Ishihara et al. *Research and Practice in Technology Enhanced Learning* (2026) 21:39
https://doi.org/10.58459/rptel.2026.21039

Research and Practice in
Technology Enhanced Learning

**RESEARCH** **Free and Open Access**

# Discovering the links between real-world activities and previous course contents: The potential of information retrieval using large language models

Manabu Ishihara [1] *, Izumi Horikoshi [2] and Hiroaki Ogata [3]

*Correspondence:
ishihara.manabu.46z@st.kyoto-u.ac.jp
Graduate School of Informatics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto 6068501, Japan
Full list of author information is available at the end of the article

**Abstract**

In experiential learning involving real-world activities, such as fieldwork and pre-service training, transferring knowledge into practice is essential. While reflection is a critical component of this process, it is challenging to review how previously learned knowledge has been utilized. To address this issue, this study connected student descriptions of real-world activities with relevant course contents using information retrieval techniques and large language models (LLMs). The validity of linking was evaluated for one approach without LLM and three approaches that employ LLMs differently. These approaches were applied to a dataset collected from a university course in Japan. There were conditions for the inclusion or exclusion of supplemental information. The results indicated the supremacy of LLM-featured approaches without supplemental information. However, we found that these performances have not yet been stable. The findings and discussions shed light on the potential of the LLM-featured retrieval approaches for data-enhanced reflection across in-class knowledge acquisition and real-world knowledge applications.

**Keywords:** Experiential learning, Reflection, Natural language processing, Large language models, Information retrieval

## Introduction

In higher education, students are expected to apply knowledge acquired through coursework to real-world situations. Educators facilitate this transition through various programs, such as fieldwork, internships, and pre-service training. These approaches often incorporate classroom activities, for example, preparatory lectures and reflective activities

(Lee et al., 2020). This aligns with the experiential learning cycle (Gibbs, 1988; Kolb, 1984), a pedagogical framework that values conceptualization through reflection along with hands-on experiences.

In practice, however, transitioning from real-world activities to reflection poses challenges. Students often struggle to identify the relevance between what they have been taught in class and their field experiences (Boud et al., 1985), despite the importance of reflective practice in bridging theoretical learning and its application beyond the classroom (Harvey et al., 2016). This cannot be overlooked because such sense-making processes are positively associated with better learning outcomes, including critical thinking and attitude toward further learning (Laird et al., 2014). Also, educators face difficulties in accessing students' thoughts (Boud et al., 1985). They need student outputs, such as reflective writings, to facilitate and evaluate the learning process in experiential learning programs. Nevertheless, manually reviewing these outputs becomes challenging, particularly when dealing with numerous detailed descriptions. The significance of real-world activities will be limited if the issues remain.

Data-enhanced approaches, especially natural language processing (NLP) techniques, can address these challenges. For example, topic modeling can identify key themes within student journals (Chen et al., 2016). Co-occurrence network analysis enables the extraction of keywords and their associations from reflective writings (Lebowitz et al., 2020; Shikama et al., 2021). Recent studies have employed large language models (LLMs) to assess the quality of reflective writings (Masikisiki et al., 2023; Nehyba & Štefánik, 2023; Zhang et al., 2024). Despite these advancements, limited research has explored how previously learned knowledge was reflected in the outcomes of experiential learning activities.

To address this gap, it is necessary to match relevant previous course contents with student descriptions. This can be framed as an information retrieval task: taking student descriptions as a query and searching for relevant course contents. Information retrieval is a fundamental research area that has contributed to Internet browsing, question-answering systems, digital libraries (Hambarde & Proença, 2023), and educational technologies such as book or material recommendation systems (Nuipian & Chuaykhun, 2023; Takii, Flanagan, et al., 2024). This study area is becoming increasingly active with the rise of LLMs (Zhu et al., 2023). However, few studies have applied information retrieval techniques and LLMs to connect student descriptions with external information.

This study aims to evaluate the applicability of information retrieval techniques involving LLMs in identifying students' mentions of previous course contents in their outcomes of real-world activities. For that, we formulated the following research questions:

RQ1: To what extent can information retrieval techniques involving LLMs correctly identify course contents relevant to student descriptions?

RQ2: Can the inclusion of supplemental information enhance the validity of information retrieval results?

## Literature review

### Learning through classroom and real-world activities

Knowledge and meaning are constructed by assimilating new experiences and reforming existing understanding (Bada, 2015). The real-world situations serve as a learning environment in terms of forging students' knowledge and skills for future applications through practical experiences (Jamil et al., 2021).

The pedagogical framework called experiential learning cycle supports educators in integrating real-world activities into their learning designs. This cycle comprises abstract conceptualization, active experimentation, concrete experience, and reflective observation (Gibbs, 1988; Kolb, 1984). It implies the importance of prior learning and post-activities to maximize takeaways from real-world activities. Thus, educators often include pre- and post-classroom activities in their learning designs. For example, fieldwork (Jemec Auflič et al., 2019), internships (Lambert Snodgrass et al., 2023), and pre-service training (Lin, 2021).

### Relating previous course contents to real-world experiences

#### *Conventional approaches*

Despite the pedagogical foundation, integrating classroom learning with real-world activities remains challenging. It can be difficult to relate "what was learned in class" to "how it was applied to real-world activities." Conventionally, educators encourage students to document their experiences and thoughts. For instance, worksheets provide a clear structure for knowledge gain in students' autonomous activities (Basten et al., 2014). Also, portfolios facilitate students' self-reflection and can work as an assessment tool for field experiences (Prince, 2004). Reflective writing helps students achieve deeper insights by re-examining their basic assumptions (Asselin, 2011; Boud et al., 1985). Such sense-making process enhances understanding of experiences and fosters an attitude of continuous update of knowledge and skills (Fernandez et al., 2015). However, it is difficult to manually review such outcomes by considering their relevance to previously learned topics.

Computational approaches, specifically NLP techniques, can automatically extract insights from numerous texts. For example, topic modeling can be used to clarify the main themes within text data. It has been employed to evaluate pre-service teachers' journals (Chen et al., 2016). Co-occurrence network analysis can depict notable associations of frequently appearing words. Prior studies explored medical students' key perceptions of pre-service training in their reflective writings (Lebowitz et al., 2020; Shikama et al., 2021). Also, machine learning classification has been used to assess the content of student teachers' reflective writings and essays (Nehyba & Štefánik, 2023; Ullmann, 2019). However, these studies did not link external information beyond student descriptions.

Some computational approaches have attempted to relate classroom learning to real-world activities. A study on language learning drew links between in-class vocabulary learning logs, stored in an e-book system, and the logs on a mobile device for out-of-class vocabulary practice (Mouri et al., 2018). These links occurred when the same words were recorded in both in-class and out-of-class settings. However, word-matching can overlook different word expressions. Another study addressed this issue by grouping similar important words into key phrases. Key phrases from student descriptions and a dataset of course contents were converted into vectors. Subsequently, the cosine similarity between two vectors was calculated. Relevant course contents emerged based on the ranking of aggregated similarity scores (Ishihara et al., 2024). Although the study merely demonstrated linking without evaluation, the method was similar to information retrieval techniques, such as news article searching (Sarwar et al., 2022).

