “Processing Information” from the LASSI-HS (Weinstein et al., 1988). In contrast to other studies (e.g., Magno, 2011), in this study, the subscale reached only a Cronbach’s alpha of 0.60.

The post-survey covered questions on cognitive load, knowledge, interest in the chatbot and the topic, learner satisfaction and perceived usefulness. Cognitive load was assessed using the 8-item instrument by Klepsch et al. (2017), which covers all three aspects of cognitive load. Knowledge was assessed through open-ended pre- and post-tests on cognitive biases, scored for breadth, depth, and precision by two raters. As a control variable, the participant’s interest in the topic and in using a chatbot was assessed via single items (Laine et al., 2020). Furthermore, seven items were included to measure affective and evaluative aspects of the participants’ experience, including adaptations from Lim et al. (2022). Students in the experimental group also rated the helpfulness of prompt suggestions and customization options.

Analysis and data preprocessing

106 of the 108 datasets were complete and included in the sample. In order to analyze the chatbot interactions, the following main codes were established deductively: *adaptation prompts*, *questions*, and *conversation with the bot* as well as the codes for changing the language level and length of the bot’s answers. Then, further subcodes were coded inductively. This way, a coding scheme was developed in the first round. The second rater coded all chats with this coding scheme. In cases of a discrepancy greater than 1, both raters discussed the case and reached an agreement. The intercoder agreement was then calculated as the weighted kappa for adaptation prompts ($\kappa = 0.94$) and questions ($\kappa = 0.91$).

The responses for the pre- and post-test were rated according to a rating scheme, then ratings with a difference greater than 1 were discussed between the two raters so that interrater agreement reached $\kappa = 0.70$ for breadth, $\kappa = 0.73$ for depth and $\kappa = 0.79$ for precision for the pre-test, and $\kappa = 0.88$ for breadth, $\kappa = 0.83$ for depth and $\kappa = 0.74$ for precision for the post-test.

Results

Sample

The participants attended comprehensive schools and gymnasiums in Germany. Of the 106 participants, 54 stated that they were female, 48 were male, 1 was genderqueer and 3 left this question unanswered. On average, they were 14.7 years old (SD = 0.81). When asked what language or languages they speak at native level, 46 listed one language, 36 named two, and 24 chose more than two. Five of the 106 participants did not list German as one of the languages they speak at native level. The chatbot protocols were investigated in terms of potential language barriers and none were found, so these cases were not excluded from the sample. Three cases were in the experimental group and two cases in the control group, so they were not distributed unevenly.

Descriptive statistics

The descriptive statistics for the moderator and control variables are listed in Table 1. There were no significant differences between the two groups on these variables.

RQ1: learner satisfaction and perceived usefulness

To address research question 1—*How do learners perceive the generative AI chatbot with and without adaptation guidance?*—the descriptive results for the items on learner satisfaction (both groups) and perceived usefulness of the support features (only experimental group) are examined.

All following items were assessed on a 7-point Likert scale. The mean values do not differ significantly between the two groups for the items on learner satisfaction. Learners do not perceive the chatbot as providing answers more adapted to their needs in the group with adaptation guidance than in the group without. Both groups show a very high overall satisfaction (control group (CG): M = 6.17; experimental group (EG): M = 6.25) and ease of use (CG: M = 6.47; EG: M = 6.53). In both groups, students felt they always knew what to ask next (CG: M = 4.4; EG: M = 4.57) and that they always received the response they needed (CG: M = 5.16; EG: M = 5.17). They moderately agreed that the

Table 1 Descriptive statistics between the conditions for moderator and control variables

Variable	Scale	Control (n = 53)		Experimental (n = 53)	
		M	SD	M	SD
Moderator variables					
SRL	1–5	3.51	0.49	3.52	0.57
Prior chatbot experience	1 (never) – 4 (daily)	1.65	0.69	1.74	0.63
Pre-test score	0–5	0.27	0.55	0.04	0.17
Control variables					
Intrinsic load	1–7	2.75	1.34	3.18	1.51
Germane load	1–7	4.9	1.32	5.21	1.07
Interest chatbot pre	1–7	4.7	1.27	4.46	1.46
Interest topic pre	1–7	5.24	1.21	5.21	1.46
Interest chatbot post	1–7	5.15	1.23	5.15	1.49
Interest topic post	1–7	5.72	1.03	5.92	1.28

chatbot responses contained too many technical terms (CG: M = 3.08; EG: M = 3.3) and were too long (CG: M = 3.53; EG: M = 3.6), but did not perceive them as too short (CG: M = 2.17; EG: M = 2.06). The experimental group rated the supportive features favorably with an average agreement of 5.40.

