Special Issue: Human Factors of AI

Human Factors
2026, Vol. 0(0) 1–28

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00187208261474518
journals.sagepub.com/home/hfs

Learning and Engaging With AI: An Exploration of the Effects of Mental Workload and Search Behavior on Short-Term Computer-Based Learning With a Chatbot

Alexandre Marois[1,2], Isabelle Lavallée[3], Gabrielle Boily[1], and Jonay Ramon Alaman[1]

Abstract

Objective: This study examined how learning, workload and search behaviors were impacted by a chatbot during a self-regulated Web search task, as opposed to more classic search engines.
Background: Artificial intelligence technologies, including chatbots, are becoming increasingly accessible. These tools have been demonstrated useful to support self-regulated learning, mostly in structured learning contexts. The reduction in workload they offer may, however, prevent key learning strategies from being deployed, especially for Web search.
Method: Sixty participants were asked to answer a set of essay questions and to rate their workload, effort deployed and literacy while either gathering information from the Internet (Web condition) or by chatting with a chatbot (LLM condition) with the possibility of verifying information on the Web. Several key search strategies and chatbot interaction measures were extracted. A surprise memory test was also presented to evaluate how they learned the content addressed in the essay questions.
Results: Measures of effort, mental workload, search behaviors and literacy differed significantly across conditions. Performance on the memory test did not vary. Multiple relationships with memory performance, including key Web search and verification behaviors, were found.
Conclusion: Chatbots may help reduce workload and short-term learning with a chatbot may be more influenced by the nature of the interaction with the tool, rather than the tool itself.
Application: Effective uses of chatbots may require learners to verify the content generated by the chatbot and to show superior engagement. Engagement-promoting learning activities should be considered when using LLM-driven agents to support self-regulated, Web-based search.

[1]École de psychologie, Université Laval, Québec, QC, Canada
[2]School of Psychology and Humanities, University of Lancashire, Preston, La, UK
[3]Département de psychologie, Université de Sherbrooke, Sherbrooke, QC, Canada

Corresponding Author:
Alexandre Marois, École de psychologie, Université Laval, Pavillon Félix-Antoine-Savard, local 1328, 2325, rue des Bibliothèques, Québec, QC G1V 0A6, Canada.
Email: alexandre.marois@psy.ulaval.ca

Keywords
AI, chatbot, learning, mental workload, web search

Key Points

- Chatbots can support self-regulated learning activities and reduce workload. Learning may however benefit from a certain amount of effort, including for self-regulated Web search tasks.
- Participants performed a Web search task with or without support from a chatbot, followed by a memory test.
- Workload and key search strategies differed across conditions, although memory performance remained similar.
- The nature of the interaction with a tool, rather than the tool itself, mainly affects the learning experience.

Introduction

The recent growth in performance and accessibility of artificial intelligence (AI) tools impacts our ways of living. Generative AI is a technology that can generate content traditionally reserved to humans with reduced time and effort. This includes technologies such as ChatGPT (OpenAI) and other tools that use natural language processing and large language models (LLM), which typically take the form of conversational agents (or chatbots) that can process and generate text similar to what humans can accomplish (Cao et al., 2023). LLM-based chatbots have been studied as tools to support knowledge-centered activities such as writing and learning (Bai et al., 2023; Rice et al., 2024), especially for structured learning activities (e.g., Kohnke et al., 2023). Yet, opportunities for learning with chatbots go way beyond scholarly uses, including punctual Web searches and other computer-based learning activities. The present study focuses on understanding how interacting with chatbots vs. Web search tools may affect short-term, computer-based self-regulated learning and on investigating the interplay with mental workload.

Computer-Based Learning, Web Searches and Mental Workload

Self-regulated learning is a type of activity highly prevalent in computer-based learning environments (Al-Shloul et al., 2024; Greene et al., 2011). It is commonly conceptualized as unfolding across a set of successive phases that involve different metacognitive processes. Classically, Zimmerman's (2000) model proposes three successive phases: (1) a forethought phase, with goal setting and strategic planning; (2) a performance phase, involving metacognitive monitoring and control; and (3) a self-reflection phase, activating self-evaluation and reflection processes. Self-regulated learning models, however, may differ in the number and taxonomy of the phases they propose (see, e.g., Pintrich, 2004; Winne & Hadwin, 1998). For instance, Cheng et al. (2013) propose a two-phase model specifically adapted to online help seeking for Web-based learning, which distinguishes *basic self-regulation* (i.e., one's awareness of insufficient knowledge and goal setting according to said knowledge) from *advanced self-regulation* (i.e., the monitoring of learning processes, including the adoption of retrieval strategies and self-evaluation outcomes). Each phase can in turn be associated with a set of trace behaviors and micro-actions performed during self-regulated learning activities (Chiu et al., 2013; Colville & Ostern, 2026; Lai et al., 2025; Lyu & Ding, 2025; Maldonado-Mahauad et al., 2018; Sun et al., 2023; Zhou, 2013). Across these steps, metacognitive abilities are engaged in managing mental workload to dynamically adjust the focus of attention towards—and resources assigned to—different pieces of information, affecting how said information is learned. In the present study, we adopt Cheng et al.'s (2013) two-phase model given our focus on the interaction with online tools for self-regulated learning.

Web search is a highly prevalent self-regulated, computer-based learning activity (Narciss et al., 2007). Typically, Web searches are conducted using search engines. These tools allow individuals to easily access information from multiple sources on the Internet. Information on the Internet is, however, quasi-infinite and structured in a complex, nonlinear and non-hierarchical way. Therefore, successful learning from Web searches requires a set of abilities, including minimal domain knowledge, self-regulation of actions and cognitive resources, technical capacities, and strategic (keyword) search behaviors (Gwizdka, 2010; von Hoyer et al., 2022; Willoughby et al., 2009), all of which can be categorized under different self-regulated learning phases as proposed by Cheng et al. (2013) (see also Chiu et al., 2013; Lai et al., 2025; Maldonado-Mahauad et al., 2018; Sun et al., 2023; Zhou, 2013). These different activities may produce an increase in mental workload (Lidstone & Lucas, 1998). As supported by the cognitive load theory (Sweller, 1988, 2011), too much information can increase cognitive load and impede performance. This has been shown in Web search research, where learners faced cognitive overload over too many sources of information (Mayer et al., 2001; Mayer & Moreno, 2003).

Yet, research also shows that encountering certain difficulties that demand mental effort during the learning process can be beneficial (Bjork & Bjork, 2020). These are known as "desirable difficulties." Desirable difficulties are useful because they activate various encoding and retrieval processes that promote learning, comprehension, and memorization. However, some difficulties are undesirable. When learners encounter too many difficulties in regulating their learning, they may achieve suboptimal memorization (Bannert et al., 2015; Kizilcec et al., 2017). Alternatively, when the demands are too low, failing to mobilize sufficient cognitive resources may also result in suboptimal learning. This is supported by empirical work on cognitive offloading, that is, the delegation of mental operations to external tools (Risko & Gilbert, 2026). Mental workload and cognitive effort are in fact considered key components of learning, even for self-regulated learning, as supported by recent theoretical perspectives that integrate cognitive load factors in self-regulated learning models (e.g., Seufert et al., 2024; Wang & Lajoie, 2023; Wirth et al., 2020; see also de Bruin et al., 2020, 2025).

According to Wang et al. (2019), chatbots can help reduce some undesirable difficulties, namely extraneous cognitive load—that is, task-irrelevant cognitive load—by providing detailed and personalized feedback to learners. For example, Schmidhuber et al. (2021) observed that participants interacting with a chatbot in a realistic enterprise setting reported lower cognitive load using the NASA-Task Load Index (NASA-TLX) questionnaire, and improved productivity as opposed to participants with no chatbot. Similar conclusions were reached by Stadler et al. (2024) using a Web search task. Participants were asked to produce recommendations for a fictional socio-scientific issue regarding nanoparticles in sunscreen, an unsettled complex topic that requires weighing competing claims, with the aid of ChatGPT or traditional Web search with Google. The ChatGPT condition reported lower measures of mental workload. Still, participants interacting with ChatGPT demonstrated lower-quality reasoning and argumentation in their recommendations. The authors suggested that this lower quality for the ChatGPT group could be explained by reduced opportunities for learning. Indeed, they outlined that interacting with more diverse and challenging sources of information (through Web search engines) can promote higher comprehension and processing of learning material, engaging learners more deeply.

A reduction in mental workload while interacting with a chatbot may thus potentially impact key self-regulated learning components and even some desirable difficulties by lowering the need to actively search for and verify information. For instance, Lyu and Ding (2025) found that 82% of the actions undertaken by chatbot users during an English writing task focused on task execution with poor engagement into planning and reflection.

However, evidence from formal learning contexts suggests that the influence of chatbots may sometimes be positive. A meta-analysis by Wu and Yu (2024) found that chatbots can improve long-term learning performance, especially among university students, because these systems can offer structured guidance, personalized feedback, and chances for iterative elaboration of concepts. At the same time, excessive reliance may also reduce cognitive engagement and long-term retention (Bai et al., 2023). Importantly, most evidence regarding the effects of chatbots is based on studies involving long-term usage (e.g., over multiple weeks during a course). Still, the literature on short-term effects of chatbots on learning is rapidly growing.

In Ju (2023), participants were asked to read an article, write a summary, and answer five recall questions about its content. Three groups were randomly formed: (a) the manual group (with no help from an AI); (b) the AI group (with help from an AI to write the summary); and (c) the active group (with reading guidance from a conversational agent). The AI group obtained a 25.1% decrease in correct recall compared to the manual group, while performance of the active group was reduced by 12%. In addition, the active and AI groups generally took less time to complete the task than the manual group. Kosmyna et al. (2025) reached similar conclusions. Participants were asked to write four essays with the help from ChatPT (LLM group), help from the Internet (Search engine group) or no support tool (Brain-only group). After each essay, participants were queried on multiple aspects, including their ability to quote excerpts from the essay they wrote and the correctness of their quotes. Participants from the LLM group generally struggled more to report direct quotes from their essays and made more errors in their quotes. EEG spectral band analysis also suggested lower engagement for the LLM group compared to both Search engine and Brain-only groups. However, it remains difficult to draw conclusions from these results given the lack of a full portrait of the questions asked and the specific focus on direct quote citations.

