

LitLinker: Supporting the Ideation of Interdisciplinary Contexts with Large Language Models for Teaching Literature in Elementary Schools

Haoxiang Fan
School of Artificial Intelligence
Sun Yat-sen University
Zhuhai, China
fanhx6@mail2.sysu.edu.cn

Changshuang Zhou*
Faculty of Education
University of Macau
Macau, Macao
mc34188@um.edu.mo

Hao Yu*
School of Artificial Intelligence
Sun Yat-sen University
Zhuhai, China
yuhao53@mail2.sysu.edu.cn

Xueyang Wu
NeurlStar
Shenzhen, China
xwuba@connect.ust.hk

Jiangyu Gu
Xiangzhou Experimental School of
Zhuhai
Zhuhai, China
3299158551@qq.com

Zhenhui Peng[†]
School of Artificial Intelligence
Sun Yat-sen University
Zhuhai, China
pengzh29@mail.sysu.edu.cn

Abstract

Teaching literature under interdisciplinary (*e.g.*, science, art) contexts that connect reading materials has become popular in elementary schools. However, constructing such contexts is challenging as it requires teachers to explore substantial amounts of interdisciplinary content and link it to the reading materials. In this paper, we develop *LitLinker* via an iterative design process involving 13 teachers to facilitate the ideation of interdisciplinary contexts for teaching literature. Powered by a large language model (LLM), *LitLinker* can recommend interdisciplinary topics and contextualize them with literary elements (*e.g.*, paragraphs, viewpoints) in the reading materials. A within-subjects study (N=16) shows that compared to an LLM chatbot, *LitLinker* can improve the integration depth of different subjects and reduce workload in this ideation task. Expert interviews (N=9) also demonstrate *LitLinker*'s usefulness for supporting the ideation of interdisciplinary contexts for teaching literature. We conclude with concerns and design considerations for supporting interdisciplinary teaching with LLMs.

CCS Concepts

• **Human-centered computing** → **Interactive systems and tools; Empirical studies in HCI.**

Keywords

Interdisciplinary contexts, ideation, elementary schools, teachers, large language models

*Both authors contributed equally to this research.

[†]Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1394-1/25/04

<https://doi.org/10.1145/3706598.3714111>

ACM Reference Format:

Haoxiang Fan, Changshuang Zhou, Hao Yu, Xueyang Wu, Jiangyu Gu, and Zhenhui Peng. 2025. LitLinker: Supporting the Ideation of Interdisciplinary Contexts with Large Language Models for Teaching Literature in Elementary Schools. In *CHI Conference on Human Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3706598.3714111>

1 Introduction

The integration of knowledge from different disciplines within literature instruction at the elementary school level has been proven to enhance student learning outcomes [34, 44]. Contextualization, as a pedagogical approach, makes the curriculum more meaningful and practical for students, which is essential for effective learning [18]. Learning in an interdisciplinary context not only enhances students' comprehensive learning abilities but also increases their motivation to learn. However, the ideation of a suitable interdisciplinary context is challenging. Firstly, human knowledge is categorized into distinct disciplines, and people with an occupation usually focus their expertise on a single field [45]. The development of an interdisciplinary context requires literature teachers to engage in the continuous exploration and evaluation of substantial amounts of information from unfamiliar fields (*e.g.*, mathematics, science, and art). Secondly, elementary school teachers frequently face constraints related to limited time outside of the classroom [13], which do not allow them to conduct comprehensive searches for interdisciplinary knowledge.

Recent advances in large language models (LLMs) show great potential to address the challenges faced by teachers in ideating interdisciplinary contexts. Such as GPT-4¹ and GLM-4², possess extensive knowledge across various domains, owing to their pre-training on large text datasets [7]. Also, LLMs' capabilities in long-context understanding [59] enable them to synthesize information from extensive textual materials effectively. Within the field of human-computer interaction (HCI), the LLM-empowered tools have emerged to facilitate interdisciplinary information exploration [68],