Overall, with and without adaptation guidance learners perceived the chatbot favorably in both groups. They attest a high usability and express high overall satisfaction. Learners in the experimental group rated the supportive features as useful. In the two groups, students equally felt that they always knew what to ask next and that the chatbot responses suited their informational needs.

RQ2: Analyses of the chatbot interactions

To answer research question 2 – *How do learners use the chatbot during research?* – patterns in the participants' chatbot interactions will be investigated. To give an overview

Table 2 Frequency of subcodes across all chat protocols (N = 106)

Code/Subcode	Instances both groups n = 106	EG n = 53	CG n = 53
Question written by the learner	492	248	244
• Question	377	176	201
• Follow-up question	115	72	43
Question (suggested by chatbot)	79	79	NA
Adaptation Prompts written by the learner	385	255	130
• Adapting the form	328	213	115
o Easy or easier	96	67	29
o Give an example	55	26	29
o Short or shorter	49	29	20
o More about this; more details	30	21	9
o Create an outline	10	10	0
o Make a table	7	7	0
• Adapting the content	44	33	11
o Everything important; most important facts; basics	17	14	3
o What else can I ask; what aspects are missing	18	17	1
o Everything you know; as much information as possible	5	1	4
Adaptation prompts suggested by the chatbot	57	57	NA
• Give an example	27	27	NA
• Say this in a simpler way	10	10	NA
• Explain this step-by-step	10	10	NA
• Illustrate this	10	10	NA
Conversation with the bot	182	45	137
• Off-topic	104	18	86
• Questions directed at the bot (e.g., "Who are you?""What is your WhatsApp?")	38	8	30
• Thank you	13	9	4
• Ok	12	2	10
• Praise	5	4	1
Changing the length or language level for the whole chat	35	35	NA
• Short answers	11	11	NA
• Simple language	14	14	NA

of the most common types of interactions in the 106 chat protocols, in Table 2, all subcodes with 5 or more instances are shown. For each subcode, the frequencies are given for both groups as well as for the control group and the experimental group.

There are a few notable differences in frequency between the group with adaptation guidance and the control group. Interestingly, the EG posed fewer self-written questions but more follow-up questions (M = 1.4) than the CG (M = 0.8). The students in the EG could also make use of the suggested questions and they did use them. Taking all forms of questions together, the EG posed significantly more questions (p < .001; M = 7.8) than the control (M = 4.8). Concerning the form of questions, 14 of the 377 questions (3.7%) were written like a search query, e.g., just "cognitive biases" and not in the form of a question like "What are cognitive biases?".

The students in the experimental group also used almost twice as many adaptation prompts, not including those adaptation prompts that were suggested by the chatbot in the EG. Looking at the individual types of adaptation prompts, they used almost all of them more often than the control group – except for "Give an example"–and they used a larger variety. One third of the participants in the control group did not use any adaptation prompt, not even "Give an example".

There is more off-topic conversation in the control group than in the group with adaptation guidance, like "Which consumer brand is the best?" or "How do I get rich fast?" as well as insults or meaningless sequences of characters. These absolute numbers, however, are spread across a minority in each group: 15% (8) of the learners in the experimental group and 23% (12) in the control group engaged in off-topic chat at least once.

Only the experimental group had access to the buttons that would change the length or level of language for the subsequent chatbot responses. These options were used only by a minority: 27% (17) of the participants in the EG used one of the options at least once. This partially contrasts with the fact that asking for an easier or shorter response was the most used adaptation prompt in both groups, apart from asking for examples.

In summary, while the participants interacted with the chatbot in a variety of different ways, some patterns emerge: Almost all of the inputs are formulated in a complete sentence; only 3.7% of the questions were asked in the form of a keyword. In the control group, a third of the learners did not use any adaptation prompt, while with adaptation guidance, all learners used at least one adaptation prompt. The adaptation prompts used most often are equivalents of "Make it easier", "Give an example", and "Make it shorter". However, the option to make all chatbot responses easier or shorter was not used very often in the experimental group. Students in the experimental group posed more questions, especially more follow-up questions. Students in the control group, on the other hand, engaged in more off-topic conversation.

