Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Availability and Effectiveness of Generative AI for Web-Based Investigative Learning

Yutaka WATANABE & Akihiro KASHIHARA
[a]*Graduate School of Informatics and Engineering,*
The University of Electro-Communications, Japan
*yutaka.watanabe@uec.ac.jp

Abstract: In Web-based learning, learners are expected to investigate information through gathering and navigating Web resources, and to construct knowledge. In our previous work, we designed a model of Web-based investigative learning, which represents a learning process as a cycle of phases. We also developed a cognitive tool called iLSB (interactive Learning Scenario Builder) for scaffolding learning process as modeled. At present, on the other hand, learners can explore information by chatting with generative AI, which generates texts in a dialogue format. However, it is unclear about the context where generative AI is available for Web-based investigative learning, and about the effective use of generative AI. In order to examine the availability and effectiveness of generative AI, we have had case studies including the comparison between the use of iLSB and the use of generative AI. The results suggest that generative AI prevents learners from investigating sufficiently compared with iLSB, and the Web-based investigative learning model contributes to the proper use of generative AI for investigative learning. Furthermore, it is suggested that the use of generative AI is effective in exploring background knowledge about questions investigated and in organizing information for knowledge construction.

Keywords: Web-based Investigative Learning, Generative AI, Information Exploration

1. Introduction

Investigative learning on the Web involves information exploration (Kashihara & Akiyama, 2017). On the Web, numerous resources for any question are available. Learners are expected to gather and navigate these resources according to their interests, and to construct knowledge related to an initial question (Hill & Hannafin, 1997; Henze & Nejdl, 2001). Learners could accordingly create their own learning scenario by deciding which resources and in what order they should investigate while navigating with knowledge construction. However, such learning scenario creation is difficult due to cognitive load (Zumbach & Mohraz, 2008). In order to support such self-directed activity, we have designed a model of Web-based investigative learning, which represents the process of investigating Web resources to construct knowledge (Kashihara & Akiyama, 2017; Kashihara, 2023). In our previous work, we have also developed a cognitive tool called iLSB, which stands for interactive Learning Scenario Builder. It provides learners with some scaffolds so that they can conduct investigative learning process as modeled. It is shown that iLSB allows learners to investigate wider and deeper information compared with a browser not having any scaffold (Kashihara & Akiyama, 2017).

On the other hand, learners can use generative AI to explore information at present. Generative AI generates texts with large language models (LLMs), which is generally trained using Web resources. It also has a text-based conversation with its users. In the context of investigative learning, learners can use generative AI to ask what they want to know regarding an initial question, and obtain information summarized for the question. They can also continue asking generative AI the sub-questions to explore further information.

However, there exists little work on in what kind of context generative AI is available and how effectively it could be used for Web-based investigative learning. In this work, we have had case studies whose purposes were to examine the availability and effectiveness of generative AI in investigative learning on the Web. The results of the case studies suggest that generative AI prevents learners from investigating sufficiently regarding an initial question compared with iLSB. It is also suggested that the model contributes to properly using generative AI for effective learning. Furthermore, it is suggested that the use of generative AI is effective when learners explore background knowledge about questions investigated to construct their knowledge.

2. Web-based Investigative Learning

2.1 Model of Web-based Investigative Learning

In order to support wider/deeper knowledge construction in Web-based investigative learning, we have designed a model of Web-based investigative learning (Kashihara & Akiyama, 2017; Kashihara, 2023) which instructs them how to create a learning scenario.

The model represents the learning process as a cycle of three phases, which are (1) searching and navigating Web resources, (2) knowledge construction, and (3) question expansion. In the first phase, learners are expected to search for Web resources regarding an initial question and navigate them. They are also expected to extract keywords that represent what they learned regarding the question. In the second phase, learners are expected to make the relationships among the segmented keywords to represent a keyword structure. This corresponds to constructing knowledge. In the third phase, learners are expected to expand the question into some related sub-questions to be investigated further, which are selected from the keywords segmented in the first phase. The selected keywords are called question keywords (q-keywords for short).

In this way, learners are expected to repeat these three phases for each sub-question until they consider the question expansion sufficient. The question expansion results in a tree called question tree as shown in Figure 1, whose root represents the initial question, and other nodes represent the sub-questions expanded. It represents the learning scenario, which implies the order of the questions learners investigated. It is viewed as an outcome of Web-based investigative learning, and it also shows how learners investigate information regarding the initial question widely and deeply. Since they could decide sub-questions to be expanded according to their needs and interest, the learning scenarios created are different from each other.