¹<https://openai.com/index/gpt-4>

²<https://github.com/THUDM/GLM-4>

lesson plan preparation based on educational theories [17], and support learners in critical and creative thinking [51, 65]. However, neither conversational applications of LLMs (e.g., ChatGPT) nor these LLM-powered tools can offer a comprehensive solution for the challenges at hand. Firstly, the process of ideating interdisciplinary contexts from a variety of reading materials necessitates that educators engage in a thorough exploration of both the context and the materials. This engagement enables teachers to discern the connections between the context and the reading materials, rather than reading the responses generated by LLMs. Without such an in-depth exploration, fostering a comprehensive understanding of the context and materials becomes challenging, which is essential for effective teaching. Secondly, the suggested interdisciplinary contexts need to be effectively integrated into the classroom environment. However, due to a limited understanding of teaching practices and the cognitive backgrounds of elementary students, LLM-generated content often fails to align with the cognitive and pedagogical demands of elementary education [17]. Consequently, the generated content may require adjustments by literature teachers or may even be unusable. Lastly, educators, particularly those who are not so familiar with LLMs, have to allocate significant time and effort in constructing complex prompts and understanding the intricate outputs generated by these systems. In summary, there is a need for an interactive tool that assists educators in effectively exploring and generating ideas for interdisciplinary contexts within the classroom.

In this paper, we focus on the teaching scenarios of Chinese language courses in China and introduce *LitLinker*³, an LLM-powered interactive system that supports teachers in ideating diverse interdisciplinary contexts for teaching literature in elementary schools. To develop *LitLinker*, we follow a user-centered design approach that involves in total 13 Chinese literature teachers in three interview sessions, prototyping, and evaluation. *LitLinker* is a web application powered by the LLM GLM-4. When users select relevant disciplines and reading materials for exploration, *LitLinker* initially recommends interdisciplinary topics and corresponding analyses based on the themes and concepts of the selected materials. Subsequently, *LitLinker* facilitates the exploration of relationships between various literary elements (e.g., paragraphs, sentences, and viewpoints within reading materials) and the selected topics. Users can bookmark topics they focused on and ask the LLM any questions related to the topics and reading materials. Finally, *LitLinker* produces a lesson plan that includes a course outline, an introductory overview, and recommended classroom activities.

We evaluate *LitLinker* through two studies that answer three research questions (RQs), i.e., how would *LitLinker* affect the interdisciplinary context exploration (RQ1) outcomes and (RQ2) process, and (RQ3) how would users perceive the usability and creative support of *LitLinker*. Experiment I is a within-subjects design that quantitatively answers these three RQs with 16 novices of Chinese literature teaching, while Experiment II is a qualitative study that gains answers to the RQs from 10 elementary Chinese literature teachers with varying teaching experience. The quantitative results indicate that *LitLinker* leads to deeper integration of different subjects in the outcome lesson plan, is deemed significantly more

satisfying and efficient, and significantly reduces task workload. The qualitative analysis reveals high satisfaction with the generated results and a strong appreciation for *LitLinker* in facilitating teachers to develop interdisciplinary courses.

In summary, this paper has three main contributions. First, we introduce *LitLinker*, an interactive system that supports the ideation of interdisciplinary contexts for literature classes. Second, through a within-subjects study and expert interviews, we provide empirical evidence of *LitLinker*'s effectiveness and usefulness. Third, based on our design process and findings, we offer design implications for future LLM-based systems that assist teachers.

In the rest of this paper, we first review literature that inspires our work (Section 2), followed by the design process (Section 3) and implementation (Section 4) of *LitLinker*. Then, we present two experiments (Section 5, Section 6) that evaluates *LitLinker*. We finally discuss the implications of our work to HCI and education communities (Section 7).

2 Related Work

2.1 Interdisciplinary and Literature Instruction

Interdisciplinary education equips students to think across subject boundaries, addressing complex global challenges and enhancing critical thinking, problem-solving, and collaboration skills [48]. Its growing importance is evident in fostering adaptable, innovative learners for a rapidly changing world [64]. By applying language in real-world contexts, disciplinary approaches promote engagement [56]. Interdisciplinary education offers key advantages in K-12 language learning, promoting critical thinking, creativity, and motivation by connecting learning to real-world issues [2, 23]. Empirical studies often focus on STEM fields, such as English [15]'s integration approach and Sarama et al. [50]'s C4L curriculum, which combines language learning with early childhood development. Also, Czerniak and Johnson [12] demonstrated how interdisciplinary science teaching enhances language development.