RQ3 and RQ4: hypothesis tests

For research questions 3 and 4, hypotheses were stated. Research question 3 asks whether there are differences between the groups (with and without adaptation guidance) concerning the difference in the number of adaptation prompts, extraneous cognitive load, and knowledge gain. It is hypothesized that adaptation guidance leads to a higher number of adaptation prompts (H3.1), reduced extraneous load (H3.2),

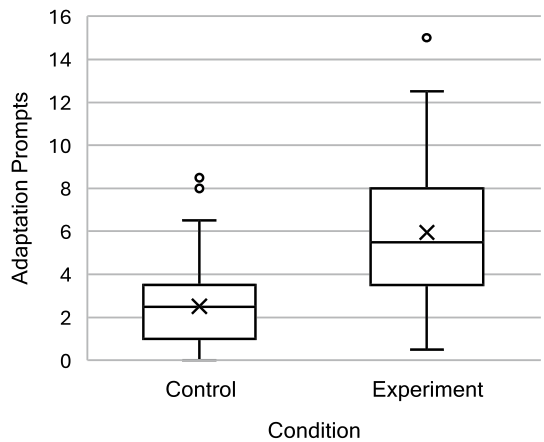

Fig. 2 Boxplot of the number of adaptations per case

Table 3 Study 1 main effects descriptive statistics and post-hoc ANOVA results

Variable	Scale	CG (n = 53)		EG (n = 53)		F(1, 104)	*p*	partial η^2
		M	SD	M	SD			
Adaptation prompts		2.51	2.1	6.0	3.32	41,085	< .001	0.283
Extraneous load	1–7	2.77	1.04	3.0	1.32	0.886	0.349	0.008
Knowledge gains	-5 – 5	2.52	1.1	2.43	0.83	0.202	0.654	0.002

and a larger difference between the pre- and post-test on knowledge, i.e., a larger knowledge gain (H3.3).

In order to test the group differences for three dependent variables and one independent variable with two values, a MANOVA was calculated. The one-factorial MANOVA showed a statistically significant difference between the two groups on the combined dependent variables, $F(3, 102) = 13.978$, $p < .001$, partial $\eta^2 = 0.291$, Wilk's $\Lambda = 0.709$.

Post-hoc univariate ANOVAs were conducted for every dependent variable. Results show a statistically significant difference between the groups with and without adaptation guidance only for the number of adaptive prompts, $F(1, 104) = 41.085$, $p < .001$, partial $\eta^2 = 0.283$, and not for extraneous load, $F(1, 104) = 0.886$, $p = .349$, partial $\eta^2 = 0.008$, or knowledge gain, $F(1, 104) = 0.202$, $p = .654$, partial $\eta^2 = 0.002$.

Therefore, hypothesis 3.1, stating that adaptation guidance leads to a higher number of adaptive prompts is supported by the data. The $\eta^2 = 0.283$ points to a large effect (Cohen, 1988). The average number of adaptation prompts per session is 6 in the experimental group (written by learner and suggested by chatbot combined) and 2.51 in the control group. Figure 2 shows the distribution for both groups.

Hypothesis 3.2, assuming a reduced extraneous load, is not supported. Learners report similar levels of extraneous load in the experimental group (M = 3.0) and in the control group (M = 2.77). Hypothesis 3.3, assuming a larger knowledge gain, is also not supported. The mean difference between pre- and post-test scores is similar between the experimental group (M = 2.43) and the control group (M = 2.52). An overview of the statistics of the hypothesized main effects is given in Table 3.

Research question 4 asks whether different kinds of prior knowledge moderate the effect of adaptation guidance on the number of adaptation prompts, extraneous load, or knowledge gains.

A moderation analysis was performed using the PROCESS macro by Hayes (2017). Conditions for this method were not met for the dependent variable adaptation prompts. For extraneous load and knowledge gain moderation analyses using this approach were conducted for the three hypothesized moderators: prior chatbot experience, self-regulation, and prior knowledge, measured via the pre-test. There were no significant differences between the groups in terms of the moderator variables. For none of these hypothesized moderations a significant interaction was found. Therefore, hypotheses 4.1 through 4.3 are not supported by the data for extraneous load and knowledge gain and could not be tested for adaptation prompts.

Discussion

This study was guided by four research questions:

(1) How do learners perceive the generative AI chatbot with and without adaptation guidance?
(2) How do learners use the chatbot during exploratory search?
(3) Does adaptation guidance affect the number of adaptation prompts, extraneous cognitive load, and knowledge increase?
(4) Do prior knowledge, chatbot experience, or self-regulated learning skills moderate the effect of adaptation guidance on the number of adaptation prompts, extraneous load, or knowledge increase?