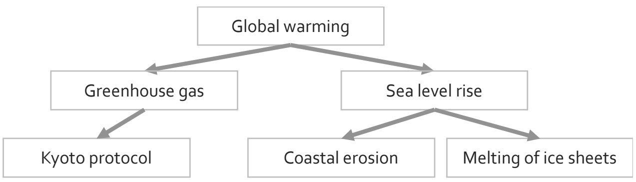

Figure 1. An Example of Learning Scenario Regarding “Global Warming”

2.2 interactive Learning Scenario Builder (iLSB)

In order to scaffold the Web-based investigative learning as modeled, we have developed iLSB (Kashihara & Akiyama, 2017; Kashihara, 2023). iLSB is an add-on system that works on Firefox, and it has three main functions to support the three phases in the model: (a) Web browser and search engine, (b) keyword repository, and (c) question tree viewer. Keyword repositories are prepared for each q-keyword. Learners can extract keywords via the Web browser to put into the keyword repository, which allows them to make their relationships for knowledge construction. The question tree viewer allows them to build up a learning scenario.

3. Generative AI for Learning

Generative AI is viewed as a tool to generate text with large language models (LLMs). As generative AI can produce sentences automatically in response to learners' questions, it is expected to be able to support several learning activities (Kasneci et al., 2023). However, generative AI would have negative effects on learners. In particular, generative AI sometimes presents inaccurate information (Lo, 2023). Moreover, if learners are too dependent on generative AI, they may not think the reliability by themselves, which prevents them from coming up with further questions (Kasneci et al., 2023).

Although generative AI could be used for Web-based instigative learning, the same concerns described above should be addressed. In order to discuss how generative AI could be used in Web-based investigative learning, let us consider functions of generative AI to be used for it. Related work has classified functions of ChatGPT for learning into three: question answering, information summarization, and collaboration facilitation (Lo, 2023). In Web-based investigative learning, learners are expected to investigate information for resolving questions. The question answering function is accordingly essential. In addition, they are expected to learn from the investigated information, which involves reorganizing the contents explored. The information summarization function would be effective for such reorganization.

There are several works that have examined the accuracy of outputs by generative AI. However, the availability of generative AI in actual learning has not been sufficiently addressed yet. Furthermore, it is unclear about the effective way to learn with generative AI involving functions such as question answering and information summarization. This paper aims to analyze how learners investigate information using generative AI, and to examine its availability and effectiveness in Web-based investigative learning. As for the availability, we analyze in what kind of context generative AI is available. As for the effectiveness, we analyze how learners investigate information when using iLSB with generative AI.

4. Case Study

4.1 Purpose

In order to examine the availability and effectiveness of generative AI used in Web-based investigative learning, we have had case studies to conduct three analyses as follows. The first and second analyses were done for evaluating the availability, and the third analysis was done for evaluating the effectiveness of generative AI in using iLSB.

(1) Analysis of investigation with generative AI before learning the model of Web-based investigative learning process compared to investigation with iLSB.
(2) Analysis of investigation with generative AI after learning the model compared to the one before learning the model.
(3) Analysis of investigation using iLSB with generative AI compared to investigation only with iLSB.

4.2 Preparation

For the case studies, we have developed a generative AI client, which allows learners to chat with generative AI. For this client, we used gpt3.5-turbo-1106, a LLM developed by OpenAI and optimized for chatting (OpenAI, 2024). In using iLSB with the generative AI client, learners are allowed to segment some keywords from its outputs to the keyword repository in iLSB.

4.3 Procedure

Participants were 26 graduate and undergraduate students in science and engineering. They were divided into three groups named iLSB-group, AI-group, and iLSB-AI-group. AI-group had 8 participants, and the other two groups had 9 participants for each group. In the case studies, the participants used Japanese to chat with generative AI.

First, all participants answered the questionnaire about experience of using generative AI. After that, the AI-group participants made investigation regarding "gender gap" using generative AI for 40 minutes (called AI-Pre-condition). However, they were allowed to finish their learning within the limited time when they thought they learned sufficiently. This was applied to all conditions. After the investigation, they answered questionnaire about the difficulty and sufficiency of learning.

Next, all participants watched an explanation video about the model of Web-based investigative learning, and made an investigation for 20 minutes for practice with iLSB. Then, the participants in the AI-group and iLSB-AI-group made an investigation for 20 minutes in order to practice learning with generative AI according to the model. The initial questions for these practices were different from "gender gap".

After that, the participants in each group investigated regarding "gender gap" for 40 minutes. The iLSB-group participants made an investigation with iLSB (called iLSB-condition), and the AI-group participants made an investigation only with generative AI (called AI-Post-condition). The iLSB-AI-group participants made an investigation with iLSB, and they were allowed to use generative AI whenever they wanted to use during investigation (called iLSB-AI-condition). They were allowed to extract keywords from generative AI's outputs into iLSB's keyword repository. After investigative learning, they answered questionnaire about the difficulty and sufficiency of their learning process. In addition, the iLSB-AI-group participants answered questionnaire about how they used generative AI.