The Present Study

Overall, the literature suggests that chatbot support in computer-based learning can help reduce mental workload. Such a reduction may be beneficial to prevent the large amount of information to manage that might induce overload. Nonetheless, this could also prevent key self-regulation cognitive processes to be deployed, rather reducing the quality of learning. While earlier work mostly focused on structured learning activities, more recent research examined the short-term and immediate effects of chatbot support on learning. A growing body of research has aimed at examining short-term outcomes such as task performance, cognitive effort, and immediate memory recall following web search/chatbot interactions (e.g., Akgun & Toker, 2025; Ju, 2023; Kosmyna et al., 2025; Stadler et al., 2024). However, these studies mostly provide a partial portrait of the effects of chatbots interactions on learning and mental workload. Critically, they also studied chatbot uses in a very restrictive and controlled context, unrepresentative of real-life interactions where it remains possible to verify the information provided by the chatbot.

The current study investigates the impact of interacting with a chatbot as opposed to relying on more traditional browsing strategies to perform a realistic computer-based Web search task. Participants were asked to answer brief essay questions on a diversity of subjects with the help from a search engine (Web condition) or from an LLM-driven agent (LLM condition) with the possibility of verifying information from the Web. Participants were also shown a surprise memory test covering the subjects addressed in the essay questions. Measures of self-reported workload were collected to assess differences across conditions and explore their relationship with memory performance. Several search process indicators (e.g., number of keywords, strategies for answering the question, Web

verification, number of sources collected) and abilities with the tools used were also examined. We hypothesized that interacting with the chatbot would reduce workload and, concurrently, reduce memory performance given the risks for cognitive offloading and decreased opportunity for desirable difficulties to be experienced, which generally allow for learning strategies to be deployed. These measures were also expected to be associated with some of the search process indicators.

Method

Participants

Sixty participants (27 women, 33 men; M_{age} = 28.93, SD = 13.36) took part in the study in exchange for a 20-CAD honorarium. The inclusion criteria were to be aged over 18 years and to be able to complete a computerized task in the lab, in French. Participants varied in terms of educational background (highest diploma: 10% high school degree; 28.3% college degree; 35% bachelors' degree; 25% master's degree; and 1.7% doctoral degree) and were similar on their level of proficiency in French (M = 9.29 points out of 10, SD = 0.84). They were randomly assigned to either the chatbot LLM condition or the Web condition (see Table 1 for the detailed sociodemographic characteristics for each condition). Recruitment was carried out through institutional emails and posters displayed throughout the Université Laval campus. This research complied with the tenets of the Declaration of Helsinki and was approved by the Institutional Review Board at Université Laval. Informed consent was obtained from each participant.

Apparatus and Material

The main task involved interacting with an HTML-based testbed interface developed specifically for this study (Figure 1; see also https://github.com/LEILAHUL/chatbot-websearch-memory-testbed). Participants were presented 12 essay questions, randomly chosen from a 39-question bank common to all participants, that each had to be answered between 75 and 100 words (e.g., "What are the main environmental challenges for sea turtles?"; see Appendix A for the full list of questions). They were asked to answer each question under 10 min using the experimental interface. Each question was developed to be answered within the word limit imposed and to require the inclusion of six to nine predefined key elements, on average. The questions were produced in accordance with the "Remembering" and "Understanding" classifications of Anderson & Krathwohl's (2001) revision of Bloom's taxonomy, meaning that they required participants to recall knowledge or construct meaning from the different sources of information they found. For instance, for the sea turtle question presented above, participants were expected to include elements such as climate change, natural predators, pollution, disease or hunting.

Participants in the LLM condition were asked to interact with a chatbot, namely the Phi-3 LLM implemented into the LM Studio interface (Figure 2). To be representative of real-life interactions, they had the possibility of verifying the chatbot's answers on the Web. At the beginning of the experiment, participants from the Web condition were shown the Google opening page. They were told they could navigate the Internet without using any AI tool. Data collection began before Google introduced AI Overviews, and the feature was deactivated once it became available using the Google Chrome Disable AI Overview | Turn Off AI Overview extension (hideaioverviews; see https://chromewebstore.google.com/detail/hide-google-ai-overviews/neibhohkbmfjninidnaoacabkjonbahn). After each essay question, participants rated the extent of efforts deployed on a scale between 1 and 5, and their prior knowledge on the subject on a 1-to-10 scale.

Once the 12 essay questions were answered, a questionnaire addressing the participants' capacity to interact with their respective tool was presented. Participants from the LLM condition completed the ChatGPT Literacy Scale (25 five-point Likert items related to one's capacity to interact with

Table 1. Depiction of the two conditions' sociodemographic characteristics

Characteristic	Condition	
	Web (n = 30)	LLM (n = 30)
Mean age (years)	27.9	29.9
SD age (years)	12.7	14.1
Gender (%)		
Female	46.7	43.3
Male	53.3	56.7
Mean level of proficiency in French (/10)	9.1	9.5
SD level of proficiency in French (/10)	0.92	0.69
Highest diploma (%)		
High school	10	10
College	33.3	23.3
Bachelor's	33.3	36.7
Master's	23.3	26.7
Doctoral	0	3.3

a chatbot; Lee & Park, 2024), whereas participants in the Web condition answered the Internet Skills Scale (26 five-point Likert items focused on skills related to Web browsing; van Deursen et al., 2015). Both questionnaires, although distinct, allow measuring to what extent participants are familiar and capable to interact with either chatbots or the Internet. Following this questionnaire, participants completed the NASA-TLX questionnaire (Hart & Staveland, 1988) to rate their perceived effort and performance during the essay task. It included six 10-point Likert subscales: mental demand, physical demand, temporal demand, effort, performance and frustration. Note that, in line with Taylor and Bolton (2026) recommendation, the individual subscales of the NASA-TLX, rather than the initial standard averaged or weighted recommended metrics, were used (see also Bolton et al., 2023; Felton et al., 2012; Galy et al., 2018). This recommendation is based on evidence that the NASA-TLX dimensions are sometimes characterized by important multicollinearity issues, and the independent subscales allow for improved diagnosticity. In addition, the Frustration and Physical subscales of the NASA-TLX were removed from the analysis because they were not relevant to the current experiment and because prior evidence suggests that they tend to contribute less consistently to the actual construct of (mental) workload (Hertzum, 2021). Then, participants were presented with an unexpected recall test consisting of 12 short-answer questions, displayed in random order. An average of

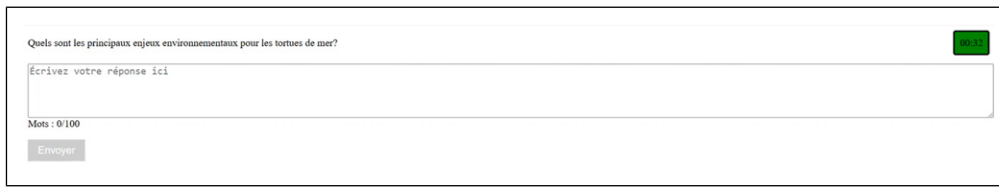

Figure 1. HTML-based testbed interface presented to display the different essay and memory questions of the main task. The prompt, originally displayed in French, reads: "What are the main environmental challenges for sea turtles?"

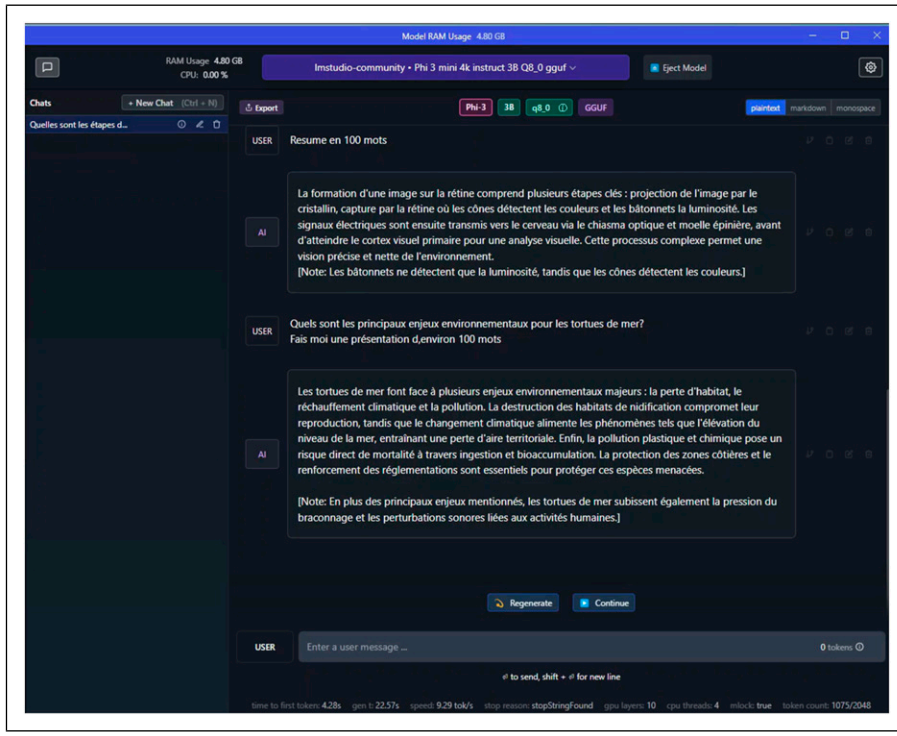

Figure 2. Depiction of the LM Studio interface used by the participants from the LLM condition to discuss with the chatbot

30.45 min (*SD* = 6.89) passed between the last essay question and the first memory question. Each question addressed content from one of the 12 essay questions (e.g., "Name a human activity that has a negative ecological impact on turtles."). Participants were expected to answer, under 1 min without any aid, with a single element that was reported in the essay question if answered correctly.

Procedure

After having signed a pre-experimental consent form and be explained the general purpose of the study, participants were presented the essay questions and the tool they had to work with (either Internet and/or the LLM chatbot). Each essay question was followed by single items asking participants to rate their effort deployed and their prior knowledge of the content addressed. Following the 12 essay questions, participants were presented with the tool literacy scale, depending on their condition. Then, they filled the NASA-TLX, followed by the surprise memory test. After the memory test, a post-experimental consent form was also presented to explain the true purpose of the study, given that the memory test was not mentioned in the initial agreement. Throughout the study and after each consent form (pre- and post-experiment), participants were informed they could freely withdraw from the study without any disadvantage. Finally, participants received their honorarium and were thanked for their participation. Figure 3 summarizes the complete procedure that participants went through.