To address these questions, an experiment was set up in which 106 secondary school students in 9th and 10th grade who conducted research for twenty minutes on the topic of "cognitive biases" with a chatbot. Half of these students were randomly assigned to the experimental group and received three types of adaptation guidance: 1) an instruction of about three minutes on tips for adaptation prompts, 2) four suggestions after each chatbot answer for what to ask next, two of them adaptation prompts and two further questions, 3) the options to change the length and language level for all the subsequent chatbot answers. The other half of the students, as the control group, did not receive such guidance for their research with the chatbot. The quantitative results from the survey were connected with the qualitative findings from the analysis of the chat protocols.

The results show that:

(1) Learners perceived the chatbot favorably in both groups and there are no significant differences in ratings such as usability and overall satisfaction between the two groups. With or without adaptation guidance, students equally stated that they always knew what to ask next and that the chatbot responses suited their information needs.
(2) Learners who received adaptation guidance instruction and had access to the adaptation features used quantitatively more and qualitatively more varied adaptation

prompts, asked more questions and follow-up questions, and engaged in less off-topic interaction than the control group.

(3) The adaptation guidance led to a significantly higher number of adaptation prompts, while extraneous load and knowledge gain did not differ significantly.
(4) Prior knowledge, chatbot experience, and self-reported self-regulated learning skills had no moderating effects on extraneous load and knowledge gain.

Perceptions of chatbot interaction

Concerning the perceptions of chatbot interaction, two results are of interest: First, students judged their ability to use chatbots for learning as high, even though they used them only a few times a month and they had had teacher instruction on chatbots only once. Second, despite differences in use of adaptation prompts, there is no significant difference regarding several items on satisfaction. On average, students judged their competence as high despite moderate prior experience, and they were readily satisfied with their chatbot interaction.

The students may have rated their skills positively because they lacked a concrete standard of effective chatbot interaction. Self-regulation requires judging one's performance against a standard (Winne, 2013). Given that genAI chatbots are relatively new and students had low prior experience, it is likely they have not yet developed knowledge of effective chatbot interaction. Since no standard for effective learning with chatbots was provided in this study, students may have set a low bar, a common pattern in self-regulated learning. Dunning et al., (2003) dub this the "double curse" of incompetence: The self-regulated learning skills that are necessary to effectively use chatbots during research are also necessary to evaluate one's own performance. Therefore, students need clearer standards for self-regulated learning with chatbots.

Interaction patterns

In terms of interaction patterns, several results stick out: The experimental group used more follow-up questions and more general questions when self-written questions and suggested questions were added up for the EG. They also used significantly more adaptation prompts. Overall, they had more interactions with the chatbot and less off-topic conversation. This suggests that the experimental group was more engaged in the task and the conversation with the chatbot. However, this interpretation is not clearly reflected in the ratings of germane load: Descriptively, the EG rated their germane load as higher ($M = 5.2$) than the control group ($M = 4.9$), but the difference is not significant.

The different types of support have been used differently and, therefore, some of them might have contributed more to this effect. For example, 83% of the participants in the EG used at least one of the next prompts suggested by the chatbot. 18% of the adaptation prompts and 19% of the questions prompted by the students were suggested by the chatbot, so this feature was used considerably often but students also did not overly rely on it. In contrast, only 27% of the students in the EG changed the language level or response length at least once by using the respective buttons. Therefore, the ability to adapt individual messages by asking the chatbot to make them "easier" or explain step-by-step might have been more helpful for the students rather than changing the length or language level for all following responses. Changing the overall length to “shorter” and the language to “simple” might have been undesirable for the students because it would reflect badly on their performance, and also because the students might have felt like they missed out on important information in this configuration.

In terms of overall interaction style, it is notable that, on average, the participants did not prompt the chatbot as if it was a search engine (i.e., with abstracted search terms), but they did not engage with it like a human interlocutor either. Many chatbot interactions were a series of questions covering the general aspects of the topic (i.e., explanation, examples, causes, consequences, and remedies of cognitive biases), with a few adaptation prompts in most conversations and single follow-up questions in many. If one removed all the user inputs and appended all the chatbot outputs from a single user session, many if not most of the chatbot outputs could be characterized as an encyclopedic article. The reasons for this lie both within the chatbot and the users. The fact that chatbots like ChatGPT are fine-tuned to be assistants rather than conversationalists often results in verbose, encyclopedic answers. The learners might quickly adapt to this affordance and enter the habit of learning from texts. While learning from expository texts is very common in formal learning, extended formal learning from a conversation is rarer. Complex, interactive SRL learning activities, as suggested by Chiu (2024), are not present in the control group and to some extent in the group with adaptation guidance.