### *Potential of information retrieval and LLMs*

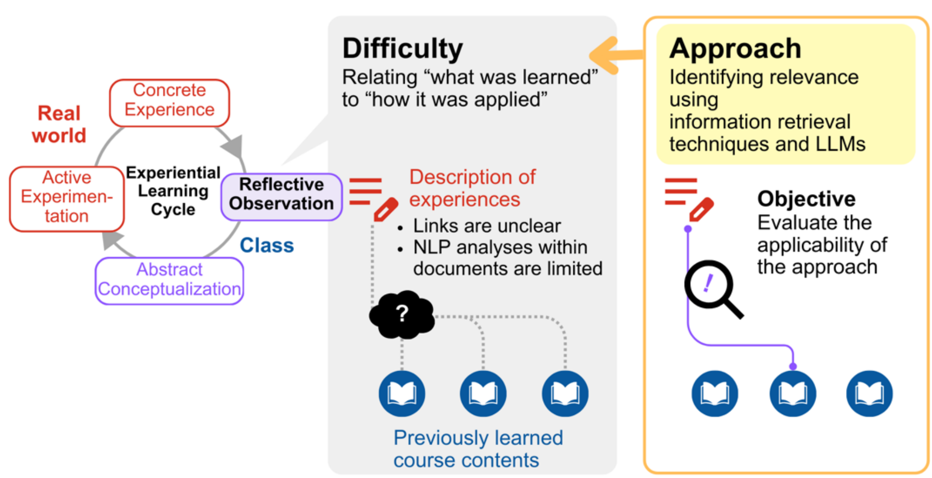
Our interest can be reframed as a task of information retrieval: "taking student descriptions as a query, searching for potentially relevant course contents, and outputting the most relevant search results." Information retrieval is a fundamental study area that has contributed to Internet browsing, question-answering systems, digital libraries, and so forth (Hambarde & Proença, 2023). In this domain, word embedding, an NLP technique that converts words into vectors, has become popular. In particular, pre-trained embedding models represent words with similar meanings as vectors with high similarity (Asudani et al., 2023). This nature is widely used to find semantically similar search targets by measuring vector similarity.

Along with the thriving embedding-based retrieval, LLMs have attracted the attention of researchers to handle complex natural languages. LLMs can overcome vocabulary gaps and guess user intents from dialogue-like inputs (Zhai, 2024). Notably, model fine-tuning is not mandatory to answer a wide range of queries. That allows users and developers to avoid retraining the models with additional sets of sample queries and responses. Instead of fine-tuning, a modern approach called Retrieval-Augment Generation (RAG) helps LLMs answer complex and knowledge-intensive search tasks by fetching additional information from external sources (Gao et al., 2023). In education, RAG is often used for question-answering systems. For instance, an LLM-based chatbot called Q-Module-Bot adopts RAG to automatically answer inquiries from online course takers by referring to relevant course content (Allen et al., 2024). Similarly, a question-answering system dedicated to dental education called DentQA uses RAG to generate answers by fetching relevant information from credible materials (Prakash & Prakash, 2024). Furthermore, LLMs can function as search agents responsible for the entire process of information retrieval (Zhu et al., 2023). Several question-answering systems simply rely on an LLM's pre-trained knowledge without any supplementary input and fine-tuning. Even so, previous studies reported that medical advice generated by their systems showed comparable accuracy to human experts in certain knowledge domains (Jo et al., 2024; Taylor et al., 2024).

However, scarce studies have addressed the applicability of information retrieval techniques and LLMs for identifying previously learned course contents from student descriptions of real-world activities. In contrast to the queries in previous studies, our queries, writings about real-world activities, may include more irrelevant information such as addresses of visited places. Also, course-related knowledge can be mentioned as additional thoughts along with the main topics. It is worth examining to what extent recent information retrieval techniques can handle such queries. Thus, we aim to assess the potential of information retrieval techniques for supporting reflection in the experiential learning cycle, as described in Figure 1.

**Fig. 1**

Objective of the study



# Method

## Data collection

To apply information retrieval techniques, we collected data from a university course in Japan, conducted during the fall semester of 2023. The course subject was human interface. This weekly course comprised 12 classes, and 41 undergraduate students attended. The classes were held in person for lectures and student presentations, with five teachers taking turns to lead. The teachers used Moodle, an online learning management system, to share lecture materials and receive assignments online. The main language was Japanese.

From the first to the fourth lectures, there were weekly assignments that required a real-world activity. Each week, students learned new course contents on human interface, accompanied by a lecture material dedicated to that week. Following each lecture, they tackled a writing assignment that required them to find and analyze interfaces in life space based on the course contents of the week. The past lecture materials were accessible on Moodle at any time. Students submitted their weekly assignments before the next class to share their findings in the class. In this study, this sequence of activities across classroom and real-world settings is referred to as “the There and Back Again process.” The course included four iterations of the There and Back Again process from the first to the fifth week.

With the consent of the teacher responsible for the second to the fifth week, we obtained the dataset from the There and Back Again process for the second and the third weeks. The dataset included lecture materials (comprising page titles and body texts), student descriptions in the assignments, and true labels assigned by the teacher to each assignment page. Details of the dataset are described in subsequent sections. Figure 2 shows an overview of the data collection, and Table 1 summarizes the dataset.

**Fig. 2**

Data collection from the There and Back Again processes

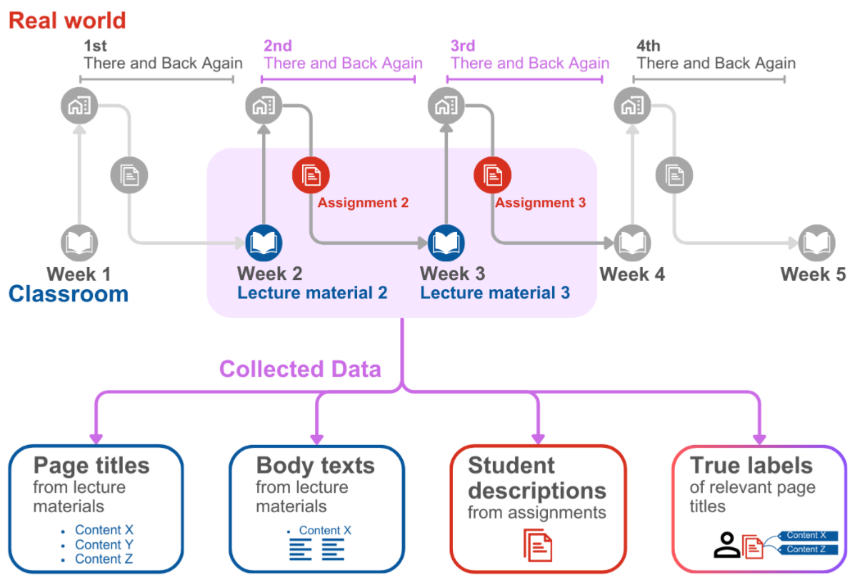



**Table 1**

Dataset

| Data | Category | Summary | Week 2 | Week 3 |
|---|---|---|---|---|
| Page titles | Course contents | Titles of the selected pages from the lecture materials | 13 items | 18 items |
| Body texts | Supplemental information | Body texts of the selected pages from the lecture materials | 13 items | 18 items |
| Student descriptions | Assignments | Outcomes of real-world activities submitted by 17 students | 22 total pages (Mean: 1.29, SD: 0.59) | 24 total pages (Mean: 1.41, SD: 0.71) |
| True labels | Course contents | Relevant course contents assigned by the teacher, including "Not Applicable" labels | 66 labels | 72 labels |

SD: Standard Deviation

## Dataset of course contents and supplemental information

We collected course contents from the lecture materials in Weeks 2 and 3, created by the teacher based on a reference book and shared as portable document format (PDF) files. We selected the pages introducing key topics on human interface. The pages related to class instructions, example cases, and lecture summaries were excluded. The titles of the selected pages were regarded as course contents, as they were supposed to represent the main ideas of each page. Eventually, 13 and 18 course contents were gathered for Weeks 2 and 3, respectively (Appendix 1). In addition, we collected the body texts corresponding to these page titles. They were regarded as supplemental information providing the details of each page.