4.4 Methods

In conducting the analyses (1) and (2), we evaluated the sufficiency and appropriateness of learning. As for the sufficiency, we measured the number of keywords as learning keyword (called l-keyword) which learners investigated to learn. For iLSB-group, segmented keywords were viewed as l-keywords. For AI-group, we viewed the following keywords as l-keywords, which were keywords the participants extracted from responses of generative AI to use in their messages, and keywords they came up with for representing the topics of the messages. The keyword extraction from the generative AI responses corresponds to keyword segmentation in the model. The l-keywords in both group is viewed as keywords which learners tried to incorporate into their knowledge. As for the appropriateness, we also measured the relevance between each l-keyword and the q-keyword representing the initial question. The relevance was assessed by three evaluators at three levels: strongly relevant, weakly relevant, or no relevance. They used reliable Web resources provided by organizations affiliated with Japan government or United Nations to assess. If two or more evaluators assessed the relevance as the same level, it resulted in the level. If all evaluators assessed it with different levels, it was evaluated as weakly relevant.

In conducting the analysis (3), we evaluated the sufficiency of learning by measuring the structure of question tree with six indicators: the number of sub-questions, the number of leaves, the maximum depth of leaves, the average depth of leaves, the degree of links from the root, and the minimum degree of question expansion (minimum QE). The minimum QE means to what extent learners investigate regarding an initial question, and QE is calculated for each node in the tree according to the following equations:

- $\mathrm{QE}(\mathrm{root}) = 1$
- $\mathrm{QE}(i) = \mathrm{QE}(\mathrm{parent}(i)) / m_{\mathrm{parent}(i)}$ (for $i \neq \mathrm{root}$)

where QE(i) is defined as QE of node i, parent(i) is defined as parent node of i, and $m_{parent(i)}$ is defined as the number of children nodes of parent(i). The smaller the minimum QE is, the more detailed the question tree is.

4.5 Hypotheses

We set up the following hypotheses H1 to H3 for each analysis.

H1: The use of generative AI without the model of Web-based investigative learning process results in insufficient learning compared to the use of iLSB.
H2: The use of generative AI with the model promotes investigating information more sufficiently.
H3: The use of iLSB with generative AI promotes proper knowledge construction and question expansion compared to the use of only iLSB.

4.6 Results

Figure 2 shows the total number of I-keywords and the number of I-keywords assessed at each relevance level in the analyses (1) and (2). In the analysis (1), we first conducted the tests for homogeneity of valiance for each indicator, and then conducted independent t-tests for homoscedastic indicators, and Welch's t-tests for heteroscedastic indicators. As a result, there was a significant difference in the total numbers of I-keywords between iLSB and AI-Pre conditions ($t(15)=5.54$, $p=0.000069$, Cohen's $d=2.69$). In other words, investigative learning in the AI-Pre-condition was less sufficient than the one in the iLSB-condition, which supports H1. In the analysis (2), we also conducted paired t-tests for each indicator. There was also a significant difference in the total numbers of I-keywords between AI-Pre and AI-Post conditions ($t(7)=2.67$, $p=0.032$, Cohen's $d=0.84$). This result suggests that learners in AI-group could investigate more information regarding the initial question due to the model, which support H2.