Analysis

Three participants were removed because of recording issues, and one participant because they failed to properly understand items from the NASA-TLX and effort deployed

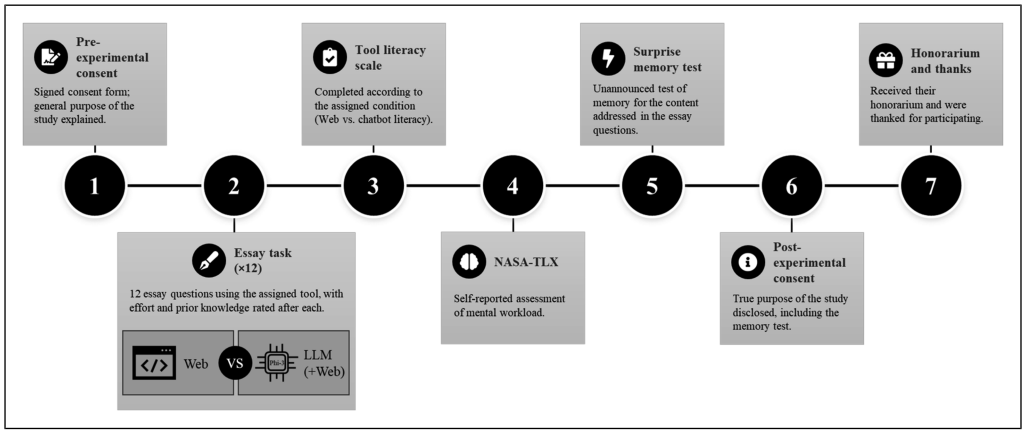

Figure 3. Depiction of the procedure that participants followed throughout the study

questionnaires due to language miscomprehension (reported after the experiment), resulting in 29 participants in the Web condition and 27 in the LLM condition. All essay and memory questions were corrected by two trained raters according to an answer sheet. Responses on the essay questions were scored against a comprehensive answer key that considered the six to nine key elements for each question. A proportional scoring scheme was applied, meaning that a response containing only half of the expected elements received approximately half of the available points. This approach reflects the breadth of information retrievable on each topic rather than a binary correct/incorrect judgment, and consequently can produce scores that may appear modest even if the information provided is factually correct.

The following dependent variables were analyzed and compared across both conditions using Mann-Whitney tests: (a) accuracy (in %) and response time (RT, in s) on the essay questions; (b) accuracy (in %) and RT (in s) for the memory questions; (c) number of queries (either on a search engine or to the chatbot); (d) % of copied and pasted questions to collect information on the essay questions; (e) number of websites consulted; (f) frequencies for the different answering strategies on the essay questions (complete copy-paste vs. copy-paste + modifications [mixed] vs. self-elaborated writing); (g) mean effort deployed (/5); (h) mean prior knowledge (/10); (i) mean score on the tool literacy questionnaire (either the ChatGPT Literacy Scale or the Internet Skills Scale, in %); and (j) mean values for the mental, temporal, effort and performance subscales of the NASA-TLX (/10). Nonparametric tests were used because of normality issues emerging in most of the variables (Shapiro-Wilks tests: $ps < .05$).

We also performed condition-specific analyses. In the Web condition, we analyzed the nature of the keywords used during the Web search from a search engine optimization perspective (Erdmann et al., 2022), focusing on the occurrence of phrases vs. primary (theme-specific) keywords vs. secondary (theme-related) keywords (Meghanathan et al., 2010). In the LLM condition, we analyzed the percentage of essay questions where participants verified information on the Web and the number of requests performed for this purpose. We also characterized the prompting approach for each essay question, according to prompt engineering terminology (e.g., Google Cloud, 2025; IBM, 2025), that is zero-shot (direct query with no context) vs. context-setting (one or few context instructions before the query) prompts. For each condition, exploratory correlation analyses were performed.

With this information, we also conducted exploratory hierarchical regression analyses

separately for each condition. Hierarchical regression was selected because it allows testing whether incrementally added sets of predictors explain additional variance beyond the preceding sets (Tabachnick & Fidell, 2013). In both conditions, memory performance was predicted from three successive blocks of variables: (1) basic self-regulation, (2) advanced self-regulation, and (3) workload. The first two blocks were derived from Cheng et al.'s (2013) two-phase model of self-regulated learning. This two-component organization was adopted because prior work has suggested that more fine-grained three-component models of self-regulated learning may be too complex to fit empirical data adequately (cf. Pintrich et al., 2000; see also Chiu et al., 2013). The third block was added to account for the hypothesized impact of workload and in coherence with previous theoretical perspectives integrating cognitive load factors to self-regulated learning models (e.g., Seufert, 2018; Wang & Lajoie, 2023; Wirth et al., 2020).

The first block, corresponding to basic self-regulation, included variables representing participants' initial knowledge state: tool literacy, prior knowledge, and accuracy on the essay questions. These variables were representative of the prior knowledge state of the participants. The second block, corresponding to advanced self-regulation, included indicators of participants' information-retrieval and response-construction processes: prompting/keyword strategy depth, response strategy depth, and number of websites visited. Prompting/keyword strategy depth and response strategy depth were computed as normalized ordinal indices reflecting the extent to which participants generated, reformulated, or elaborated their own queries and answers. Within each index, categories were rank-ordered according to their assumed level of participant-generated elaboration, with higher-ranked categories receiving higher scores. These scoring values were used to preserve the ordinal structure of the categories and should not be interpreted as proportional weights indicating that one category was twice or three times as important as another. The indices were computed following equations (1)–(3):

$$\text{Keyword depth}_{\text{Web}} = \frac{(\%\,\text{copy-paste} \times 1) + (\%\,\text{phrases} \times 2) + (\%\,\text{primary keywords} \times 3) + (\%\,\text{secondary keyword} \times 4)}{4} \quad (1)$$

$$\text{Prompt depth}_{\text{LLM}} = \frac{(\%\,\text{copy-paste} \times 1) + (\%\,\text{zero} - \text{shot prompts} \times 2) + (\%\,\text{context} - \text{setting prompts} \times 3)}{3} \quad (2)$$

$$\textit{Response depth} = \frac{(\%\,\textit{copy-paste} \times 1) + (\%\,\textit{mixed} \times 2) + (\%\,\textit{self-elaborated} \times 3)}{3} \quad (3)$$

where Keyword $\text{depth}_{\text{Web}}$ refers to the depth of participants' Web-search queries in the Web condition across all Web searches, Prompt $\text{depth}_{\text{LLM}}$ refers to the depth of participants' prompts submitted to the LLM in

the corresponding condition, and Response depth refers to the degree of elaboration of participants' responses to the essay questions. In equation (1), % copy-paste refers to the amount of Web search queries that was directly copied from the interface to the web browser, % phrases refers to the searches formulated as a sentence reformulating the initial question, % primary keywords refers to searches that relied only on primary keywords, and % secondary keywords refers to the amount of Web search queries that included also secondary keywords. In equation (2), % copy-paste refers to the percentage of prompts that directly contained the initial question, zero-shot prompts refer to situations where the participants formulated a single prompt to receive answer on the question, whereas context-setting prompts refers to situations where participants provided one or few context instructions before the query. In equation (3), % copy-paste refers to questions whose answer was completely copied from the web platform to the interface and was not modified by the user before submitting, % mixed refers to the percentage of answers combining copied or externally generated material with participant-generated content, whereas % self-elaborated refers to the percentage of answers that were completely formulated by the participant. These composite variables allowed multiple micro-actions related to information retrieval and response construction to be summarized while minimizing the number of predictors entered into the regression models. For each index, the weighted ordinal sum was divided by the maximum possible category score for that index, allowing scores to be expressed relative to the maximum level of strategic elaboration. Higher values therefore indicated greater participant-generated manipulation of information. The third block, corresponding to workload, included the NASA-TLX subscales of performance assessment, temporal pressure, mental load, and effort. This block was entered last to assess whether perceived workload explained additional variance in memory performance after accounting for basic and advanced self-regulation. Adjusted-R^2 measures computed from both Wherry's (1931) classic and Stein's (1960) cross-validation formula were reported.

Given the low sample size and the risks of overgeneralization with a small amount of data points, we also report Bayesian equivalent tests. The Bayesian analysis focused on reporting Bayes factors of the alternate hypothesis (BF_{10}) following the classification scheme proposed by Jeffreys (1961). More specifically, we report the BF_{10} values based on data augmentation algorithm with five chains of 1,000 iterations as well as qualitative interpretation of said values. Bayesian-equivalent linear regressions using the full set of predictors used in the last step of the hierarchical regressions were ran using the beta binomial method because it is considered more conservative with complex models. All statistical analyses were performed using JASP 0.96.0.0.

Results

Confirmatory Analyses: Comparison Across Conditions

The trained raters reached excellent agreement on the correction of the essay and memory questions (ICC = .94 and .95 for the essay and memory questions, respectively). Raters then came to an agreement to set the final scores (see Figure 4). The mean accuracy measures on the essay and memory questions, collapsed across all participants, were marginally positively correlated (r_S = .31, p = .055). Accuracy on the essay questions followed a normality pattern (W = .979, p = .659) whereas the accuracy measures of the memory questions was positively skewed and violated normality of the distribution (W = .932, p = .022).

Table 2 presents the dependent variables commonly collected across both conditions, along with the results of the Mann-Whitney tests. Significant differences emerged mostly between the conditions regarding the measures of Web search and mental workload.