## Dataset of student descriptions

For student descriptions of the real-world activities, we collected PDF files of writing assignments submitted by 17 students consented to the use of their data. Most descriptions were written in Japanese, with occasional English terms and several images of human interfaces. From the PDF files, we extracted texts and page numbers using the Python library PDFMinerLoader through the LangChain framework. After normalizing the texts, we manually removed personal information such as student identifications, names, and affiliations. We also eliminated irrelevant texts transcribed from the images. The total, mean, and standard deviation of pages in Assignment 2 were 22 pages, 1.29 pages, and 0.59 pages, respectively. For Assignment 3, these values were 24 pages, 1.41 pages, and 0.71 pages, respectively.

## Dataset of true labels

To evaluate the information retrieval approaches, we asked a teacher to assign true labels to the assignment pages. The teacher led the classes from Week 2 to 5, including creating the lecture materials and designing the There and Back Again process. We obtained the true labels solely from the teacher instead of involving other annotators. This is because the labeling required to have highly specific contexts of the course and subject knowledge.

We provided the teacher with instructions on labeling and spreadsheets. According to the instructions, the teacher assigned up to three labels of relevant course contents for each page of Assignments 2 and 3 on the spreadsheets. The spreadsheets contained the page-by-page text of preprocessed student descriptions. There were also columns for pseudo-student numbers, indices of PDF files, week numbers, and page numbers. The teacher selected up to three labels using the drop-down lists. For Assignment 2, the drop-down lists suggested only the course contents from Week 2. For Assignment 3, only the course contents from Week 3 appeared. In case of no relevant course contents on a page, the teacher could select "-" to indicate not applicable (N/A). The original PDF files of the assignments and lecture materials were available for reference. As a result, we collected 66 and 72 labels for Assignments 2 and 3, respectively, including the N/A labels.

## Information retrieval approaches

For RQ1, we employed three types of LLM-featured approaches: RAG, Query Rewriting RAG, and LLM Agent. Also, we set up one approach named Embedding that solely relies on a pre-trained embedding model without LLM. The details are explained in the following sections. Each approach takes student descriptions per assignment page as a query and returns up to three labels of course content or an N/A label. For the queries from Assignment 2, relevant course contents in Week 2 were returned. For those from Assignment 3, relevant course contents in Week 3 were returned.

These approaches were defined by following the two-stage model of information retrieval that consists of the retrieval and ranking stages. The retrieval stage broadly fetches candidate results, and the ranking stage filters them for the final outputs (Hambarde & Proença, 2023). In addition, part or the entire stages in the LLM-featured approaches are enhanced by LLMs as rewriter, retriever, reranker, reader, or search agent (Zhu et al., 2023). A rewriter LLM refines the query. A retriever LLM considers the intent of query then generates document identifiers. A reranker LLM selects final outputs according to certain conditions in its prompt. A reader LLM formats the final outputs in a favorable way. A search agent LLM handles the entire retrieval process.

The LLM-featured approaches consistently used the same model called generative pre-trained transformer 4 omni (GPT-4o). This model is equipped with a wide range of domain knowledge such as mathematical problem-solving (Lin et al., 2024), legal knowledge (Katz et al., 2023), and medical knowledge (Takagi et al., 2023). It also performed well for open-ended question-answering tasks in biology, earth science, and physics (Rodrigues et al., 2024). To obtain consistent outputs, we minimized its parameter of temperature, randomness of outputs, as zero.

For RQ2 regarding the supplemental information, we ran the approaches under two conditions: whether the search targets include supplemental information or not. We considered the body texts of lecture materials as supplemental information. This was expected to enrich the context of the selected page titles. The retrieval approaches could more accurately identify relevant course contents by considering the full content of material pages. Figure 3 and Table 2 show the overview of approaches.

**Fig. 3**

Process of information retrieval per approach

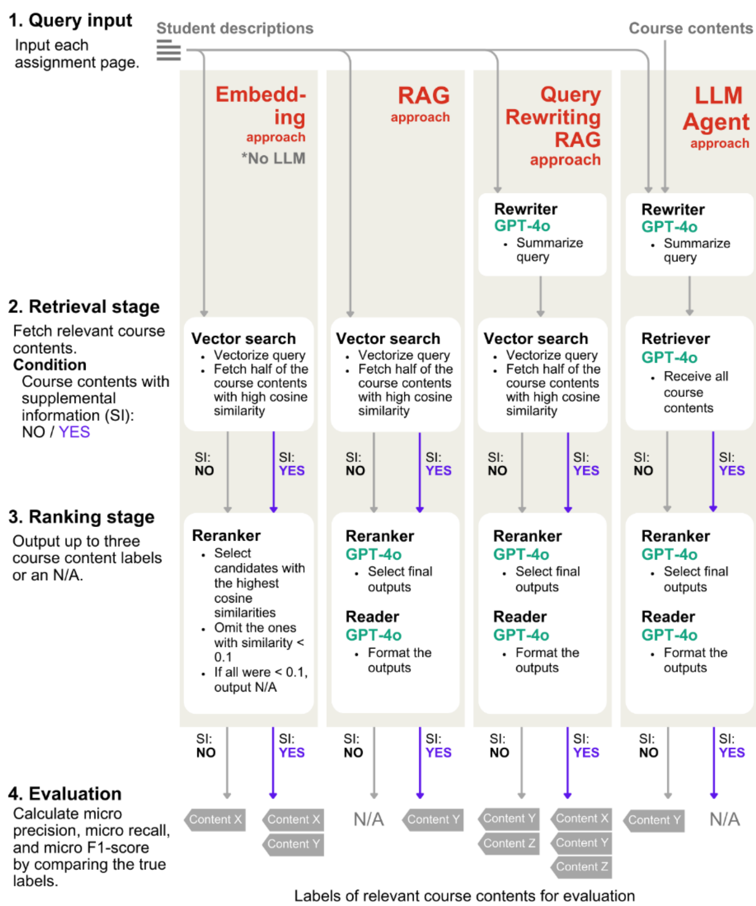



**Table 2**

Information retrieval approaches

| Approach | Supplemental information | LLM modules in the retrieval process |
|---|---|---|
| Embedding | No (Search for page titles) | Rewriter: None<br>Retriever: None<br>Reranker: None<br>Reader: None |
| | Yes (Search for page titles with body texts) | |
| RAG | No | Rewriter: None<br>Retriever: None<br>Reranker: GPT-4o for selecting final outputs<br>Reader: GPT-4o for formatting final outputs |
| | Yes | |
| Query Rewriting RAG | No | Rewriter: GPT-4o for query summarization<br>Retriever: None<br>Reranker: GPT-4o for selecting final outputs<br>Reader: GPT-4o for formatting final outputs |
| | Yes | |
| LLM Agent | No | Rewriter: GPT-4o for query summarization<br>Retriever: GPT-4o for receiving course contents<br>Reranker: GPT-4o for selecting final outputs<br>Reader: GPT-4o for formatting final outputs |
| | Yes | |

### *Embedding approach*

As a benchmark, this approach identifies relevant course contents through vector search with no enhancement of LLMs. First, we prepared a vector database using the Python library langchain_chroma and stored the course contents vectorized by OpenAI's text-embedding-3-small model. This pre-trained embedding model can take long text inputs and convert each input into a vector of 1,536 dimensions. It was also used for vectorizing the queries. In the retrieval stage, half of the stored course contents were fetched from the vector database based on high cosine similarity to a query vector. The cosine similarity represents the semantic similarity between the query and each search target. The range of cosine similarity is from -1 to 1. A value of 1 denotes that the two vectors have the same meaning. A value of 0 indicates that the vectors are unrelated. A value of -1 means that the two vectors have opposite meanings. In the ranking stage, up to three results with high cosine similarity scores were returned for each query. If the top three course contents showed low cosine similarity scores less than 0.1, the candidates were omitted. If all candidates were omitted, an N/A was indicated.