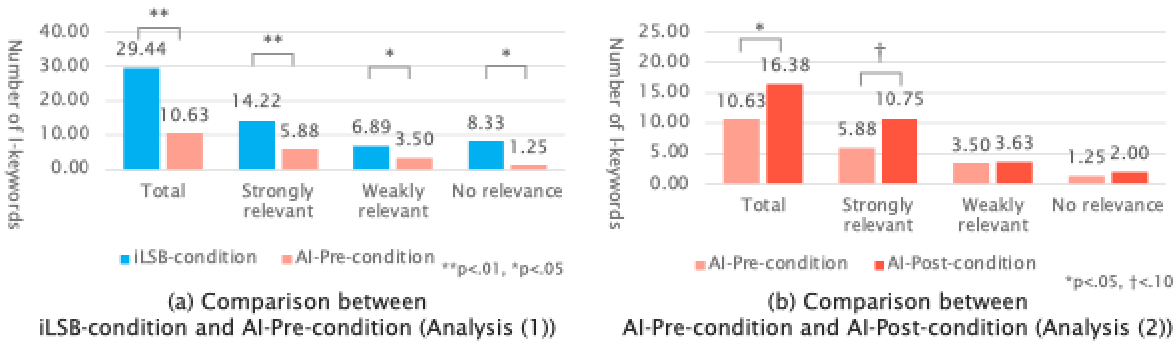

Figure 2. Number of I-keyword at Each Relevance Level

In the analysis (3), we conducted the same test as that of the analysis (1) for iLSB and iLSB-AI conditions. The minimum QE in iLSB-condition was 0.06 and that in iLSB-AI-condition was 0.02, and there was a significant difference between them after arcsine transformation ($t(16)=2.55$, $p=0.025$, Cohen's $d=1.20$). Also, the degree of links in iLSB-condition was 3.89 and that in iLSB-AI-condition was 2.67, and there was a tendency of significant difference ($t(16)=1.87$, $p=0.083$, $d=0.08$). These results suggest that participants in iLSB-condition investigated more deeply and widely compared with iLSB-AI-condition. In other words, iLSB with generative AI prevents learners from executing question expansion compared with the use of iLSB, which rejects H3.

4.7 Discussion

According to Figure 2 (a), there were significant differences between the number of I-keywords at all levels of relevance in iLSB and AI-Pre conditions (strongly relevant: $t(15)=4.33$, $p=0.00098$, Cohen's $d=2.03$, weakly relevant: $t(15)=3.12$, $p=0.096$, Cohen's $d=1.56$, no relevance: $t(15)=2.76$, $p=0.014$, Cohen's $d=1.35$). These results show that the participants using iLSB investigated strongly relevant information more, but they concurrently investigated weakly relevant or no relevant information. In order to support learners who tend to investigate such information, we have already developed the iLSB's function for diagnosing the appropriateness of sub-questions expanded (Sato et al, 2019). According to Figure 2 (b), in addition, there was a significant difference between the numbers of strongly relevant I-keywords in AI-Pre and AI-Post conditions ($t(7)=2.24$, $p=0.060$, $d=0.92$). This shows that the participants investigated a greater number of strongly relevant information through learning the model.

In the analysis (3), H3 was rejected. According to the questionnaire about the difficulties in using generative AI integrated with iLSB, it is considered that learners in iLSB-AI-group had difficulties in expanding more questions than iLSB-group because they were not presented how and when to use generative AI in the investigative learning process.

In addition, we analyzed how learners could use generative AI effectively by evaluating the validity of each sub-question in the learning scenarios. In this evaluation, three evaluators used the same Web resources as the ones used for the relevance evaluation to judge whether sub-questions relevant to the initial question and their parent questions were valid. According to the questionnaire about how learners in iLSB-AI group used generative AI, we found that learners who expanded many valid sub-questions tended to use generative AI for gathering background knowledge before investigation on the Web, and organizing what they learned from Web resources to construct their knowledge.

5. Conclusion

In this paper, we have ascertained the availability and effectiveness of using generative AI in Web-based investigative learning through three analyses. As a result of the analyses, it is suggested that learners could investigate information less sufficiently when using generative AI compared with iLSB. It is also suggested that the model promoted investigating more information, especially strongly relevant information when using generative AI. In addition, it is suggested that the use of generative AI in iLSB is effective in some situations. In the future, we should discuss an effective way to use generative AI in detail by following these findings.

References

Henze, N., & Nejdl, W. (2001). Adaptation in open corpus hypermedia. *International Journal of Artificial Intelligence in Education, 12*(4), 325–350.

Hill, J. R., & Hannafin, M. J. (1997). Cognitive strategies and learning from the world wide web. *Educational Technology Research and Development, 45*(4), 37–64. https://doi.org/10.1007/BF02299682

Kashihara, A. (2023). Design of Web-Based Investigative Learning. *Transactions of Japanese Society for Information and Systems in Education, 40*(2), 105-116. https://doi.org/10.14926/jsise.40.105 (In Japanese)

Kashihara, A. & Akiyama, N. (2017). Learning Scenario Creation for Promoting Investigative Learning on the Web. *The Journal of Information and Systems in Education, 15*(1), 62-72. https://doi.org/10.12937/ejsise.15.62

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103,* 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lo, C. K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

OpenAI. (2024) Models - OpenAI Platform. https://platform.openai.com/docs/models/

Sato, Y., Kashihara. A., Hasegawa, S., Ota, K., & Takaoka, R. (2019). Promoting reflection on question decomposition in Web-based investigative learning. *Proceedings of the 27th International Conference on Computers in Education (ICCE2019)*, 1, 75-80.

Zumbach, J., & Mohraz, M. (2008). Cognitive load in hypermedia reading comprehension: Influence of text type and linearity. Computers in Human Behavior, 24(3), 875–887. https://doi.org/10.1016/j.chb.2007.02.015