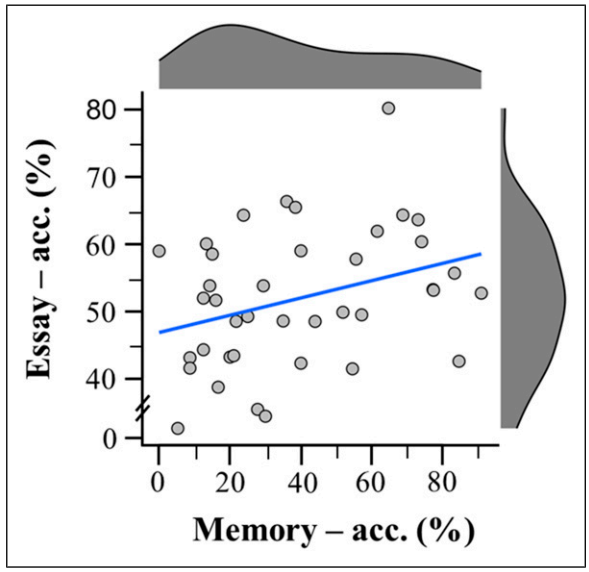

Figure 4. Distribution and density of the accuracy (acc., in %) measures for all the essay and memory questions, averaged across all participants

Participants used a larger number of queries/prompts and copied-pasted less frequently the questions from the experiment interface in the Web condition, and they explored a larger number of websites and Web pages than those in the LLM condition. The Bayesian analyses provided strong to extreme evidence favoring *H*1, supporting the significant differences found by the frequentist tests. Measures of effort deployed were also significantly higher in the Web condition compared with the LLM condition, with moderate evidence favoring *H*1 from the Bayesian analysis. The same pattern was observed for the mental effort subscale of the NASA-TLX, but with anecdotal evidence from the Bayesian analysis. Literacy with the Web was significantly higher than that of participants with respect to LLM tools, with very strong Bayesian evidence. Other differences failed to reach significance. Considering the results of the Bayesian analyses, these non-significant findings can be divided into two groups. Accuracy in the essay questions, the frequency of questions answered with a mixed strategy, the NASA-TLX performance subscale, and prior knowledge showed moderate evidence for the absence of a group difference. In contrast, essay and memory questions response times, accuracy in the memory questions, questions answered exclusively with a copy-paste and self-elaborated strategies, and the NASA-TLX temporal and effort subscales showed only anecdotal evidence for the absence of a group difference.

Exploratory Analyses: Condition-Specific Analyses

Web Condition. Multiple significant Spearman correlations (r_S) were found between memory performance and other measures, including with RT of the essay questions ($r_S = .40$, $p = .034$), the % of copy-paste answers on the essay questions ($r_S = -.64$, $p < .001$), the % of self-elaborated answers produced on the essay questions ($r_S = .51$, $p = .007$), the number of websites consulted ($r_S = .68$, $p < .001$), the mean prior knowledge of the subjects addressed ($r_S = .39$, $p = .037$), RT on the memory questions ($r_S = -.38$, $p = .041$), and accuracy on the essay questions ($r_S = .55$, $p = .002$).

A three-step hierarchical regression was run to predict performance on the memory questions (see Table 3). The first step,

Table 2. Depiction of the mean values (and SD) for the different variables common to both conditions (Web vs. LLM conditions) and results of the Mann-Whitney tests

	Condition					
Variable	Web	LLM	*W*	r_{rb}	BF_{10}	*BF* qualitative
Performance						
Essay – RT (s)	241.6 (95.9)	208.1 (108.1)	488.5	−.25	0.65	Anecdot. H_0
Essay – acc. (%)	53.0 (9.9)	54.1 (7.7)	420.0	.07	0.27	Moderate H_0
Memory – RT (s)	17.5 (7.8)	19.5 (7.2)	312.0	.20	0.61	Anecdot. H_0
Memory – acc. (%)	50.6 (0.2)	43.3 (0.2)	451.5	−.15	0.41	Anecdot. H_0
Research/LLM interactions						
N queries	1.8 (0.5)	1.4 (0.4)	577.0**	−.47	10.95	Strong H_1
Copy-paste questions (%)	13.9 (24.9)	33.2 (25.5)	171.5***	.56	16.42	Strong H_1
N websites consulted	1.5 (0.6)	0.6 (0.7)	666.0***	−.70	263.15	Extreme H_1
Essay answer strategies						
Exclusive copy-paste answers (%)	49.4 (35.0)	58.2 (36.8)	327.0	.17	0.53	Anecdot. H_0
Mixed answers (%)	27.8 (24.7)	31.3 (31.0)	385.0	.02	0.28	Moderate H_0
Self-elaborated answers (%)	22.9 (35.8)	10.5 (22.1)	470.5	−.20	0.49	Anecdot. H_0
Effort and workload						
Efforts deployed (/5)	2.8 (0.8)	2.2 (0.8)	525.5*	−.39	3.46	Moderate H_1
NASA-TLX – Mental (/10)	5.7 (2.3)	4.4 (1.8)	513.5*	−.31	1.60	Anecdot. H_1
NASA-TLX – Temporal (/10)	5.4 (2.6)	4.6 (2.4)	464.0	−.19	0.56	Anecdot. H_0
NASA-TLX – Effort (/10)	4.2 (1.8)	3.5 (1.9)	477.0	−.22	0.60	Anecdot. H_0
NASA-TLX – Performance (/10)	6.2 (1.8)	6.1 (1.8)	407.0	−.04	0.30	Moderate H_0
Knowledge and literacy						
Prior knowledge (/10)	2.2 (1.1)	2.5 (1.4)	362.0	.08	0.32	Moderate H_0
Tool literacy (%)	80.9 (11.5)	67.3 (12.1)	627.0***	−.60	51.86	V. strong H_1

Note. r_{rb} = Ranked biserial correlation (effect size measure). The *BF* qualitative column displays the interpretation of the Bayesian analysis based on the BF_{10} value, according to Jeffreys (1961) classification scheme, depending on whether the value provides anecdotal (anecdot.), moderate, strong, very strong (v. strong) or extreme evidence favoring the null (H_0) or the alternative (H_1) hypothesis.
*p < .05 **p < .01 ***p < .001.

focusing on basic self-regulation (tool literacy, prior knowledge and essay questions accuracy), significantly predicted memory performance, $F(3, 25) = 3.90$, $p = .020$, adjusted $R^2_{Wherry} = .24$, adjusted $R^2_{Stein} = .11$, with a marginal impact of accuracy on the essay questions ($p = .064$). The second step, involving the addition of advanced self-regulation metrics (keywords strategy depth, response strategy depth and number of websites visited), further increased the amount of variance predicted, $F(6, 22) = 9.01$, $p < .001$, adjusted $R^2_{Wherry} = .63$, adjusted $R^2_{Stein} = .51$. At this step, prior knowledge, tool literacy, response strategy depth, and the number of webpages visited significantly contributed to the model. The third and final step of the hierarchical regression model, including the mental workload metrics, marginally improved the prediction of memory performance, $F(10, 18) = 8.05$, $p < .001$, adjusted $R^2_{Wherry} = .72$, adjusted $R^2_{Stein} = .54$. At this step, prior knowledge, tool literacy, response strategy depth, the number of webpages visited, NASA-TLX temporal pressure and NASA-TLX effort significantly impacted the prediction of memory performance. A Bayesian-equivalent regression analysis comprised of the 10 predictors included in the full hierarchical regression model produced a $BF_M = 2.86$ ($p_{(M|\text{data})} = .223$). This model was 459.6 times more likely than the null model. Anecdotal evidence favoring the inclusion of the variables

Table 3. Summary of the hierarchical linear regression performed to predict memory performance for the Web condition

Variable	*b*	95% CI		*SE*	β	R^2	ΔR^2
		Lower	Upper				
Step 1: Basic self-regulation (F = 3.90, p = .020)						.32	.32*
Constant	−0.275	−0.838	0.287	0.039	-		
Tool literacy	0.003	−0.004	0.010	0.003	.183		
Prior knowledge	0.045	−0.026	0.117	0.035	.233		
Essay – acc.	0.008†	-4.86×10^{-6}	0.016	0.004	.370		
Step 2: Basic self-regulation + advanced self-regulation (F = 9.01, p < .001)						.71	.39***
Constant	−0.535*	−0.979	−0.092	0.214	-		
Tool literacy	0.005*	2.04×10^{-4}	0.010	0.002	.283		
Prior knowledge	0.051*	8.83×10^{-4}	0.101	0.024	.263		
Essay – acc.	-5.55×10^{-4}	−0.007	0.006	.003	−.026		
Keywords depth	4.17×10^{-4}	-3.44×10^{-4}	0.001	3.67×10^{-4}	.141		
Response depth	0.001*	1.03×10^{-4}	0.002	4.43×10^{-4}	.323		
N websites consulted	0.164**	0.062	0.266	0.049	.466		
Step 3: Basic self-regulation + advanced self-regulation + workload (F = 8.05, p < .001)						.82	.11†
Constant	−0.614*	−1.176	−0.052	0.267	-		
Tool literacy	0.006*	9.65×10^{-4}	0.011	0.002	.340		
Prior knowledge	0.050*	0.005	0.096	0.022	.259		
Essay – acc.	−0.002	−0.008	0.004	0.003	−.095		
Keywords depth	6.87×10^{-4}	-1.10×10^{-4}	0.001	3.79×10^{-4}	.232		
Response depth	0.001**	4.80×10^{-4}	0.002	4.16×10^{-4}	.428		
N websites consulted	0.180***	0.088	0.272	0.044	.511		
NASA-TLX – Mental	−0.016	−0.044	0.012	0.013	−.175		
NASA-TLX – Temporal	−0.027*	−0.052	−0.001	0.012	−.330		
NASA-TLX – Effort	0.054**	0.015	0.092	0.018	.455		
NASA-TLX – Performance	−0.013	−0.045	0.019	0.015	−.109		

†p < .07 *p < .05 **p < .01 ***p < .001.

in the model was found for the keywords strategy depth ($BF_{\text{inclusion}}$ = 1.83), accuracy on the essay questions ($BF_{\text{inclusion}}$ = 1.22), NASA-TLX mental workload ($BF_{\text{inclusion}}$ = 1.40), NASA-TLX temporal pressure ($BF_{\text{inclusion}}$ = 2.72), NASA-TLX effort ($BF_{\text{inclusion}}$ = 2.67) and NASA-TLX performance estimation ($BF_{\text{inclusion}}$ = 1.13). Moderate evidence was found for the response depth ($BF_{\text{inclusion}}$ = 7.48), prior knowledge ($BF_{\text{inclusion}}$ = 3.44) and tool literacy ($BF_{\text{inclusion}}$ = 3.18), while very strong evidence was found for the inclusion of the number of websites visited ($BF_{\text{inclusion}}$ = 32.08).