### *RAG approach*

This approach features a reranker LLM in the ranking stage that enables RAG techniques. RAG is a combination of vector search and a reranker LLM. By retrieving domain-specific knowledge from an external vector database, the reranker can more accurately select final outputs in tasks that require local or contextual knowledge. In the retrieval stage of this approach, as well as the Embedding approach, the cosine similarity between a query and each course content in the vector database will be measured. The course contents with high similarities to the query will be passed to the ranking stage. Then, the query and the retrieved course contents will be inserted in the prompt for GPT-4o (Appendix 2). According to the prompt, the reranker GPT-4o selects the final outputs, and the reader GPT-4o formats them as labels of course contents or an N/A label.

### *Query Rewriting RAG approach*

This approach is an extension of the RAG approach above. It involves an additional GPT-4o module as a rewriter that modifies queries in a retrieval-friendly way. Query rewriting is a common technique for question-answering systems to improve retrieval performance by restating ambiguous queries (Vakulenko et al., 2021). In our approach, a rewriter GPT-4o takes each query and summarizes the analytical part of student descriptions within 100 characters in Japanese. It intends to minimize redundant information in the raw queries. Using summarized queries, the subsequent vector search could more accurately find relevant candidate course contents by paying attention to important mentions in student descriptions.

### *LLM Agent approach*

The last approach fully relies on LLM modules. GPT-4o serves as a search agent and carries out the entire process. The rewriter module summarizes queries as well as the Query Rewriting RAG approach. The summarized queries will be passed to the retrieval stage. Here, the retriever module receives both a summarized query and a list of course contents instead of conducting a vector search. In the ranking stage, the reranker module considers all course contents to determine the final outputs, whereas other approaches consider half of the candidates. Lastly, the reader module formats the outputs.

### Evaluation

For each approach, we evaluated the validity of the information retrieval results, that is, the performance of relating previous course contents to student descriptions. We calculated micro precision, micro recall, and micro F1-score using the Python library sklearn. Generally, precision indicates the extent to which a classifier does not classify false labels as true. Recall denotes how well a classifier finds all true labels. The F1-score is a harmonic mean of the precision and recall scores that indicates the integrated performance. Micro versions of the precision, recall, and F1-score can be used to evaluate the overall performance of several multi-label classifications based on the counts of total true positives, false positives, and false negatives.

## Results

### Validity of the approaches

Regarding RQ1, we confirmed that the LLM-featured approaches resulted clearly better than the Embedding approach. For the linking tasks, identifying relevant course contents without LLM was not effective. More importantly, we found that the performances of the approaches differed for Assignments 2 and 3, as shown in Table 3. For Assignment 2, micro F1-scores, the overall validity indicator, remained under 0.60 in all approaches. The highest micro F1-score of 0.59 belonged to the LLM Agent approach that had no supplemental information. As for Assignment 3, the micro F1-scores were above 0.60, except the results of the Embedding approach. The highest F1-score of 0.78 was observed in the RAG approach, which had no supplemental information. Our LLM-featured approaches showed favorable validities for Assignment 3 with 18 search targets. Compared to that, the results for Assignment 2 with 13 search targets were clearly limited. Intuitively, a larger number of search targets makes retrieval tasks difficult, yet the results of the LLM-featured approaches indicated an opposite tendency.

**Table 3**

Overall results

| Approach | Supplemental Information | Assignment 2 | | | Assignment 3 | | |
|---|---|---|---|---|---|---|---|
| | | Micro precision | Micro recall | Micro F1-score | Micro precision | Micro recall | Micro F1-score |
| Embedding | No | 0.25 | 0.34 | 0.29 | 0.02 | 0.03 | 0.02 |
| | Yes | 0.22 | 0.29 | 0.25 | 0.05 | 0.08 | 0.06 |
| RAG | No | 0.54 | 0.50 | 0.52 | 0.73 | **0.83** | **0.78** |
| | Yes | 0.54 | **0.58** | 0.56 | 0.69 | 0.78 | 0.73 |
| Query Rewriting RAG | No | 0.52 | 0.45 | 0.48 | 0.74 | 0.80 | 0.77 |
| | Yes | 0.53 | 0.47 | 0.50 | 0.69 | 0.72 | 0.71 |
| LLM Agent | No | **0.61** | **0.58** | **0.59** | 0.69 | 0.55 | 0.61 |
| | Yes | 0.57 | 0.55 | 0.56 | **0.76** | 0.70 | 0.73 |

It is possible that the features of the approaches affected the validity. The two RAG-based approaches surpassed the LLM Agent approach for Assignment 3, whereas the agent performed better than others for Assignment 2. The potential cause is the number of candidates filtered in the retrieval stage. The RAG-based approaches narrowed down half of the course contents in the stage. Possibly, these approaches omitted irrelevant search targets well when relevance could easily be inferred from queries. When the course contents were subtly mentioned or hard to infer from descriptions, the filtering process may not work adequately. In such cases, it would be better to take into account all candidates, as the LLM Agent approach did.

Regarding the use of an LLM as a query rewriter, it might not be impactful for the validity. The Query Rewriting RAG and the LLM Agent approaches summarized the queries to focus on the students' analytical points for each assignment page. However, the results did not show any notable tendency attributed to the feature. Possibly, relatively short lengths of descriptions did not need summarization to discern relevant points to the course contents. Rather than the presence of a rewriter LLM, the main difference was derived from the retrieval stage, as mentioned above.

The validity could also be influenced by the abstractness of course contents. In fact, the reranker LLM overly found relevance to specific course contents, according to the frequency of the outputs and the true labels. For the task of Assignment 2, the course contents "Mental Model," "Interaction Design," and "Affective Computing" were observed, even though the true labels did not include them (Figure 4). For Assignment 3, "Gesture Interface" and "WIMP" appeared in the same way (Figure 5). Also, several course contents were excessively output more than the actual numbers of the true label, such as "Telepresence" in Assignment 2 and "Brain-Computer Interface" in Assignment 3. It is possible that the wording of such course contents has general meanings in LLMs, and that made the reranker LLM to overly detect relevance from many parts of descriptions.

**Fig. 4**

Frequency of course content labels for Assignment 2 in the LLM-featured approaches