Overall, both groups showed an interaction style that could be likened to generating an (adapted) encyclopedic article on the topic. The experimental group used more prompts overall, especially more follow-up questions and adaptation prompts, while the control group showed more off-topic conversation which signals a difference in engagement between the two groups.

Main and moderating effects of adaptation guidance

Adaptation prompts

In terms of the main effects hypothesized in this study, only the difference in number of adaptation prompts was significant. The participants in the experimental condition, i.e., with adaptation guidance, used twice as many adaptation prompts. They also used a larger variety of adaptation prompts. For example, participants in the EG asked for an outline or had the chatbot make a table, which was not asked in the control group. Overall, a third of the participants without adaptation guidance did not use any adaptation prompt, not even “Give an example.” This indicates that even though the students reported high confidence in their skill to effectively use chatbots for learning, they do not make use of the basic capabilities when unassisted.

Conversely, given that the students in the experimental group had only received a very short instruction on adaptation prompts and prompt suggestions, this small intervention produced a large effect. This shows that extensive “prompt training” might not be necessary, but that targeted, short instruction prior to a task can be effective.

Extraneous load

The fact that extraneous load did not differ significantly between the groups is in line with research by Abbasi et al. (2019) and contributes empirical evidence to the emerging research that looks into cognitive load while learning with chatbots. Extraneous load (ECL) might not be elevated because chatting with a chatbot is intuitive at the surface level. The participants' requests were usually fulfilled by the chatbot, so they usually experienced the interaction as successful. This is also reflected in the high ratings of ease of use and satisfaction, as discussed above. Because the students did not perceive the versatility the chatbot affords, they had a smaller range of possible next prompts and were not overwhelmed by the plentitude of interaction possibilities. The effect of trial-and-error resulting in high ECL, as predicted by cognitive load theory (Sweller, 2020), does not occur because the range of options is not perceived by the learners.

Generally, instructional prompts can lead to an unintended increase in cognitive load (Horz et al., 2009). In this study, in neither of the three types of load there was a significant difference. The descriptively higher level of ICL in the EG indicates that learners judged the task as slightly more complex than in the control group. In addition to being supportive, the instruction, and maybe also the supportive interface features, raised the task difficulty: Students in the control group did not perceive the task demand of adapting the chatbot responses to their needs. Through the instruction and the supportive features, functioning as reminders, i.e., as instructional prompts, the experimental group likely perceived added task demands in terms of metacognition and self-regulation. This is in line with previous research from Huang et al. (2015) where instructional prompts did reduce ECL but also increased GCL.

Knowledge gains

The fact that the differences between pre- and post-test scores were neither significantly higher nor lower in the experimental group (M = 2.43) than in the control group (M = 2.52) is interesting because the chatbot interactions did show significant differences in terms of adaptation prompts, and also, descriptively in overall engagement operationalized by the number of questions posed overall and the fewer instances of off-topic chat in the experimental group. The larger number of adaptations might have benefitted one aspect of learning but impaired another: Berthold et al. (2011) found evidence that explanation prompts can be double-edged in that they increase conceptual knowledge and detailedness of explanations, while they reduce procedural knowledge and number of tasks performed. Similarly, in this study, the adapted chatbot responses might have led to more readability but also to less challenging content which led to shallower learning. This could have resulted in differential effects on the three rated aspects of the knowledge tests, breadth, depth, and precision. However, there are no significant differences between the conditions when these aspects are tested separately.

The lack of significant difference between the conditions also shows that adaptation guidance does not lead to significantly *lower* knowledge gains. This result would have also been plausible under the assumption that the students requested chatbot answers that were below the cognitive capacities of the students.

Moderating effects

Moderating effects of prior knowledge were hypothesized with three aspects of prior knowledge: domain knowledge, chatbot experience, and SRL skills. There were no moderation effects found for either of these aspects. That is, irrespective of the level of prior knowledge on these three aspects, the students used significantly more adaptation prompts and did not differ significantly in terms of learning gains and extraneous load. Students with lower prior knowledge did not benefit more strongly from the adaptation guidance. Prior domain knowledge and prior chatbot experience were low overall and showed little variance, which was possibly a reason for the lack of moderation effect.

Limitations

There are several limitations pertaining to the task, intervention design, and study design. First, there might have been unintended demand characteristics to the instructional pointers on adaptation prompts that undermine an effect on extraneous load. The hypothesized versatility of the chatbots in terms of ways to adapt content to their needs might have not been perceived in the control group, because the students in the control group did not perceive that the task demands such metacognitive activities to the same degree as the students in the experimental group. The experimental group was not just given hints on *how* to use adaptation prompts but also *that* it is beneficial to use them. This is reflected in the descriptively higher intrinsic and germane cognitive load in the experimental group compared to the control group. If the control group does not perceive self-regulation to be necessary, then there is not a high level of extraneous load that needs reduction to begin with (Seufert, 2018). However, as there is no significant difference between the groups in intrinsic or germane cognitive load, this effect, if present, is not very pronounced, i.e., the experimental group did not perceive the task with adaptation guidance as substantially more complex and did not invest substantially more effort.