LLM Condition. Seventeen (vs. 10) participants verified at least once the information, for a mean of 4.2 questions out of 12 (SD = 4.8), with an average of 0.9 prompt by participant (SD = 1.0). In addition to the 33.2% of prompts that copied and pasted the question shown in the study, the remaining 66.8% (SD = 25.7) used a zero-shot prompting strategy and no context-setting prompting was used. Correlation analyses were also conducted within the LLM condition. The following variables were related to memory performance: the % of copy-paste answers on the essay questions (r_S = −.45, p = .018), the number of verification requests (r_S = .43, p = .025), and the number of websites consulted (r_S = .45, p = .018). Contrary to the Web condition, memory performance was not

related to accuracy on the essay questions (r_S = .07, p = .717) and to prior knowledge (r_S = .35, p = .073). Mean memory performance for the participants who verified at least once the information (n = 17, M = 51.2%, SD = 0.05) was significantly higher than the performance of those who did not (n = 10, M = 29.8%, SD = 0.21), W = 39.5, p = .024, r_{rb} = −.54. The Bayesian-equivalent test provided anecdotal evidence favoring H_1 (BF_{10} = 1.82).

A three-step hierarchical regression was carried out to predict accuracy on the memory questions (see Table 4). The first step focusing on basic self-regulation failed to significantly predict memory performance, $F(3, 23) = 1.45$, p = .255, adjusted R^2_{Wherry} = .05, adjusted R^2_{Stein} = −.12, with the effect of prior knowledge being marginally significant (p = .061). The addition of advanced self-regulation metrics significantly improved the model and yielded a significant prediction of memory performance, $F(6, 20) = 4.50$, p = .005, adjusted R^2_{Wherry} = .44, adjusted R^2_{Stein} = .24. Here, prior knowledge, response strategy depth and the number of websites consulted significantly contributed to the model. The addition of workload metrics at the final step also led to the significant prediction of memory performance, $F(10, 16) = 3.12$, p = .021, adjusted R^2_{Wherry} = .45, adjusted R^2_{Stein} = .05, but

Table 4. Summary of the hierarchical linear regression performed to predict memory performance for the LLM condition

		95% CI					
Variable	*b*	Lower	Upper	*SE*	β	R^2	ΔR^2
Step 1: Basic self-regulation (F = 1.45, p = .255)						.16	.16
Constant	0.238	−0.520	0.996	0.366	-		
Tool literacy	−0.003	−0.011	0.006	0.004	−.131		
Prior knowledge	0.069†	−0.003	0.142	0.035	.411		
Essay – acc.	0.004	−0.010	0.017	0.006	.112		
Step 2: Basic self-regulation + advanced self-regulation (F = 4.50, p = .005)						.57	.42**
Constant	−0.491	−1.199	0.218	0.340	-		
Tool literacy	-9.73×10^{-4}	−0.008	0.006	0.003	−.050		
Prior knowledge	0.070*	0.011	0.128	0.028	.416		
Essay – acc.	0.002	−0.010	0.015	0.006	.070		
Prompts depth	0.002	−0.001	0.006	0.002	.240		
Response depth	0.002*	1.07×10^{-4}	0.003	7.36×10^{-4}	.363		
N websites consulted	0.130*	0.013	0.247	0.056	.383		
Step 3: Basic self-regulation + advanced self-regulation + workload (F = 3.12, p = .021)						.66	.09
Constant	−0.621	−1.402	0.159	0.368	-		
Tool literacy	−0.002	−0.010	0.006	.004	−.095		
Prior knowledge	0.058	−0.019	0.135	0.036	.345		
Essay – acc.	−0.002	−0.078	0.014	0.007	−.055		
Prompts depth	0.003	-8.07×10^{-4}	0.007	0.002	.322		
Response depth	0.002*	3.61×10^{-4}	0.004	8.62×10^{-4}	.483		
N websites consulted	0.135*	0.004	0.266	0.062	.396		
NASA-TLX – Mental	−0.043	−0.101	0.015	0.027	−.324		
NASA-TLX – Temporal	0.014	−0.055	0.082	0.032	.142		
NASA-TLX – Effort	0.013	−0.073	0.099	0.041	.104		
NASA-TLX – Performance	0.050	−0.022	0.122	0.034	.378		

†p < .07 *p < .05 **p < .01 ***p < .001.

failed to improve prediction. Only the response strategy depth and the number of websites consulted significantly impacted the prediction of memory performance. A Bayesian-equivalent regression analysis comprised of the 10 predictors included in the full hierarchical regression model produced a $BF_M = 1.11$ ($p_{(M|\text{data})} = .100$). This model was 3.068 times more likely than the null model. Anecdotal evidence favoring the inclusion of the variables in the model ($BFs_{\text{inclusion}} < 2.28$) was found for almost all the predictors, excepting for response depth ($BF_{\text{inclusion}} = 4.13$) and the number of websites visited ($BF_{\text{inclusion}} = 3.05$), which both yielded moderate evidence favoring their inclusion.

Discussion

The goal of this study was to examine how learning, workload and search behaviors were impacted by an LLM agent during a self-regulated computer-based learning task, as opposed to a more classic Web search strategy. Participants answered essay questions while either looking for information on the Web or by using an LLM agent (with the possibility of verifying the answers provided by the LLM on the Web), followed by a surprise memory test. Participants in the Web condition reported significantly higher mental workload and effort deployed compared to those in the LLM condition. Additionally, participants in the LLM condition submitted fewer queries to the agent than those in the Web condition did to the search engine, were more likely to copy and paste the full question rather than reformulate it into keywords, and consulted fewer websites. This reduction is consistent with the availability of an additional source of information in the LLM condition, where consulting external websites primarily served a verification role, which was expected but not required to complete the task. Tool literacy was lower in the LLM condition as opposed to that of the Web condition. Despite these differences, performance on the surprise memory test did not differ significantly across conditions. Exploratory correlations and hierarchical regression analyses highlighted multiple relationships across the different variables, with the possibility to significantly predict memory performance in both conditions with slightly different patterns. In both conditions, the extent to which participants were involved in the essay question answer (i.e., response strategy depth) as well as the number of websites consulted during the research contributed to improving accuracy on the memory questions. Yet, for the Web condition only, significant contributions were also found for the prior knowledge of the subjects addressed (positive relationship), the tool literacy (positive relationship), measures of perceived temporal pressure (negative relationship) and measure of perceived efforts (positive relationship).

These results support the idea that chatbots may reduce the mental workload and effort required to find information on the Internet for self-regulated learning. This is consistent with previous work conducted on AI and learning (Bai et al., 2023; Jose et al., 2025; Kosmyna et al., 2025; Rice et al., 2024; Wu & Yu, 2024). By automating the steps of finding information, the extraneous (irrelevant) cognitive load required to navigate through the Web is reduced, which allows further cognitive resources to be deployed on the main task (e.g., for reading, managing and processing information; Wang et al., 2019). Contrary to our hypothesis, the lower mental demands found for the LLM condition did not induce poorer learning or essay performance. Interpreted against Jeffreys' (1961) Bayesian approach, the essay accuracy measure provided moderate evidence favoring the null hypothesis, whereas other performance measures only reflected anecdotal evidence (i.e., BF_{10} between 0.33 and 1). The apparent absence of difference on memory performance should thus be taken with caution and do not, on their own, warrant strong evidence of equivalence across the two conditions. Nonetheless, these results contrast with Ju (2023) and Kosmyna et al. (2025), which showed poorer learning for chatbot users. This inconsistency may be due to the possibility for

participants in the LLM condition to verify the information on the Internet. Further analyses showed that the participants who verified at least once the information for the essay questions obtained a higher memory performance than those who did not. This supports the idea that Web verification might play a role in how participants learned through the essay task with the LLM agent. Correlation analyses for the LLM condition also showed that both the number of verification requests produced, and the number of websites consulted, were positively and significantly related to memory performance.

Exploratory correlation and regression analyses outlined how effort, mental workload and some search micro-action behaviors were particularly related to memory performance. For example, the response strategy depth measure for the essay questions was positively related to memory performance in both conditions. A larger number of websites consulted was also related to higher memory performance. These findings align with previous evidence that micro-actions performed during web searches and/or chatbot interactions are related to self-regulated learning processes (Chiu et al., 2013; Colville & Ostern, 2026; Lai et al., 2025; Lyu & Ding, 2025; Maldonado-Mahauad et al., 2018; Sun et al., 2023; Zhou, 2013). They are also consistent with the principles of desirable difficulties (Bjork & Bjork, 2020) and the idea that effort and engagement may contribute to optimal learning outcomes. This engagement required more time and may have increased elaboration of the information and improved the memory trace. In fact, a significant relationship was found with the essay questions RT for many search actions in the Web condition (frequency of copy-paste answer strategy: $r_S = -.50$, $p = .006$; effort deployed: $r_S = .43$, $p = .022$; NASA-TLX mental workload: $r_S = .51$, $p = .005$; NASA-TLX effort: $r_S = .42$, $p = .023$) and in the LLM condition (frequency of copy-paste answer strategy: $r_S = -.66$, $p < .001$; verification frequency: $r_S = .54$, $p = .003$; effort deployed: $r_S = .50$, $p = .007$; NASA-TLX mental workload: $r_S = .43$, $p = .025$). This additional effort might have increased the elaboration of information, strengthening the memory trace. Information verification and increased engagement during the essay questions may facilitate the activation of pre-existing networks of semantic associations, a mechanism widely associated with improved memory representations (e.g., Craik, 2002; Lockhart & Craik, 1990). However, the actual number of verifications for the LLM condition was low ($M = 4.19$, $SD = 4.77$, across the 12 essay questions). Such verification, which requires effort, may have been insufficient to translate into significant cognitive effort for the LLM condition, explaining why no NASA-TLX measure contributed to predict memory performance in this group. On the opposite, systematic Web search in the Web condition allowed proper self-regulated learning processes to activate, from the efforts involved in searching for relevant information to producing an appropriate response. Overall, the fact that memory performance in both conditions could be predicted from research micro-actions but that some learning components contributed to memory performance only for the Web condition suggests that learning may not be solely influenced by the tool, but rather by how one interacts with it.

Practical Implications

The results of this study have practical implications for learning and educational contexts, particularly in settings where learners rely on self-regulated learning, such as higher education, professional development, and informal knowledge and skill acquisition. The lower mental demand reported in the LLM condition suggests that chatbots may indeed help to reduce extraneous load (Wang et al., 2019) potentially easing the cognitive burden experienced in these learning contexts. However, effective use requires learners to verify the content generated by the chatbot and to show superior engagement in the task, which can enhance content accuracy and retention. This is also the case for classic Web search methods; however, these methods

typically demand more user-driven effort to locate, verify, and integrate relevant information. On the contrary, gathering information from an LLM agent does not necessarily require high effort or engagement. To ensure optimal learning and proper use of these tools, more engaging integrations should be envisioned. Methods inspired by the ICAP framework (Interactive, Constructive, Active, and Passive; Chi & Wylie, 2014), which proposes different levels of cognitive engagement activities, could represent a great way to adapt how LLM-driven agents are being used to support short-term self-regulated learning. In line with this idea, Romero (2025) provides a similar framework, focused specifically on a six-level engagement model for AI tool interactions for educational purposes (see also Kohnke, 2023, for a list of engaging microlearning activities involving chatbots).