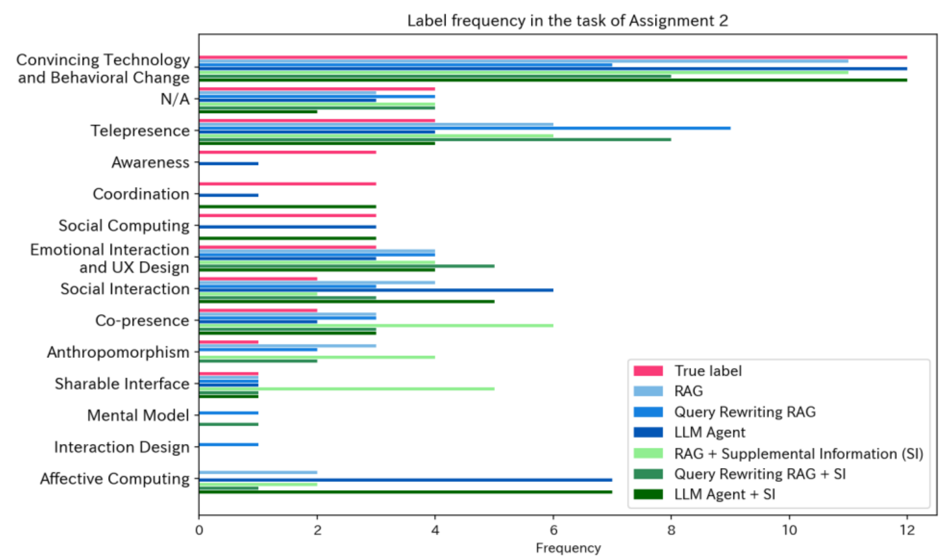
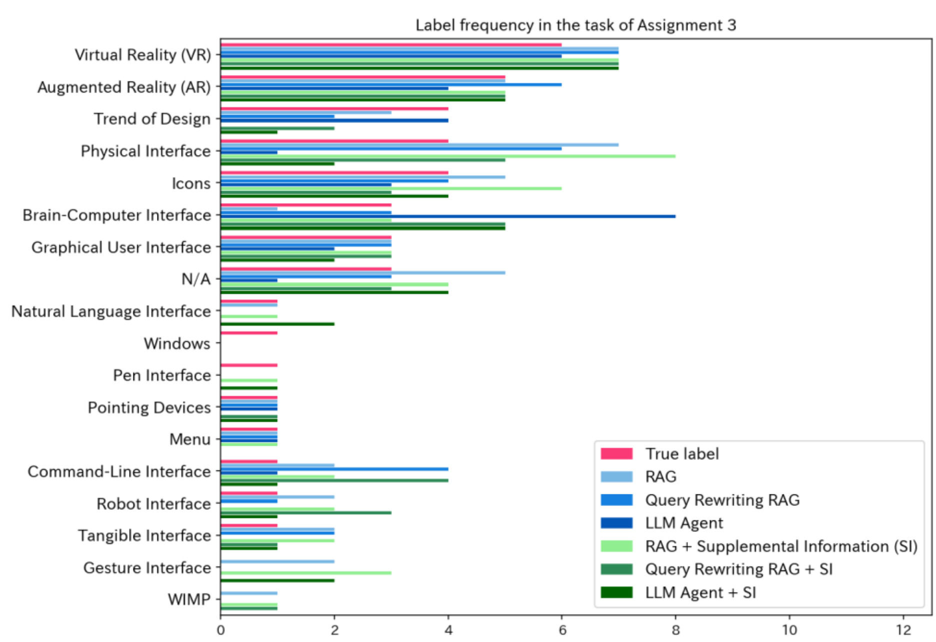



**Fig. 5**

Frequency of course content labels for Assignment 3 in the LLM-featured approaches



The potential reason for the higher validity in the task for Assignment 3 is that the wording of course contents was inferable from the descriptions of actual interfaces compared to Assignment 2. The course contents for Assignment 3 were the genre of human interface. That might be easier for LLMs to link the genre and described interfaces. On the other hand, associating the mentions about actual interfaces with the abstract concepts for Assignment 2, social and emotional factors, seemed to be more difficult. To reduce LLMs' excessive sensitivity to certain course contents, a dataset of course contents or supplemental information may need to include articulate information about how the concepts would be represented as real-world entities and functions.

### Influence of supplemental information

Given that the high abstractness of course contents makes LLMs find more erroneous relevance, that was supposed to be mitigated by supplemental information. For RQ2, we expected that the body texts in the lecture materials articulate the meanings of course contents as supplemental information. Nonetheless, we observed that the approaches without supplemental information yielded the best micro F1-scores for both Assignments 2 and 3.

Interestingly, the use of supplemental information showed positive and negative effects depending on the approaches and tasks. The micro F1-scores of the two RAG-based approaches increased for Assignment 2 but decreased for Assignment 3 when these approaches had supplemental information. Possibly, detailed information of the lecture facilitated the RAG-based approaches when their search targets contained a number of abstract course contents. Such information would offer clues to distinguish relevant course contents to queries. Contrary, the LLM Agent approach showed a negative impact on Assignment 2 and a positive impact on Assignment 3. The supplemental information might be noisy when the agent seeks relevance by considering all course contents that frequently include abstract concepts. In the case of Assignment 3, the definitions or examples of the genre of human interface in supplemental information worked as expected, although the scores did not exhibit the highest validity.

## Discussion

### Key findings for RQ1

For RQ1 concerning the validity of information retrieval approaches involving LLMs, we confirmed that the LLM-featured approaches showed much better validity compared to the embedding-only approach. Nonetheless, their performances are not stable. This resonates with that the educational use of LLMs for complex classification tasks has not yet been promising (Yan et al., 2024). The LLM Agent approach performed better than the RAG-based approaches for the tasks of relating abstract concepts to the descriptions of actual interfaces. The RAG-based approaches performed well for the tasks of relating the genre of human interface to the student descriptions of actual interfaces.

From the results, we recognized the importance of the clarity of course contents as search targets. Abstract concepts for Assignment 2 possibly increased the difficulty level of inferring relevant course contents from the descriptions of real-world entities. The definitions of such concepts may require further clarification, rather than simply being extracted from lecture materials. Otherwise, abstract wordings can lead LLMs to find relevance excessively. In this regard, refinement of the course contents is a way to increase the validity of information retrieval for identifying pieces of knowledge application from student descriptions of real-world activities.

## Key findings for RQ2

The supplemental information in this study did not contribute to the best score for each task, despite partial positive effects. It was slightly beneficial for the RAG-based approaches to deal with abstract concepts in the task for Assignment 2. Also, supplemental information enhanced the LLM Agent approach in the task for Assignment 3. However, we found that supplemental information can be disruptive. The results of the RAG-based approaches for Assignment 3 indicated lower scores compared to those without supplemental information. The deterioration was also observed in the results of the LLM Agent approach for Assignment 2.

That implies, extracting the body texts from the lecture materials was not enough. In this study, the amount of extracted body texts varied depending on the page because several pages introduced concepts with images instead of text. Therefore, the impact differently appeared. To maximize the benefits of supplemental information, maintaining the quantity and quality of information should be considered.

## Implications for educational practice

To leverage the linking approaches with LLMs in educational practices, at this moment, educators may need to adjust their materials and instructions. In the context of the human interface course, it is recommended to prepare a list overviewing the course contents with an equal quantity of explanations that include definitions of concepts, effects on human behavior, and example interfaces. Constructing a course content dataset with detailed information could reduce the excessive sensitivity of LLMs to certain search targets. On the instruction side, it would be effective to explicitly ask students to develop analyses focusing on both visible and functional features and how these features affect user behaviors. That would facilitate the retriever and reranker LLMs to select appropriate course contents corresponding to the mentions of specific characteristics of certain concepts.

Despite the need for improvement, retrieval results could assist reflection by indicating students' well-exercised knowledge of course contents. An example scenario is that each student in a human interface course submits an analysis of actual interfaces via a learning management system. An information retrieval function behind the system identifies relevant previous course contents to the input. The query and the search results will be linked as part of networked data. The network can include identifiers of lecture materials. That enables a dashboard to suggest frequently mentioned course contents in the analyses of actual interfaces. On top of that, there will be recommendations of relevant materials and reflective questions triggering further learning, such as "You seemed to actively refer to Telepresence in the lecture material for Week 2 and Virtual Reality in the lecture material for Week 3. Can you state your common interests across these concepts? Can you find an example around you that represents the common interests?"

Such feedback could facilitate the sense-making process across the coursework and real-world activities. That would also stimulate self-monitoring and metacognitive awareness, which are the factors of self-regulated learning (Schraw, 1998). Further, in light of the fact that experiential learning is a cyclic process, such tangible suggestions would be the glue for future experiences. Students will be able to review their pieces of knowledge applications to form new objectives for upcoming real-world activities. This would accelerate the cycle of experiential learning that fosters continuous knowledge updating (Fernandez et al., 2015) and leads to higher learning outcomes (Laird et al., 2014).

Retrieval results could also be beneficial for educators in terms of grasping the class-wide tendency of knowledge application. Accumulating the related course contents from students' linking data can depict what kind of course contents were frequently mentioned by the students. Such salient course contents can be regarded as common interests among students and are expected to be stimulating topics of group discussions in subsequent class activities. The least related course contents are also insightful in terms of directing students' attention to knowledge that they have not yet exercised.