Second, while the premise of the study is close to an authentic setting, i.e., conducting research on an unfamiliar topic with the help of a chatbot, there is a limitation to its ecological validity: Students did not have access to web search and were restricted to the chatbot. In everyday use, students would likely combine the chatbot with web search and other tools, possibly to differing degrees. Future studies could explore allowing web search which could shed light on how students switch between chatbot and web search and how they synthesize the results.

Third, the sample size is not large enough to detect small effects. As this is applied research, it is not targeted at detecting small effects, if the intervention requires significant resources. The intervention in this study, however, does not require significant resources and, therefore, even small effects would be of interest so that a larger sample would have been desirable. Future studies could repeat this intervention with a larger sample to investigate smaller effects.

Fourth, it is important to note the short-term nature of this study. The research focused on a single, 20-min interaction with a chatbot, which may not fully capture the potential long-term effects of adaptation guidance. While the findings show immediate benefits in terms of increased use of adaptation prompts, one cannot draw conclusions about how this might impact learning outcomes or self-regulated learning skills over extended periods of use and in other contexts. Future research could explore the long-term effects of adaptation guidance, potentially through longitudinal studies that track students' use of chatbots, their adaptation strategies, and learning outcomes over time.

Practical implications

Some design considerations can be inferred from this study. First, even a short introduction on adaptation prompts like "Give me a summary" or "Make it easier", in combination with easily implementable additions to the chatbot interface suffice to help students adapt the chatbot responses to their needs more often. More specifically, the automatically generated prompt suggestions were used, but not overused, and were perceived as useful. In contrast, the feature to change the language level and length for all subsequent outputs was not used as often. Possibly, there is an aspect of self-disclosure when students opt for shorter answers in simple language. Alternatively, students might have felt they missed out on important information if they had simplified all the chatbot's answers and therefore preferred targeted adaptations to individual messages. Still, the feature was perceived as useful and could be more relevant for learners with lower language skills.

Conclusion

This study investigated K-12 students' perceptions and interaction patterns with generative AI chatbots during a research task, as well as the effectiveness of adaptation guidance in promoting more effective use of these tools. The findings reveal several insights into how secondary school students engage with genAI chatbots in an authentic scenario.

First, students generally perceived the chatbots favorably and reported high levels of satisfaction and ease of use, regardless of whether they received adaptation guidance. This perception of the chatbot as "easy", suggests an underinvestment of mental effort (Salomon, 1984; Stadler et al., 2024) and, combined with students' moderate prior experience, a potential overestimation of their ability to effectively use these tools for learning. The underutilization of core affordances, such as adapting answers to the learner's needs becomes apparent in the fact that a third of the students without adaptation guidance did not query the chatbot for any adaptation. Second, the results demonstrate that even brief instruction on adaptation prompts can significantly influence how students interact with chatbots. Students who received adaptation guidance used a greater variety and a significantly higher number of adaptation prompts, asked more follow-up questions, and engaged in less off-topic conversation. This indicates that modest interventions can potentially enhance students' interaction with chatbots. Using the full range of chatbot capabilities in self-regulated learning might still be "tough", but this can be effectively alleviated by targeted support.

This study has approximated a scenario in which students research new information with the help of a chatbot, which is a major use case of chatbots among students (e.g., von Garrel & Mayer, 2023). Future studies could investigate this type of learning task in a more natural setting, and they could interview students to gain a more nuanced understanding of learner perceptions and interaction patterns. Furthermore, similar targeted interventions could be designed to help with further aspects of self-regulated learning, such as investing an appropriate amount of mental effort or choosing effective strategies for planning, performing, and evaluating the learning process. Overall, this line of research can contribute to a deeper understanding of how genAI chatbots can enhance learning and what conditions foster successful learning with these tools.

Acknowledgements
Not applicable

Author contributions
The author has planned, designed, and conducted the study and written this article.

Funding
Open Access funding enabled and organized by Projekt DEAL. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials
The datasets used and analyzed during the current study are available from the author on reasonable request.

Declarations

Competing interests
The author declares that they have no competing interests.