Limitations

The results of the present study should be interpreted considering a few limitations. As stated earlier, the correlation and regression analyses were exploratory. The regression analyses were performed using a set of predictors which was too large for the number of participants in each condition. Furthermore, no p-value correction was performed despite the large number of tests carried out. This situation in which multiple correlations are conducted to each explore independent questions, as opposed to jointly responding to a single omnibus null hypothesis, correspond to an "individual testing" context in which adjustments for multiple testing is not required (cf. Garcia-Perez, 2023). Still, this exploratory purpose, in addition to the relatively low sample size of each condition, means that our results may not necessarily generalize to new data. This was particularly evident with the LLM condition where the adjusted R^2_{Stein} values were low and even, in one case, negative, suggesting that the model would likely fail to outperform a simple mean-based prediction when applied to a new independent sample. As such, these exploratory analyses should be considered with caution. The goal of the present study was to explore potential mechanisms of explanation for memory performance. Because exploratory analyses carry a greater risk of false-positive findings (Simmons et al., 2011) we interpreted these results cautiously and clearly identified their exploratory nature, in line with reporting standards (APA, 2024). The present findings should therefore be viewed as preliminary evidence, providing a basis for future studies designed to test the robustness of these relationships in larger and in samples with different characteristics more representative of the general population.

Another limitation to consider is the possibility for participants in the LLM condition to verify the information. Indeed, this reduces the internal validity of the manipulations and the between-condition comparisons. However, this was intentionally allowed to ensure sufficient ecological validity. Furthermore, it also offers the opportunity to evaluate potential interactions between both tools which, for scientific and real-life applications, might be more useful. Future studies should consider the addition of an LLM-only condition and an LLM-assisted condition in which systematic Web verification is mandatory to truly isolate the effect of chatbots and information verification on learning.

Finally, different choices could have been made regarding the measures and questionnaires used, including, for example: (a) the tool literacy questionnaires (which varied between conditions, thus reducing the validity of a direct comparison between them); (b) the measures focused on research and response strategy depth (which was computed from different micro-actions performed by participants); (c) the presentation of frequent effort-related items (which results differed from the general NASA-TLX effort subscale, potentially due to the problems inherent to the NASA-TLX questionnaire; Bolton et al., 2023; Felton et al., 2012; Galy et al., 2018; Taylor & Bolton, 2026); and (d) the actual essay and memory questions (which were used for the first time and not necessarily

validated). Many of these choices were driven by the intention to reduce as much as possible the length of the study while guaranteeing complementarity of the different measures. Although research on chatbot-assisted learning is burgeoning (e.g., Akgun & Toker, 2025; Ju, 2023; Kosmyna et al., 2025; Stadler et al., 2024), there is still substantial variability across study designs and measures. Therefore, the study was characterized by an important exploratory component, and it remains difficult to rely solely on validated/normed questionnaires and metrics. Further studies are thus required to improve our comprehension of the effects of chatbots and cognitive effort on learning and, in turn, to establish reliable strategies for testing this phenomenon.

Conclusion

In conclusion, the current study offers an exploration of the interaction of workload and search behavior during a short-term self-regulated learning task. The results supported the idea that chatbots may help reduce mental workload and perceived effort, and outlined how different actions with both a chatbot and/or the Web might influence the extent to which memory is impacted during a Web search task. This study helps refine our view of chatbot interactions for short-term learning and highlights how the nature of the interaction with the tool, rather than the tool itself, may affect the learning experience. Future studies could explore how the different actions performed while browsing the Web and/or interacting with an LLM agent relate to Cheng et al.'s (2013) and other similar models' self-regulated learning phases and assess their impact on learning.

Author Notes

Part of this work was previously published in the *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* [Marois, A., Lavallée, I., Boily, G., Ramon Alaman, J., Desrosiers, B., & Lavoie, N. (2025). Chatbot memory: Uncovering how mental effort and chatbot interactions affect short-term learning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *69*(1), 2120-2126. https://doi.org/10.1177/10711813251358242] by Sage Journals. The present paper extends the analyses and proposes the analysis of a different set of participants, as well as new results and different research objectives.

Acknowledgments

The authors thank all participants who took part in the study. We also want to thank Bérénice Desrosiers, Noémie Lavoie, Laurie Dionne, Olivier Morin, Noémie Trelles-Boucher, Laurent Berthod and Jeanne Nicole for their support.

ORCID iDs

Alexandre Marois https://orcid.org/0000-0002-4127-4134

Jonay Ramon Alaman https://orcid.org/0000-0002-8642-0422

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by a grant provided by the Fonds de soutien à la recherche of Université Laval and a Discovery grant from the Natural Sciences and Engineering Research Council of Canada (NSERC), awarded to Alexandre Marois. Isabelle Lavallée and Gabrielle Boily were both supported by an Undergraduate Student Research Award, provided by NSERC.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

All data generated or analyzed during this study are included in the supplementary information files.

References

Akgun, M., & Toker, S. (2025). Modeling changes in individuals' cognitive self-esteem with and without access to search tools. *ArXiv*. https://doi.org/10.48550/arXiv.2501.10517

Al Shloul, T., Mazhar, T., Abbas, Q., Iqbal, M., Ghadi, Y. Y., Shahzad, T., Mallek, F., & Hamam, H. (2024). Role of activity-based learning and ChatGPT on students' performance in education. *Computers and Education: Artificial Intelligence*, *6*, 100219. https://doi.org/10.1016/j.caeai.2024.100219

American Psychological Association [APA]. (2024). *APA style JARS – JARS-Quant*. American Psychological Association. https://apastyle.apa.org/jars/quant-table-1.pdf

Anderson, L., & Krathwohl, D. A. (2001). *Taxonomy for learning, teaching and assessing: A revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Bai, L., Liu, X., & Su, J. (2023). ChatGPT: The cognitive effects on learning and memory. *Brain-X*, *1*(3), Article e30. https://doi.org/10.1002/brx2.30

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

Bjork, R. A., & Bjork, E. L. (2020). Desirable difficulties in theory and practice. *Journal of Applied Research in Memory and Cognition*, *9*(4), 475–479. https://doi.org/10.1016/j.jarmac.2020.09.003

Bolton, M. L., Biltekoff, E., & Humphrey, L. (2023). The mathematical meaninglessness of the NASA task load index: A level of measurement analysis. *IEEE Transactions on Human-Machine Systems*, *53*(3), 590–599. https://doi.org/10.1109/THMS.2023.3263482

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *arXiv*. https://doi.org/10.48550/arXiv.2303.04226

Cheng, K.-H., Liang, J.-C., & Tsai, C.-C. (2013). University students' online academic help seeking: The role of self-regulation and information commitments. *The Internet and Higher Education*, *16*, 70–77. https://doi.org/10.1016/j.iheduc.2012.02.002

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chiu, Y. L., Liang, J. C., & Tsai, C. C. (2013). Internet-specific epistemic beliefs and self-regulated learning in online academic information searching. *Metacognition Learning*, *8*(3), 235–260. https://doi.org/10.1007/s11409-013-9103-x

Colville, S., & Ostern, N. (2026) (In press). Making informed trust judgements: The role of metacognition in ChatGPT interactions. *Behaviour & Information Technology*, 1–16. https://doi.org/10.1080/0144929X.2026.2651114

Craik, F. I. M. (2002). Levels of processing: Past, present. and future? *Memory*, *10*(5–6), 305–318. https://doi.org/10.1080/09658210244000135

de Bruin, A. B. H., Janssen, E. M., Waldeyer, J., & Stebner, F. (2025). Cognitive load and challenges in self-regulation: An introduction and reflection on the topical collection. *Educational Psychology Review*, *37*(3), 65. https://doi.org/10.1007/s10648-025-10042-2

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., & Baars, M. (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

Erdmann, A., Arilla, R., & Ponzoa, J. M. (2022). Search engine optimization: The long-term strategy of keyword choice. *Journal of Business Research*, *144*, 650–662. https://doi.org/10.1016/j.jbusres.2022.01.065

Felton, E. A., Williams, J. C., Vanderheiden, G. C., & Radwin, R. G. (2012). Mental workload during brain–computer interface training. *Ergonomics*, *55*(5), 526–537. https://doi.org/10.1080/00140139.2012.662526

Galy, E., Paxion, J., & Berthelon, C. (2018). Measuring mental workload with the NASA-TLX needs to examine each dimension rather than relying on the global score: An example with driving. *Ergonomics*, *61*(4), 517–527. https://doi.org/10.1080/00140139.2012.662526

Garcia-Perez, M. A. (2023). Use and misuse of corrections for multiple testing. *Methods in Psychology*, *8*, 100120. https://doi.org/10.1016/j.metip.2023.100120

Google Cloud. (2025). *What is prompt engineering*. Google Cloud. https://cloud.google.com/discover/what-is-prompt-engineering

Greene, J. A., Moos, D. C., & Azevedo, R. (2011). Self-regulation of learning with computer-based learning environments. *New Directions for Teaching and Learning*, *126*(2011), 107–115. https://doi.org/10.1002/tl.449

Gwizdka, J. (2010). Distribution of cognitive load in Web search. *Journal of the American Society for Information Science and Technology*, *61*(11), 2167–2187. https://doi.org/10.48550/arXiv.1005.1340

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hertzum, M. (2021). Reference values and subscale patterns for the task load index (TLX): A meta-analytic review. *Ergonomics*, *64*(7), 869–878. https://doi.org/10.1080/00140139.2021.1876927

IBM. (2025). *The 2025 guide to prompt engineering*. IBM. https://www.ibm.com/think/prompt-engineering#605511093

Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford University Press.

Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology*, *16*, 1550621. https://doi.org/10.3389/fpsyg.2025.1550621

Ju, Q. (2023). Experimental evidence on negative impact of generative AI on scientific learning outcomes. *SSRN. Research Square*. https://doi.org/10.2139/ssrn.4567696

Kizilcec, R. F., Pérez-Sanagustin, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behaviour and goal attainment in massive open online courses. *Computers & Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

Kohnke, L. (2023). Microlearning with chatbots. In *Using technology to design ESL/EFL microlearning activities. Springer Briefs in Education*. Springer. https://doi.org/10.1007/978-981-99-2774-6_7

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, *54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I., & Maes, P. (2025). Your brain on ChatGPT: Accumulation of cognitive debt when using an AI assistant for essay writing task. *arXiv*. https://doi.org/10.48550/arXiv.2506.08872

Lai, J. W., Qiu, W., Thway, M., Zhang, L., Jamil, N. B., Su, C. L., Ng, S. S. H., & Lim, F. S. (2025). Leveraging process-action epistemic network analysis to illuminate student self-regulated learning with a Socratic chatbot. *Journal of Learning Analytics*, *12*(1), 32–49. https://doi.org/10.18608/jla.2025.8549

Lee, S., & Park, G. (2024). Development and validation of ChatGPT literacy scale. *Current Psychology*, *43*(21), 18992–19004. https://doi.org/10.1007/s12144-024-05723-0

Lidstone, J., & Lucas, K. B. (1998). Teaching and learning research methodology from interactive multimedia programs: Postgraduate students' engagement with an innovative program. *Journal of Educational Multimedia & Hypermedia*, *7*(2–3), 237–261. https://eric.ed.gov/?id=EJ570655.

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, *44*(1), 87–112. https://doi.org/10.1037/h0084237

Lyu, Y., & Ding, R. (2025). Discovering self-regulated learning patterns in chatbot-powered education environment. *arXiv*. https://doi.org/10.48550/arXiv.2510.01275

Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, *80*, 179–196. https://doi.org/10.1016/j.chb.2017.11.011

Marois, A., Lavallée, I., Boily, G., Ramon Alaman, J., Desrosiers, B., & Lavoie, N. (2025). Chatbot memory: uncovering how mental

effort and chabot interactions affect short-term learning. *Proceedings of the Human Factors and Ergonomics Society*, *69*(1), 2120–2126. https://doi.org/10.1177/10711813251358242

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Meghanathan, N., Kostyuk, N., Isokpehi, R., & Cohly, H. (2010). An algorithm to self-extract secondary keywords and their combinations based on abstracts collected using primary keywords from online digital libraries. *International Journal of Computer Science and Information Technology*, *2*(3), 93–101. https://doi.org/10.5121/ijcsit.2010.2307

Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, *23*(3), 1126–1144. https://doi.org/10.1016/j.chb.2006.10.006

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43–97). Buros Institute of Mental Measurements. https://digitalcommons.unl.edu/burosmetacognition/3?utm_source=digitalcommons.unl.edu%2Fburosmetacognition%2F3&utm_medium=PDF&utm_campaign=PDFCoverPages

Rice, S., Crouse, S. R., Winter, S. R., & Rice, C. (2024). The advantages and limitations of using ChatGPT to enhance technological research. *Technology in Society*, *76*, 102426. https://doi.org/10.1016/j.techsoc.2023.102426

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Romero, M. (2025). From consumption to co-creation: A systematic review of six levels of AI-enhanced creative engagement in education. *Multimodal Technologies and Interaction*, *9*(10), 110. https://doi.org/10.3390/mti9100110

Schmidhuber, J., Schlögl, S., & Ploder, C. (2021). Cognitive load and productivity implications in human-chatbot interaction. In 2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS), Magdeburg, Germany, pp. 1–6. https://doi.org/10.1109/ICHMS53169.2021.9582445

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., Hamm, V., Vogt, A., & Riemer, V. (2024). The interplay of cognitive load, learners' resources and self-regulation. *Educational Psychology Review*, *36*(2), 50. https://doi.org/10.1007/s10648-024-09890-1

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

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

Stein, C. (1960). Multiple regression. In I. Olkin (Ed.), *Contributions to probability and statistics* (pp. 264–305). Stanford University Press.

Sun, J. C.-Y., Liu, Y., Lin, X., & Hu, X. (2023). Temporal learning analytics to explore traces of self-regulated learning behaviors and their associations with learning performance, cognitive load, and student engagement in an asynchronous online course. *Frontiers in Psychology*, *13*, 1096337. https://doi.org/10.3389/fpsyg.2022.1096337

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *The psychology of learning and motivation: Cognition in education* (pp. 37–76). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.

Taylor, S. S., & Bolton, M. L. (2026). Workload does not work: On the many problems of the NASA-TLX. *IEEE Transactions on Human-Machine Systems*, *56*(3), 682–687. https://doi.org/10.1109/THMS.2026.3664262

Van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2015). Development and validation of the Internet Skills Scale (ISS). *Information, Communication & Society*, *19*(6), 804–823. https://doi.org/10.1080/1369118X.2015.1078834

von Hoyer, J., Hoppe, A., Kammerer, Y., Otto, C., Pardi, G., Rokicki, M., Yu, R., Dietze, S., Ewerth, R., & Holtz, P. (2022). The search as learning spaceship: Toward a comprehensive model of psychological and technological facets of search as learning. *Frontiers in Psychology*, *13*, 827748. https://doi.org/10.3389/fpsyg.2022.827748

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

Wang, Z., Gong, S.-Y., Xu, S., & Hu, X. E. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Computers & Education*, *136*, 130–140. https://doi.org/10.1016/j.compedu.2019.04.003

Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *The Annals of Mathematical Statistics*, *2*(4), 440–457. https://doi.org/10.1214/aoms/1177732951

Willoughby, T., Anderson, S. A., Wood, E., Mueller, J., & Ross, C. (2009). Fast searching for information on the Internet to use in a learning context: The impact of domain knowledge. *Computers & Education*, *52*(3), 640–648. https://doi.org/10.1016/j.compedu.2008.11.009

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Taylor & Francis. https://psycnet.apa.org/record/1998-07283-011

Wirth, J., Stebner, F., Trypke, M., Schuster, C., & Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load. *Educational Psychology Review*, *32*(4), 1127–1149. https://doi.org/10.1007/s10648-020-09568-4

Wu, R., & Yu, Z. (2024). Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *British Journal of Educational Technology*, *55*(1), 10–33. https://doi.org/10.1111/bjet.13334

Zhou, M. (2013). Using traces to investigate self-regulatory activities: A study of self-regulation and achievement goal profiles in the context of web search for academic tasks. *Journal of Cognitive Education and Psychology*, *12*(3), 287–305. https://doi.org/10.1891/1945-8959.12.3.287

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Author Biographies

Alexandre Marois is an assistant professor at the École de psychologie of Université Laval, Québec (Canada) and a visiting fellow at the School of Psychology and Humanities of University of Lancashire. He obtained his PhD in experimental psychology from Université Laval in 2019.

Isabelle Lavallée is a PhD student at the Département de psychologie of Université de Sherbrooke, Sherbrooke (Canada). She obtained her BA in psychology from Université Laval in 2023.

Gabrielle Boily is a master's student at the École de psychologie of Université Laval, Québec (Canada). She obtained her BA in psychology from Université Laval in 2024 and an LL. B. from Université Laval in 2025.

Jonay Ramon Alaman is a PhD Student at the at the École de psychologie of Université Laval, Québec (Canada). He received his Master's degree in Aerospace Engineering (2020) from ISAE – SUPAERO (Toulouse, France).

Appendix

Table A1. Questions (Original French and English translation) Used for the Essay Questions and the Memory Test.

Question		Essay question	Memory question
1	French	*Quels sont les trois types de roches et comment se forment-elles?*	*Quel est le type de roches formées à partir de lave ou de magma solidifié?*
	English	What are the three types of rocks and how are they formed?	What types of rocks are formed from lava or solidified magma?
2	French	*Quelles sont les étapes d'une éruption volcanique?*	*Nommez une étape de l'éruption volcanique qui précède l'étape de l'éruption.*
	English	What are the stages of a volcanic eruption?	Name a stage of volcanic eruption that precedes the eruption stage.
3	French	*Expliquez le principe de la photosynthèse et comment les plantes utilisent la lumière pour produire de l'énergie.*	*Dans le processus de photosynthèse, en quelle substance l'énergie lumineuse est-elle transformée?*
	English	Explain the principle of photosynthesis and how plants use light to produce energy.	In the process of photosynthesis, what substance is light energy converted into?
4	French	*Expliquez le phénomène d'apparition d'un mirage supérieur et d'un mirage inférieur.*	*Quel type de mirage peut faire en sorte que des objets situés à l'horizon paraissent inversés?*
	English	Explain the phenomenon of upper mirages and lower mirages.	What type of mirage can cause objects on the horizon to appear inverted?
5	French	*Comment se déroule le processus de transformation du sable en verre plat?*	*Lors de la transformation du sable, comment se nomme le procédé durant lequel le verre est étalé sur une surface liquide pour obtenir une épaisseur uniforme?*
	English	How does the process of transforming sand into flat glass work?	When processing sand, what is the name of the process in which glass is spread over a liquid surface to obtain a uniform thickness?
6	French	*Quelles sont les principales caractéristiques des étoiles de type O et quelles conséquences ont-elles sur leur cycle de vie?*	*Les étoiles de type O émettent une grande quantité de quel type de rayonnement?*
	English	What are the main characteristics of O-type stars, and how do they affect their life cycle?	What type of radiation does an O-type star emit?