## Implications for research

Our study expands the means for understanding the relationships between students' knowledge acquisition and applications, particularly in activities involving real-world settings. Prior studies of mobile and ubiquitous learning have analyzed the out-of-class experiences, but the discussions are still scarce in terms of learning processes across different settings and supporting sense-making (Pishtari et al., 2020). The linking approaches in this study could address the remaining issues and improve learning designs. Capturing pieces of links over time can represent the temporal tendency of knowledge application. It would enrich personalized reflection designs with the links between in-class and out-of-class learning. Further, exploring the associations between the tendency of knowledge application and performance scores would accelerate discussions of effective patterns of knowledge application in experiential learning.

Moreover, the linking approaches can produce the synergy with the studies of learner models. Learner models aim to provide personalized support tailored to each student's learning paths and comprehension levels of learning topics. Researchers have developed learner models for online learning (Ain et al., 2024) and technology-enhanced classrooms (Long & Aleven, 2017; Takii, Liang, et al., 2024). Nonetheless, experiential learning activities have been discussed separately from this domain. The use of LLM-featured information retrieval could enrich learner models by linking the footprints of knowledge application in real-world activities. This would enable more precise personalized support for classroom and online learning settings. Also, considering the importance of consistent learning scenarios across curricular contexts and out-of-class experiences (Hsu et al., 2016), the combination of our approach and learner models could enable personalized mobile learning. Using a mobile device, customized real-world challenges can be proposed based on individuals' learning models, which preserve the records of previous classroom and real-world activities. Developing such initiatives could establish data-enhanced experiential learning that accelerates the cycle of knowledge acquisition, application, and conceptualization.

## Limitations

### Dataset preparation

This study has several limitations. First, the dataset was collected over a limited range. The subject was focused only on human interface. Also, the descriptions in the assignments were relatively brief, as the average length was under two pages (Table 1). Longer descriptions of different subjects are required to confirm the generalizability of this approach. Second, this study did not maintain the granularity of data on the course contents and supplemental information. Some page titles and body texts of lecture materials may have provided irrelevant nuance or scarce context in the information retrieval process. Future studies should use a well-defined dataset. LLMs could be useful for that purpose. For instance, automatically structuring the search target dataset as a knowledge graph (Zhu et al., 2024). Third, the dataset for evaluation was limited to two assignments, and the labeling was performed on each assignment page. As this study highly relied on one teacher's recognition, particularly those who led the There and Back Again process, it was difficult to obtain a substantial amount of true labels. For more precise evaluation, it would be preferable to acquire a greater number of true labels with detailed granularity, such as paragraph by paragraph or sentence by sentence.

### Use of LLMs

The LLM-featured approaches in this study have not yet fully exploited advanced techniques. To enhance performance, refining the prompt and modifying the configuration of LLMs would be effective. For example, incorporating a retriever LLM that can conduct web searches (Liu et al., 2023) can enrich supplemental information. Also, the parameters of the LLM can be further optimized, not limited to the temperature settings. Furthermore, advanced techniques such as graph retrieval-augmented generation (Edge et al., 2024) and the chain of thought method (Wei et al., 2022) may contribute to performance.

In addition, this study did not consider the images included in the assignments and the lecture materials. In view of recent advancements in LLMs capable of processing both text and images, visual information can be used for supplemental information (Caffagni et al., 2024). While the applicability of LLM-featured information retrieval approaches for relating course contents to descriptions of real-world activities was partially confirmed, its full potential remains to be explored through further research.

### Interpretation of captured links

Even though the retrieval approaches could accurately draw links between course contents and student descriptions, each link would not indicate the depth of learning. Student outputs may reflect different levels of knowledge application. According to the revised version of Bloom’s taxonomy, knowledge-building achievements can be classified as remembering, understanding, applying, analyzing, evaluating, and creating (Krathwohl, 2002). Although we briefly discussed the links as a medium of reminder and enhancement of metacognitive awareness, this study does not cover the interpretation of the links. To address the levels of understanding of the course contents, as mentioned earlier, it is worth integrating the idea of learner models.

## Conclusion

This study addressed the challenge of reflection in experiential learning, specifically the difficulty in identifying relevance between descriptions of real-world activities and previous course contents. Three information retrieval approaches involving LLMs and one approach without LLM were evaluated through multi-label classification tasks using a university course dataset. The results indicated that the RAG approach achieved the overall highest micro F1-score of 0.78. However, the performances of our approaches were changeable depending on the use of LLMs and tasks. Despite limitations, there are ways to increase the linking validity. With improvements, the linking approach could contribute to understanding and enhancing the process of knowledge acquisition and application involving real-world activities. It could enable more effective experiential learning by incorporating data-enhanced reflection and fostering awareness for future experiences.

## Appendix 1: Course contents

| Week (Lecture topic) | Header ID | Course content (original) | Course content in English (reference translation) |
|---|---|---|---|
| Week 2 (Social emotional interaction) | 2-6 | インタラクション・デザイン | Interaction Design |
| | 2-10 | メンタルモデル | Mental Model |
| | 2-19 | ソーシャルインタラクション | Social Interaction |
| | 2-20 | ソーシャル・コンピューティング | Social Computing |
| | 2-28 | テレプレゼンス（Telepresence) | Telepresence |
| | 2-31 | コプレゼンス（Co-presence) | Co-presence |
| | 2-32 | 協調（Coordination) | Coordination |
| | 2-33 | 気づき（Awareness) | Awareness |
| | 2-34 | 共有可能なインターフェース | Shareable Interface |
| | 2-43 | エモーショナル・インタラクションとUX デザイン | Emotional Interaction and UX Design |
| | 2-44 | Affective Computing | Affective Computing |
| | 2-50 | 擬人化（Anthropomorphism) | Anthropomorphism |
| | 2-52 | 説得する力のあるテクノロジーと行動変容 | Convincing Technology and Behavioral Change |
| Week 3 (Various Interfaces) | 3-16 | Command-Line インターフェース | Command-Line Interface |
| | 3-20 | Graphical User インターフェース | Graphical User Interface |
| | 3-21 | WIMP | WIMP |
| | 3-22 | ポインティングデバイス | Pointing Devices |
| | 3-24 | ウインドウ | Windows |
| | 3-25 | メニュー | Menu |
| | 3-27 | アイコン | Icons |
| | 3-32 | デザインのトレンド | Trend of Design |
| | 3-34 | Virtual Reality (VR): 仮想現実 | Virtual Reality (VR) |
| | 3-37 | Augmented Reality (AR): 拡張現実 | Augmented Reality (AR) |
| | 3-42 | ペンインタフェース | Pen Interface |
| | 3-43 | ジェスチャーインタフェース | Gesture Interface |
| | 3-44 | テーブルトップインタフェース | Tabletop Interface |
| | 3-45 | ロボットインタフェース | Robot Interface |
| | 3-46 | Brain-Computer インタフェース | Brain-Computer Interface |
| | 3-47 | 自然言語インタフェース | Natural Language Interface |
| | 3-48 | タンジブルインタフェース | Tangible Interface |
| | 3-49 | 物理的インタフェース | Physical Interface |

## Appendix 2: Prompts for GPT-4o in the ranking stage

| Condition | Prompt in Japanese (original) | Prompt in English (reference translation) |
|---|---|---|
| Common part | あなたは先生であり、「生徒の作文の中に、過去の授業内容に言及している部分があるか」を判定するプロである。<br>以下の作文について、授業内容に言及している部分がある場合、最大 3 つ、授業内容の冒頭の ID を出力せよ。<br>言及している部分がない場合、「-」と出力せよ。説明は出力しないこと。<br><br>出力例: 1_1, 1_2, 1_3<br>出力例: -<br><br>【作文】<br>{student_descriptions}<br><br>【授業内容】<br>{*} | You are a teacher and a professional who assesses “whether any part of a student description refers to previous course contents.”<br>For the following descriptions, if there are any mentions of course contents, output up to three header IDs of the course contents.<br>If there is no mention, output “-”. Do not output any explanation.<br><br>Output example: 1_1, 1_2, 1_3<br>Output example: -<br><br>[descriptions]<br>{student_descriptions}<br><br>[Course contents]<br>{*} |
| Supplemental information: NO | 【授業内容】<br>{page_titles} | [Course contents]<br>{page_titles} |
| Supplemental information: YES | 【授業内容】<br>{body_text} | [Course contents]<br>{body_text} |

**Abbreviations**
GPT-4o: Generative Pre-trained Transformer 4 omni; LLM: Large Language Model; N/A: Not Applicable; NLP: Natural Language Processing; PDF: Portable Document Format; RAG: Retrieval-Augment Generation; SD: Standard Deviation.