Received: 29 October 2024 Accepted: 31 March 2025
Published online: 08 May 2025

References

Abbasi, S., Kazi, H., & Hussaini, N. (2019). Effect of chatbot systems on student's learning outcomes. *Sylwan, 163*(10), 49–63.

Abdelhalim, S. M. (2024). Using ChatGPT to promote research competency: English as a foreign language undergraduates' perceptions and practices across varied metacognitive awareness levels. *Journal of Computer Assisted Learning, 40*(3), 1261–1275. https://doi.org/10.1111/jcal.12948

Antonenko, P. D., Dawson, K., & Sahay, S. (2017). A framework for aligning needs, abilities and affordances to inform design and practice of educational technologies. *British Journal of Educational Technology, 48*(4), 916–927. https://doi.org/10.1111/bjet.12466

Azevedo, R., Feyzi Behnagh, R., Duffy, M., Harley, J., & Trevors, G. (2012). Metacognition and self-regulated learning in student-centered learning environments. In: S. Land & D. Jonassen (Eds.), *Theoretical foundations of student-center learning environments* (2nd ed., pp. 171–197). Routledge. https://doi.org/10.4324/9780203813799

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., Cloude, E., Dever, D., Wiedbusch, M., Wortha, F., & Cerezo, R. (2022). Lessons learned and future directions of MetaTutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology, 13*, 813632. https://doi.org/10.3389/fpsyg.2022.813632

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior, 27*(1), 69–75. https://doi.org/10.1016/j.chb.2010.05.025

Chiu, T. K. F. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of ChatGPT. *Educational Technology Research and Development, 72*, 2401–2416. https://doi.org/10.1007/s11423-024-10366-w

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: A theoretical framework and research Agenda. *Educational Psychology Review, 32*(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83–87. https://doi.org/10.1111/1467-8721.01235

Gentner, N., & Seufert, T. (2020). The double-edged interactions of prompts and self-efficacy. *Metacognition and Learning, 15*(2), 261–289. https://doi.org/10.1007/s11409-020-09227-7

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (Second edition.). The Guilford Press.

Horz, H., Winter, C., & Fries, S. (2009). Differential benefits of situated instructional prompts. *Computers in Human Behavior, 25*(4), 818–828. https://doi.org/10.1016/j.chb.2008.07.001

Huang, K., Chen, C.-H., Wu, W.-S., & Chen, W.-Y. (2015). Interactivity of question prompts and feedback on secondary students' science knowledge acquisition and cognitive load. *Educational Technology & Society, 18*(4), 159–171.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., & Krusche, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Klar, M. (2024). How should we teach chatbot interaction to students? A pilot study on perceived affordances and chatbot interaction patterns in an authentic K-12 setting. Proceedings of DELFI Workshops 2024. https://doi.org/10.18420/delfi2024-ws-22. Gesellschaft für Informatik e.V.

Klaykaew, K. K., & Prasittichok, P. (2024). The effects of the ADMIRE program on the learning motivation and self-regulated learning of university students. *European Journal of Educational Research, 13*(1), 197–205. https://doi.org/10.12973/eu-jer.13.1.197

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Kotek, H., Dockum, R., & Sun, D. (2023). Gender bias and stereotypes in Large Language Models. In: *Proceedings of The ACM Collective Intelligence Conference*, 12–24. https://doi.org/10.1145/3582269.3615599

Laine, E., Veermans, M., Gegenfurtner, A., & Veermans, K. (2020). Individual interest and learning in secondary school STEM education. *Frontline Learning Research, 8*(2), 90–108.

Lim, L., Lim, S. H., & Lim, R. W. (2022). Measuring learner satisfaction of an adaptive learning system. *Behavioral Sciences, 12*(8), 264. https://doi.org/10.3390/bs12080264

Magno, C. (2011). Validating the academic self-regulated learning scale with the motivated strategies for learning questionnaire (MSLQ) and learning and study strategies inventory (LASSI). *The International Journal of Educational and Psychological Assessment, 7*(2), 56–73.