With the introduction of the new curriculum standards by China's Ministry of Education in 2022, interdisciplinary curriculum design was formally included in the "expansive learning task group" of Chinese language education, explicitly outlining requirements for interdisciplinary teaching⁴. Guided by these new standards, interdisciplinary teaching in primary school Chinese language classes can bring a fresh perspective to Chinese language education. By the design of diverse interdisciplinary activities, course content is integrated to cultivate students' active learning and inquiry skills. Additionally, reading instruction fosters students' interdisciplinary thinking abilities, promotes cross-subject integration, and stimulates students' imagination. However, few works focus on integrating literature with other subjects and customizing technologies for diverse teaching environments. Teachers could face challenges in preparing for interdisciplinary literature teaching due to inadequate professional development [3, 61]. Our work is motivated by the trend and benefits of interdisciplinary literature teaching and aims to help teachers ideate interdisciplinary contexts for teaching literature in elementary schools.

³It is open-sourced and available on <https://github.com/fanhaoxiang1/LitLinker>

⁴<http://www.moe.gov.cn/srcsite/A26/s8001/202204/W020220420582344386456.pdf>

2.2 Interactive Systems that Support Brainstorming and Ideation

Interactive systems have been widely used to facilitate the creative thinking processes, including art design [20, 24], writing [21], and learning [29]. Creativity Support Tools (CSTs) provide numerous advantages due to their interactive nature. For instance, C2Ideas [27], an innovative system for designers to ideate color schemes, provides users with an interactive workflow that aligns with traditional interior design methods to facilitate the design process. Users are required to input their initial design intentions and customize the intermediate results, after which the system generates good outcomes. Furthermore, in the field of supporting creative activities within interdisciplinary environments, DiscipLink [68] assists users in making sense of information by generating exploratory questions, expanding queries, and extracting themes and connections among academic papers. Users can freely explore the carefully designed Orientation View, Exploration View, and Collection View of the user interface, a layout that we also use.

More and more creativity support interactive systems are being designed for specific tasks in particular professions [10, 33, 37, 47]. For example, AngleKindling [47] is designed to assist journalists in exploring diverse angles for reporting on press releases. This work is similar to ours, as the creativity support also comes from analyzing large amounts of textual material, and our work also leverages the capabilities of LLMs. There also has been a growing trend toward the integration of human-computer interaction (HCI) and education. This integration has led to the development of interactive systems aimed at enhancing educational outcomes by promoting users' creative thinking [66]. NaCava [60] is a mobile interactive system designed to facilitate nature-inspired creativity for children. It enhances children's multidimensional observation and engagement with nature by encouraging the collection of multimodal materials and using them in a creation process.

However, few studies focus on developing creativity support tools for teachers to help them design classroom activities to promote student learning, especially in interdisciplinary setups. In our work, we create *LitLinker* aimed at teachers to help them ideate interdisciplinary contexts in education, offering creativity support to facilitate the ideation process.

2.3 Large Language Models in Education

Large Language Models (LLMs) are increasingly utilized across various educational domains due to their capacity to facilitate personalized learning experiences and enhance accessibility for all learners [40]. They support self-learning to enhance personal skills in diverse areas, including programming [63], problem-solving ability [30], and writing skills [52]. Additionally, they also support various activities within classroom environments, such as helping with homework, reviewing [40], and assisting students in project-based learning [67]. Specifically, within the K-12 educational domain, LLMs have demonstrated their effectiveness in providing assistance across multiple subjects, including foreign languages, programming, and mathematics [41]. This capability is attributed to their pre-training on large datasets across various fields [7]. Recent work also suggested that designing multiple LLM agents to

simulate human collaboration can enhance the quality of the outputs on domain-specific tasks, such as translation [58] and software development [14]. Inspired by these works, we leverage the potential of LLMs and deploy them as agents to generate precise and comprehensive interdisciplinary information in our study.