**Acknowledgements**
The authors would like to acknowledge all the contributors to this study, the reviewers, and the invitation from the ICCE 2024 committee.

**Author’s contributions**
Ishihara and Horikoshi collected the dataset. Ishihara conducted data analysis and drafted the manuscript. Horikoshi and Ogata provided insights. Ogata provided supervision of the study. All authors read and approved the final manuscript.

**Author’s information**
Manabu Ishihara is a doctoral student at the Graduate School of Informatics, Kyoto University, Japan. His research interests are data-enhanced learning design and learning analytics across in-class and out-of-class learning settings.

Izumi Horikoshi was an assistant professor at the Academic Center for Computing and Media Studies and the Graduate School of Informatics, Kyoto University, Japan. She currently works as a senior research fellow at Uchidayoko Institute for Education Research. Her research interests include learning analytics and classroom visualization for formative assessment and reflection.

Hiroaki Ogata is a full Professor at the Academic Center for Computing and Media Studies, Kyoto University, Japan. His research interests include learning analytics, evidence-based education, educational data mining, educational data science, computer supported ubiquitous and mobile learning, and CSCL.

**Funding**
This work was partly supported by CSTI SIP Program Grant Number JPJ012347 and JSPS KAKENHI Grant Number 23H00505.

**Availability of data and materials**
The data and materials cannot be shared because it falls outside the scope of the data providers' consent.

## Declarations

**Competing interests**
The authors declare that they have no competing interests.

**Author details**
[1] Graduate School of Informatics, Kyoto University, Japan

[2] Uchidayoko Institute for Education Research, Uchida Yoko Co., Ltd., Japan

[3] Academic Center for Computing and Media Studies, Kyoto University, Japan

Received: 6 January 2025 Accepted: 20 August 2025
Published online: 3 February 2026

**References**

Ain, Q. U., Chatti, M. A., Meteng Kamdem, P. A., Alatrash, R., Joarder, S., & Siepmann, C. (2024, March). Learner modeling and recommendation of learning resources using personal knowledge graphs. In *Proceedings of the 14th Learning Analytics and Knowledge Conference* (pp. 273–283). ACM. https://doi.org/10.1145/3636555.3636881

Allen, M., Naeem, U., & Gill, S. S. (2024). Q-module-bot: A generative AI-based question and answer bot for module teaching support. *IEEE Transactions on Education*, *67*(5), 793–802. https://doi.org/10.1109/te.2024.3435427

Asselin, M. E. (2011). Using reflection strategies to link course knowledge to clinical practice: The RN-to-BSN student experience. *Journal of Nursing Education*, *50*(3), 125–133. https://doi.org/10.3928/01484834-20101230-08

Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: A review. *Artificial Intelligence Review*, *56*(9), 1–81. https://doi.org/10.1007/s10462-023-10419-1

Bada, S. O. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, *5*(6), 66–70. https://api.semanticscholar.org/CorpusID:37780480

Basten, M., Meyer-ahrens, I., Fries, S., & Wilde, M. (2014). The effects of autonomy-supportive vs. controlling guidance on learners' motivational and cognitive achievement in a structured field trip. *Science Education*, *98*(6), 1033–1053. https://doi.org/10.1002/sce.21125

Boud, D., Keogh, R., & Walker, D. (Eds.), (1985). *Reflection: Turning experience into learning*. Routledge.

Caffagni, D., Cocchi, F., Moratelli, N., Sarto, S., Cornia, M., Baraldi, L., & Cucchiara, R. (2024). Wiki-LLaVA: Hierarchical retrieval-augmented generation for multimodal LLMs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1818–1826). IEEE. https://doi.org/10.1109/cvprw63382.2024.00188

Chen, Y., Yu, B., Zhang, X., & Yu, Y. (2016, April). Topic modeling for evaluating students' reflective writing: A case study of pre-service teachers' journals. In *Proceedings of the sixth International Conference on Learning Analytics & Knowledge* (pp. 1–5). ACM. https://doi.org/10.1145/2883851.2883951

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., & Larson, J. (2024). *From local to global: A graph RAG approach to query-focused summarization*. arXiv. http://arxiv.org/abs/2404.16130

Fernandez, S. C., Chelliah, K. K., & Halim, L. (2015). A peek into oneself: reflective writing amongst undergraduate Medical Imaging students. *Reflective Practice*, *16*(1), 109–122. https://doi.org/10.1080/14623943.2014.982524

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). *Retrieval-Augmented Generation for large Language Models: A survey*. arXiv. http://arxiv.org/abs/2312.10997

Gibbs, G. (1988). *Learning by doing: A guide to teaching and learning methods*. Further Education Unit.

Hambarde, K. A., & Proença, H. (2023). Information retrieval: Recent advances and beyond. *IEEE Access*, *11*, 76581–76604. https://doi.org/10.1109/access.2023.3295776

Harvey, M., Coulson, D., & McMaugh, A. (2016). Towards a theory of the ecology of reflection: Reflective practice for experiential learning in higher education. *Journal of University Teaching & Learning Practice*, *13*(2), 7–27. https://doi.org/10.53761/1.13.2.2

Hsu, T.-Y., Kuo, F.-R., Liang, H.-Y., & Lee, M.-F. (2016). A curriculum-based virtual and physical mobile learning model for elementary schools in museums. *The Electronic Library*, *34*(6), 997–1012. https://doi.org/10.1108/EL-08-2015-0146

Ishihara, M., Horikoshi, I., & Ogata, H. (2024, November). Linking real-world experiences with course contents: A text mining approach toward effective "there and back again." In Kashihara, A. et al. (Eds.), *Proceedings Vol. 1 of the*

*32nd International Conference on Computers in Education* (pp. 450–459). Asia-Pacific Society for Computers in Education. https://doi.org/10.58459/icce.2024.4881

Jamil, M. G., Alam, N., Radclyffe-Thomas, N., Islam, M. A., Moniruzzaman Mollah, A. K. M., & Rasel, A. A. (2021). Real world learning and the internationalisation of higher education: Approaches to making learning real for global communities. In Morley, D. A., & Jamil, M. G. (Eds.), *Applied Pedagogies for Higher Education* (pp. 107–132). Springer Nature. https://doi.org/10.1007/978-3-030-46951-1_6

Jemec Auflič, M., Kumelj, Š., Peternel, T., & Jež, J. (2019). Understanding of landslide risk through learning by doing: case study of Koroška Bela community, Slovenia. *Landslides*, *16*(9), 1681–1690. https://doi.org/10.1007/s10346-018-1110-1

Jo, E., Song, S., Kim, J.-H., Lim, S., Kim, J. H., Cha, J.-J., Kim, Y.-M., & Joo, H. J. (2024). Assessing GPT-4's performance in delivering medical advice: Comparative analysis with human experts. *JMIR Medical Education*, *10*(1), e51282. https://doi.org/10.2196/51282

Katz, D. M., Bommarito, M. J., Gao, S., & Arredondo, P. (2023). GPT-4 passes the bar exam. *Philosophical Transactions of the Royal Society A*, *382*(2270), 20230254. https://doi.org/10.2139/ssrn.4389233