Mogavi, R. H., Deng, C., Kim, J. J., Zhou, P., Kwon, Y. D., Metwally, A. H., Tlili, A., Bassanelli, S., Bucchiarone, A., Gujar, S., & Nacke, L. E. (2024). ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans, 2*(1), 100027. https://doi.org/10.1016/j.chbah.2023.100027

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

Ngo, T. T. A. (2023). The perception by university students of the use of ChatGPT in education. *International Journal of Emerging Technologies in Learning (iJET), 18*(17), 4–19. https://doi.org/10.3991/ijet.v18i17.39019

Playfoot, D., Quigley, M., & Thomas, A. G. (2024). Hey ChatGPT, give me a title for a paper about degree apathy and student use of AI for assignment writing. *The Internet and Higher Education, 62*, 100950. https://doi.org/10.1016/j.iheduc.2024.100950

Pozdniakov, S., Brazil, J., Abdi, S., Bakharia, A., Sadiq, S., Gašević, D., Denny, P., & Khosravi, H. (2024). Large language models meet user interfaces: The case of provisioning feedback. *Computers and Education: Artificial Intelligence, 7*, 100289. https://doi.org/10.1016/j.caeai.2024.100289

Salomon, G. (1984). Television is 'easy' and print is 'tough': The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology, 76*(4), 647–658. https://doi.org/10.1037/0022-0663.76.4.647

Schuster, C., Stebner, F., Leutner, D., & Wirth, J. (2020). Transfer of metacognitive skills in self-regulated learning: An experimental training study. *Metacognition and Learning, 15*(3), 455–477. https://doi.org/10.1007/s11409-020-09237-5

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

Seufert, T. (2020). Building bridges between self-regulation and cognitive load—An invitation for a broad and differentiated attempt. *Educational Psychology Review, 32*(4), 1151–1162. https://doi.org/10.1007/s10648-020-09574-6

Shoufan, A. (2023). Exploring students' perceptions of ChatGPT: Thematic analysis and follow-up survey. *IEEE Access, 11*, 38805–38818. https://doi.org/10.1109/ACCESS.2023.3268224

Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, 108386. https://doi.org/10.1016/j.chb.2024.108386

Stebner, F., Schuster, C., Dicke, T., Karlen, Y., Wirth, J., spsampsps Leutner, D. (2019). The effects of self-regulation training on self-regulated learning competencies and cognitive load. In: S. Tindall-Ford, S. Agostinho, spsampsps J. Sweller (Eds.), *Advances in Cognitive Load Theory* (1st ed., pp. 194–208). Routledge. https://doi.org/10.4324/9780429283895-16

Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. *Computers and Education: Artificial Intelligence, 7*, 100259. https://doi.org/10.1016/j.caeai.2024.100259

Sweller, J. (2011). Cognitive Load Theory. In *Psychology of Learning and Motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development, 68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Theophilou, E., Koyutürk, C., Yavari, M., Bursic, S., Donabauer, G., Telari, A., Testa, A., Boiano, R., Hernandez-Leo, D., Ruskov, M., Taibi, D., Gabbiadini, A., & Ognibene, D. (2023). Learning to prompt in the classroom to understand AI limits: A pilot study. In R. Basili, D. Lembo, C. Limongelli, & A. Orlandini (Eds.), *AIxIA 2023—Advances in artificial intelligence* (pp. 481–496). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-47546-7_33

Thompson, P. (2013). The digital natives as learners: Technology use patterns and approaches to learning. *Computers & Education, 65*, 12–33. https://doi.org/10.1016/j.compedu.2012.12.022

Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments, 10*(1), 15. https://doi.org/10.1186/s40561-023-00237-x

von Garrel, J., & Mayer, J. (2023). Artificial Intelligence in studies—Use of ChatGPT and AI-based tools among students in Germany. *Humanities and Social Sciences Communications, 10*, 799. https://doi.org/10.1057/s41599-023-02304-7

Wang, T., & Lajoie, S. P. (2023). How does cognitive load interact with self-regulated learning? A dynamic and integrative model. *Educational Psychology Review, 35*(3), 69. https://doi.org/10.1007/s10648-023-09794-6

Weinstein, C. E., Zimmermann, S. A., & Palmer, D. R. (1988). 3–Assessing learning strategies: the design and development of the lassi. In C. E. Weinstein, E. T. Goetz, & P. A. Alexander (Eds.), *Learning and study strategies* (pp. 25–40). Academic Press. https://doi.org/10.1016/B978-0-12-742460-6.50009-8

Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Boonyakitanont, P., & Jinowat, N. (2024). Designing chatbots in language classrooms: An empirical investigation from user learning experience. *Smart Learning Environments, 11*(1), 32. https://doi.org/10.1186/s40561-024-00319-4

Winne, P. H. (2013). Learning strategies, study skills, and self-regulated learning in postsecondary education. In M. B. Paulsen (Ed.), *Higher education: Handbook of theory and research* (Vol. 28, pp. 377–403). Springer. https://doi.org/10.1007/978-94-007-5836-0_8

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models* (No. arXiv:2309.01219). Preprint: arXiv. https://doi.org/10.48550/arXiv.2309.01219

Publisher's Note