LLMs also promote instructional activities for teachers. For instance, RetLLM-E [39] provides instructional support by delivering context-aware, high-quality answers to student questions using LLMs. This research illustrates that retrieved context can significantly increase the quality of LLM-generated responses, thereby informing the advantages of utilizing retrieval-augmented generation [32] for precise instruction. Furthermore, LessonPlanner [17], an LLM-driven tool that formulates structured lesson plans based on educational theories, has been empirically validated to assist teachers in increasing the efficiency of lesson preparation and the quality of lesson plans. The objectives of our work are similar to theirs, as both emphasize the quality of outcomes and the user experience.

Previous studies have demonstrated the potential of LLMs in education, but few works focus on teachers' preparation for interdisciplinary instruction scenarios. In our work, we conduct an iterative design with 13 teachers to understand their views on how LLMs can facilitate interdisciplinary literature instruction and support them.

3 Design Process and Principles of *LitLinker*

Our work aims to support elementary literature teachers in effectively identifying suitable interdisciplinary contexts for their instructions, which can be used in their later lesson plans and classroom activity designs. Our design process and evaluation of *LitLinker* involve in total of 17 Chinese language teachers (Table 1) in an elementary school in mainland China. Specifically, in the design process, we involved E1-E7 in the foundational study and I1-I6 in the evaluation of the prototype. In the evaluation of *LitLinker* with teachers (*i.e.*, Experiment II), we involved E1-E5 again and E8-11. In Experiment II, I1-I6 also contributed findings about the unchanged features between *LitLinker* and its prototype.

3.1 Design Process

We generally followed a user-centered approach to plan our design process. First, to understand users' practices and involve them in the design of *LitLinker*, we conducted three sessions of semi-structured interviews with an experienced literature teacher E1, who led a seven-member (E1 - E7) interdisciplinary literature course design team that indirectly contributed to the interviews. We were not able to have direct discussions with E2 - E7 due to their inconvenience during the semester. Then, we developed a workable prototype of *LitLinker* and evaluated it with another six teachers (I1 - I6). We gathered their insights to inform our design goals for *LitLinker* presented in this paper.

3.1.1 Foundational Study. We closely worked with E1 to identify the practices, challenges and needs for support of ideating interdisciplinary contexts for teaching literature in elementary schools. Over the past two years, E1 has spearheaded a team of seven individuals (E1-E7) in the exploration and implementation of interdisciplinary literature instruction within elementary school Chinese courses.

Table 1: 17 Chinese language teachers participated in the iterative design process and expert interviews (i.e., Experiment II). Among them, there were 6 males and 11 females, with teaching experience ranging from 3 to 29 years. Two participants did not provide information on their teaching experience. This table also includes their experience in reading projects and interdisciplinary projects.

Involvement	ID	Gender	Teaching Experience (years)	Participation in Reading Projects	Participation in Interdisciplinary Projects
<i>Foundational Study & Experiment II</i>	E1	F	5	Y	Y
	E2	M	7	Y	Y
	E3	F	14	Y	Y
	E4	M	6	Y	Y
	E5	M	4	Y	Y
<i>Foundational Study</i>	E6	F	-	Y	Y
	E7	F	-	Y	Y
<i>Evaluation of Prototype (re-usable findings are presented in Experiment II)</i>	I1	F	27	Y	Y
	I2	F	8	Y	Y
	I3	F	5	N	N
	I4	M	11	Y	Y
	I5	F	5	Y	N
	I6	M	5	Y	N
<i>Expert Interviews</i>	E8	M	3	N	N
	E9	F	6	Y	N
	E10	F	29	N	Y
	E11	F	17	Y	N

To gain a comprehensive understanding of user needs, we progressively conducted three sessions of semi-structured interviews with E1 in April, June, and July 2024, lasting 38 minutes, 53 minutes, and 45 minutes, respectively. Before each session, we communicated the purpose of the interview to E1 and requested that she engage with her team to compile records of their meetings for discussing the topics in the intended interview. We documented each interview session with E1 through audio and video recordings.