(continued)

Table A1. (continued)

Question		Essay question	Memory question
7	French	*Quels sont les éléments traditionnels caractéristiques d'un hanbok coréen et comment ces éléments varient-ils en fonction du genre?*	*Quel élément du hanbok est commun au costume féminin et au costume masculin?*
	English	What are the traditional elements characteristic of a Korean *hanbok*, and how do these elements vary according to gender?	Which element of the *hanbok* is common to both women's and men's clothing?
8	French	*Que raconte L'Avare de Molière?*	*Qui est Cléante par rapport à Harpagon, le personnage principal de la comédie L'Avare?*
	English	What is *L'Avare* by Molière about?	Who is Cléante in relation to Harpagon, the main character of the comedy *L'Avare*?
9	French	*Quelles sont les différences entre l'anthropologie culturelle et l'anthropologie physique?*	*Quel type d'anthropologie s'intéresse entre autres aux langues et aux structures sociales?*
	English	What are the differences between cultural anthropology and physical anthropology?	What type of anthropology focuses on languages and social structures, among other things?
10	French	*Quels sont les principaux secteurs économiques de l'Islande en dehors de l'industrie de la pêche, et comment contribuent-ils à l'économie du pays?*	*En Islande, la production de quel matériau constitue une activité économique importante?*
	English	What are Iceland's main economic sectors outside of the fishing industry, and how do they contribute to the country's economy?	In Iceland, the production of which material is an important economic activity?
11	French	*Qu'est-ce qu'était l'apartheid en Afrique du Sud et que stipulait sa Pass Law?*	*Quelles étaient les conséquences infligées à un individu qui transgressait la Pass Law durant l'apartheid en Afrique du Sud?*
	English	What was apartheid in South Africa and what did its *Pass Law* stipulate?	What were the consequences for an individual who violated the *Pass Law* during apartheid in South Africa?
12	French	*Qu'est-ce que Motown Records et nommez 5 artistes interprètes qui y sont associés.*	*Nommez un artiste interprète phare du Motown.*
	English	What are Motown Records, and name five artists associated with it.	Name a leading Motown recording artist.
13	French	*Quel est le cycle de vie d'une étoile?*	*Comment se nomme le type de nuage gigantesque de gaz et de poussières dans lequel les étoiles naissent?*
	English	What is the life cycle of a star?	What is the name of the gigantic cloud of gas and dust in which stars are born?
14	French	*Qu'est-ce que l'Expo 67 et comment les structures construites pour cette dernière sont actuellement utilisées?*	*Après l'Expo 67, en quoi a été rénové le pavillon de la France?*
	English	What was Expo 67 and how are the structures built for it currently being used?	After Expo 67, what was the France Pavilion renovated into?

(continued)

Table A1. (continued)

Question		Essay question	Memory question
15	French	*Quels sont les principaux enjeux environnementaux pour les tortues de mer?*	*Nommez une activité humaine ayant un impact écologique négatif pour les tortues de mer.*
	English	What are the main environmental challenges affecting sea turtles?	Name a human activity that has a negative ecological impact on sea turtles.
16	French	*Comment les zostères peuvent-elles être utilisées dans le cadre de la restauration écologique?*	*Complétez l'information suivante : Les zostères permettent la réduction de _____ dans les zones côtières.*
	English	How can eelgrass be used in ecological restoration?	Complete the following information: Eelgrass helps reduce _____ in coastal areas.
17	French	*Quelles sont les caractéristiques de la végétation poussant dans le Bouclier canadien?*	*Nommez une essence de feuillus spécifique que l'on retrouve dans les forêts boréales du Bouclier canadien.*
	English	What are the characteristics of the vegetation growing in the Canadian Shield?	Name a specific hardwood species found in the boreal forests of the Canadian Shield.
18	French	*Quelles sont les caractéristiques distinctives entre les deux espèces de rhinocéros africains, soit entre le rhinocéros blanc (Ceratotherium simum) et le rhinocéros noir (Diceros bicornis)?*	*Dans quelle zone de l'Afrique peut-on retrouver le rhinocéros blanc?*
	English	What are the distinguishing characteristics between the two species of African rhinoceros, namely the white rhinoceros (Ceratotherium simum) and the black rhinoceros (Diceros bicornis)?	In which part of Africa can white rhinos be found?
19	French	*Qu'est-ce que le mycélium et quelles sont ses fonctions chez les plantes?*	*Nommez une fonction essentielle du mycélium.*
	English	What is mycelium and what are its functions in plants?	Name one essential function of mycelium.
20	French	*Quelles sont les différences entre le sapin et l'épinette?*	*Nommez une caractéristique spécifique des cônes produits par les épinettes par rapport aux cônes produits par les sapins.*
	English	What are the differences between fir and spruce trees?	Name one specific characteristic of cones produced by spruce trees compared to cones produced by fir trees.
21	French	*Comment apparaissent les aurores polaires et pourquoi leur coloration varie-t-elle?*	*Quel est le phénomène astronomique provenant de l'espace à l'origine d'une aurore polaire?*
	English	How do polar auroras occur, and why does their coloration vary?	What astronomical phenomenon originating in space causes a polar aurora?
22	French	*Décrivez ce qui caractérise le genre littéraire du haïku.*	*Combien de syllabes contient le 2e vers d'un haïku?*
	English	Describe what characterizes the haiku as a literary genre.	How many syllables does the second line of a haiku contain?

(continued)

Table A1. (continued)

Question		Essay question	Memory question
23	French	*Comment le personnage principal du livre Le Comte de Monte-Cristo obtient-il ses richesses, et qu'accomplira-t-il avec sa fortune?*	*De qui Edmond Dantès reçoit-il sa fortune lui permettant de devenir le Comte de Monte-Cristo?*
	English	How does the main character of the book *Le Comte de Monte-Cristo* acquire his riches, and what will he accomplish with his fortune?	From whom does Edmond Dantès receive the fortune that allows him to become the Count of Monte Cristo?
24	French	*Qui est Oscar Wilde et quels ont été les événements décisifs l'ayant poussé à s'exiler de l'Angleterre?*	*Dans quel pays s'est exilé Oscar Wilde?*
	English	Who was Oscar Wilde, and what were the decisive events that drove him into exile from England?	In which country did Oscar Wilde live in exile?
25	French	*Décrivez la scène dépeinte dans L'Incrédulité de saint Thomas du Caravage.*	*Qu'est-ce que le doigt de saint Thomas touche dans la peinture L'Incrédulité de saint Thomas du Caravage?*
	English	Describe the scene depicted in Caravaggio's *The Incredulity of Saint Thomas*.	What does Saint Thomas's finger touch in Caravaggio's painting *The Incredulity of Saint Thomas*?
26	French	*Qu'est-ce que la Pax Mongolica et quel a été son impact sur le commerce entre les différentes sociétés de l'Eurasie?*	*Quelle route a été revitalisée lors de la Pax Mongolica?*
	English	What was the *Pax Mongolica*, and what was its impact on trade between the various societies of Eurasia?	Which route was revitalized during the *Pax Mongolica*?
27	French	*Qui est Simone Veil et quel a été son apport en tant que ministre de la Santé?*	*Que dépénalise la Loi Veil, préparée par Simone Veil en tant que ministre de la Santé?*
	English	Who was Simone Veil, and what was her contribution as Minister of Health?	What did the *Loi Veil*, drafted by Simone Veil as Minister of Health, decriminalize?
28	French	*Quels pays s'opposaient dans la Guerre de Sept Ans et quelles ont été les répercussions de la fin de cette guerre sur l'Empire français en Amérique et en Europe?*	*Qui est sorti victorieux de la Guerre de Sept Ans?*
	English	Which countries fought against each other in the Seven Years' War, and what were the repercussions of the end of this war for the French Empire in America and in Europe?	Who emerged victorious from the Seven Years' War?
29	French	*Qu'est-ce qu'El Niño et quel est son impact sur les côtes du Pacifique?*	*Dans quelle partie de l'océan Pacifique la température de l'eau est-elle anormalement élevée lors du phénomène El Niño?*
	English	What is El Niño, and what is its impact on the Pacific coasts?	In which part of the Pacific Ocean is the water temperature abnormally high during the El Niño phenomenon?

(continued)

Table A1. (continued)

Question		Essay question	Memory question
30	French	*Résumez l'histoire légendaire de Spartacus le gladiateur.*	*Qui, finalement, vaincu Spartacus et écrasa sa rébellion d'esclaves?*
	English	Summarize the legendary story of Spartacus the gladiator.	Who ultimately defeated Spartacus and crushed his slave rebellion?
31	French	*Décrivez l'histoire racontée dans le film Bienvenue à Gattaca de Andrew Niccol.*	*Quel est le prénom du personnage principal du film Bienvenue à Gattaca?*
	English	Describe the story told in Andrew Niccol's film *Gattaca*.	What is the first name of the main character in the film *Gattaca*?
32	French	*Quelles sont, selon l'UNESCO, les valeurs universelles exceptionnelles qui font du Vieux-Québec un site protégé?*	*Nommez un élément du Vieux-Québec qui lui donne sa valeur exceptionnelle selon l'UNESCO.*
	English	According to UNESCO, what are the outstanding universal values that make Old Québec a protected site?	Name an element of Old Québec that gives it its outstanding value according to UNESCO.
33	French	*Qu'est-ce que le Conseil de sécurité de l'ONU, quels pays y siègent et quels sont ses pouvoirs?*	*Nommez un membre permanent du Conseil de sécurité de l'ONU, autre que les États-Unis.*
	English	What is the UN Security Council, which countries sit on it, and what are its powers?	Name a permanent member of the UN Security Council other than the United States.
34	French	*Expliquez comment se forme une tornade et quels sont les facteurs climatiques qui contribuent à leur formation?*	*Nommez un facteur climatique favorisant l'apparition de tornades.*
	English	Explain how a tornado forms and what climatic factors contribute to their formation?	Name a climatic factor that promotes the formation of tornadoes.
35	French	*Qui est Banksy et pourquoi est-il reconnu?*	*Nommez un thème des œuvres de Banksy.*
	English	Who is Banksy, and why is he well known?	Name a theme found in Banksy's works.
36	French	*Quels sont les types de nuages catégorisés selon leur hauteur et quelles conditions météorologiques peuvent-ils indiquer?*	*Comment se nomme la catégorie de nuages situés le plus en hauteur?*
	English	What are the types of clouds categorized according to their altitude, and what weather conditions can they indicate?	What is the name of the category of clouds located at the highest altitude?
37	French	*Quel est le plus grand désert froid du monde et quelles sont ses caractéristiques géographiques principales (faune et flore, climat, position géographique)?*	*Quel est le type de climat présent dans le désert de l'Antarctique?*
	English	What is the largest cold desert in the world, and what are its main geographical characteristics (fauna and flora, climate, geographical location)?	What type of climate is found in the Antarctic desert?

(continued)

Table A1. (continued)

Question		Essay question	Memory question
38	French	*Quelles sont les étapes de la formation d'une image sur la rétine?*	*Quel est l'organe qui contribue à focaliser les rayons lumineux dans l'œil [sur la rétine] et qui permet une bonne vision de près et de loin en se contractant ou dilatant?*
	English	What are the stages in the formation of an image on the retina?	Which organ helps to focus light rays in the eye [on the retina] and allows good near and distance vision by contracting or dilating?
39	French	*Qu'est-ce que la paréidolie visuelle et qu'est-ce qui la cause?*	*Nommez une cause de la paréidolie visuelle.*
	English	What is visual pareidolia, and what causes it?	Name a cause of visual pareidolia.

Note. The question list originally contained 40 questions, but one question was removed following pilot testing.