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Laird, T. F. N., Seifert, T. A., Pascarella, E. T., Mayhew, M. J., & Blaich, C. F. (2014). Deeply affecting first-year students' thinking: Deep approaches to learning and three dimensions of cognitive development. *The Journal of Higher Education*, *85*(3), 402–432. https://doi.org/10.1353/jhe.2014.0017

Lambert Snodgrass, L., Ghahremani, M., & Hass, M. (2023). Live-learn-work: Experiential learning and cultural intelligence in the internship abroad. *Journal of Global Education and Research*, *7*(1), 50–63. https://doi.org/10.5038/2577-509x.7.1.1120

Lebowitz, A., Kotani, K., Matsuyama, Y., & Matsumura, M. (2020). Using text mining to analyze reflective essays from Japanese medical students after rural community placement. *BMC Medical Education*, *20*(1), 38. https://doi.org/10.1186/s12909-020-1951-x

Lee, H., Stern, M. J., & Powell, R. B. (2020). Do pre-visit preparation and post-visit activities improve student outcomes on field trips? *Environmental Education Research*, *26*(7), 989–1007. https://doi.org/10.1080/13504622.2020.1765991

Lin, T.-H. (2021). Revelations of service-learning project: Multiple perspectives of college students' reflection. *PLOS ONE*, *16*(9), e0257754. https://doi.org/10.1371/journal.pone.0257754

Lin, Y., Yang, E. F.-Y., Wu, J.-S., Yeh, C. Y. C., Liao, C.-Y., & Chan, T. (2024). Enhancing students' authentic mathematical problem-solving skills and confidence through error analysis of GPT-4 solutions. *Research and Practice in Technology Enhanced Learning*, *20*, 034. https://doi.org/10.58459/rptel.2025.20034

Liu, X., Lai, H., Yu, H., Xu, Y., Zeng, A., Du, Z., Zhang, P., Dong, Y. & Tang, J. (2023, August). WebGLM: Towards an efficient web-enhanced question answering system with human preferences. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 4549–4560). ACM. https://doi.org/10.1145/3580305.3599931

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, *27*(1), 55–88. https://doi.org/10.1007/s11257-016-9186-6

Masikisiki, B., Marivate, V., & Hlophe, Y. (2023, November). Investigating the efficacy of large language models in reflective assessment methods through chain of thought prompting. In *Proceedings of the 4th African Human Computer Interaction Conference* (pp. 44–49). ACM. https://doi.org/10.1145/3628096.3628747

Mouri, K., Uosaki, N. & Ogata, H. (2018). Learning analytics for supporting seamless language learning using e-book with ubiquitous learning system. *Journal of Educational Technology & Society*, *21*(2), 150–163. https://www.jstor.org/stable/26388390

Nehyba, J., & Štefánik, M. (2023). Applications of deep language models for reflective writings. *Education and Information Technologies*, *28*(3), 2961–2999. https://doi.org/10.1007/s10639-022-11254-7

Nuipian, V., & Chuaykhun, J. (2023, December). Book recommendation system based on course descriptions using cosine similarity. In *Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval* (pp. 273–277). ACM. https://doi.org/10.1145/3639233.3639335

Pishtari, G., Rodríguez‐Triana, M. J., Sarmiento‐Márquez, E. M., Pérez‐Sanagustín, M., & Väljataga, T. (2020). Learning design and learning analytics in mobile and ubiquitous learning: A systematic review. *British Journal of Educational Technology*, *51*(4), 1078–1100. https://doi.org/10.1111/bjet.12944

Prakash, K., & Prakash, R. (2024). An artificial intelligence-based dental semantic search engine as a reliable tool for dental students and educators. *Journal of Dental Education*, *88*(9), 1257–1266. https://doi.org/10.1002/jdd.13560

Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education*, *93*(3), 223–231. https://doi.org/10.1002/j.2168-9830.2004.tb00809.x

Rodrigues, L., Dwan Pereira, F., Cabral, L., Gašević, D., Ramalho, G., & Ferreira Mello, R. (2024). Assessing the quality of automatic-generated short answers using GPT-4. *Computers and Education: Artificial Intelligence*, *7*(100248), 100248. https://doi.org/10.1016/j.caeai.2024.100248

Sarwar, T. B., Noor, N. M., & Saef Ullah Miah, M. (2022). Evaluating keyphrase extraction algorithms for finding similar news articles using lexical similarity calculation and semantic relatedness measurement by word embedding. *PeerJ Computer Science*, *8*(e1024), e1024. https://doi.org/10.7717/peerj-cs.1024

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, *26*(1/2), 113–125. https://doi.org/10.1023/a:1003044231033

Shikama, Y., Chiba, Y., Yasuda, M., Stanyon, M., & Otani, K. (2021). The use of text mining to detect key shifts in Japanese first-year medical student professional identity formation through early exposure to non-healthcare hospital staff. *BMC Medical Education*, *21*(1), 389. https://doi.org/10.1186/s12909-021-02818-1

Takagi, S., Watari, T., Erabi, A., & Sakaguchi, K. (2023). Performance of GPT-3.5 and GPT-4 on the Japanese medical licensing examination: Comparison study. *JMIR Medical Education*, *9*(1), e48002. https://doi.org/10.2196/48002

Takii, K., Flanagan, B., Li, H., Yang, Y., Koike, K., & Ogata, H. (2024). Explainable ebook recommendation for extensive reading in K-12 EFL learning. *Research and Practice in Technology Enhanced Learning*, *20*, 027. https://doi.org/10.58459/rptel.2025.20027

Takii, K., Liang, C., & Ogata, H. (2024, November). Open knowledge and learner model: Mathematical representation and applications as learning support foundation in EFL. In Kashihara, A. et al. (Eds.), *Proceedings Vol. 1 of the 32nd International Conference on Computers in Education* (pp. 595–604). Asia-Pacific Society for Computers in Education. https://doi.org/10.58459/icce.2024.4902

Taylor, W. L., 4th, Cheng, R., Weinblatt, A. I., Bergstein, V., & Long, W. J. (2024). An artificial intelligence chatbot is an accurate and useful online patient resource prior to total knee arthroplasty. *The Journal of Arthroplasty*, *39*(8S1), S358–S362. https://doi.org/10.1016/j.arth.2024.02.005

Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, *29*(2), 217–257. https://doi.org/10.1007/s40593-019-00174-2

Vakulenko, S., Longpre, S., Tu, Z., & Anantha, R. (2021, March). Question rewriting for conversational question answering. In *Proceedings of the 14th ACM international conference on web search and data mining* (pp. 355–363). ACM. https://doi.org/10.1145/3437963.3441748

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022, November). Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems* (pp. 24824–24837). ACM. https://dl.acm.org/doi/10.5555/3600270.3602070

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, *55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Zhai, C. (2024, July). Large language models and future of information retrieval: Opportunities and challenges. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 481–490). ACM. https://doi.org/10.1145/3626772.3657848

Zhang, C., Hofmann, F., Plößl, L., & Gläser-Zikuda, M. (2024). Classification of reflective writing: A comparative analysis with shallow machine learning and pre-trained language models. *Education and Information Technologies*, *29*(16), 21593–21619. https://doi.org/10.1007/s10639-024-12720-0

Zhu, Y., Wang, X., Chen, J., Qiao, S., Ou, Y., Yao, Y., Deng, S., Chen, H., & Zhang, N. (2024). LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *World Wide Web*, *27*(5), 1–23. https://doi.org/10.1007/s11280-024-01297-w

Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Chen, H., Liu, Z., Dou, Z., & Wen, J.-R. (2023). *Large language models for information retrieval: A survey*. arXiv. http://arxiv.org/abs/2308.07107

## Publisher's Note

The Asia-Pacific Society for Computers in Education (APSCE) remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.