In Session 1, we asked E1 to present their current practices of teaching literature in interdisciplinary contexts, with previously developed lesson plans and assignments in her team. The discussion also focused on the potential of AI (e.g., what do you think AI can support you (in your lesson planning in previous)), and an interactive system to facilitate the design of interdisciplinary literature contexts, including the integration of art and history into the assignments of literature reading. After this session, two authors brainstormed potential features of a system for supporting the ideation of interdisciplinary contexts and sent E1 a document that explains these features. We requested E1 to engage in a discussion with her group members to identify any additional or incorrect points about the potential system. In Session 2, E1 came back with positive feedback from her team on each potential feature. We asked her to further explain their general process for designing interdisciplinary contexts as a team, emphasizing the distinct responsibilities and cognitive processes of each teacher involved. Additionally, she presented the proposed interaction model. After this session, E1 had a group meeting with her team and came up with a template that defines the anticipated outcomes of our system. In Session 3, we introduced how a system works utilizing LLM agents to simulate roles in a team for ideating the interdisciplinary contexts, as

suggested by E1 in Session 2. We presented two example outcomes produced using our predefined prompts and intermediate outputs to ask for her opinions on these prompts and outputs (e.g., whether these intermediate outputs were helpful? If the prompts aligned well with your thoughts?).

3.1.2 Development and Evaluation of LitLinker Prototype. After Session 3, two of the authors utilized the thematic analysis method to analyze the transcribed recordings and all textual content derived from the foundational study. The analysis yielded four summarized Design Principles as described in the following Section 3.3. We then worked on the implementation of a workable prototype that chains different LLM agents in a structured process to help teachers think of interdisciplinary contexts for teaching the literature materials in the textbooks. We evaluated our workable prototype with another six elementary school Chinese language teachers (I1 - I6, 3 Male, 3 Female), as shown in Table 1. Each evaluation lasted approximately 30 to 45 minutes and comprised four parts: (1) an introduction to the background, which included the concepts of interdisciplinary literature instruction, and the theory of contexts of instruction; (2) a brief tutorial on the prototype; (3) a think-aloud study in which participants freely explored the prototype and spoke out their thoughts; and (4) a semi-structured interview for their comments and suggestions on the prototype. With the participants' consent, we conducted the evaluation offline and recorded audio and video.

At this stage, we assessed the prototype's usability (e.g., whether different functions were well-integrated), user perception (e.g., user interaction with the prototype and any additional cognitive load), and the quality of the system's outcomes. Furthermore, we collected suggestions for improving the prototype, particularly regarding user interface design and additional functionalities. These feedback

and suggestions are presented in Section 3.2, which inform the design principles (Section 3.3) of *LitLinker*. For the six teachers' (I1-6) feedback on the same features in the prototype and final version of *LitLinker*, we incorporate it in the results of Expert Interviews in Section 6.

3.2 Findings

Two of the authors utilized thematic analysis [5] to inductively code and summarize the practices, challenges, requirements, and concerns from transcribed recordings in the design process. The key themes are shown below. One author first iteratively coded the data, while the other carefully reviewed the codes to ensure accuracy. After discussions, they reached a consensus and identified six primary themes. These findings are shown below.

Finding 1: In practice, teachers usually engage in reverse thinking when ideating interdisciplinary context. Our teachers mentioned two intellectual paradigms. One paradigm referred to as “forward thinking”, resembles “deductive reasoning”, in which teachers create abstract connections from a limited number of reading materials (typically 3-5 texts) and develop a concrete and reasonable context. The other paradigm, which is termed “reverse thinking” and analogous to “inductive reasoning”, is a more habitual cognitive process employed by teachers. Teachers would like to first select interdisciplinary contexts that they deem suitable and then identify appropriate texts from a broader text pool, after which they refine the connections between the identified text and the context. *“For us, a good context often arises from a sudden inspiration, which we then backtrack to complete the ideation of what texts can connect to this context and how”* (E1).

Finding 2: It is challenging to identify the connections between the established context and the reading materials in the process of “reverse thinking”. Teachers must evaluate the effectiveness of the connections in enhancing students’ understanding of both literature and its associated subjects, as well as in stimulating their interest. *“It is quite difficult and usually takes a long time for our team to ensure that based on the reading materials, our teaching activities connected by the context can indeed help students gain knowledge”* (E1).

Finding 3: Teachers require support at three levels of granularity when analyzing the reading materials and contexts. The first level is in-depth single-text analysis, which *“explains how the elements of a given article relate to the context”* (E1). The second level is pairwise comparison, where comparative reading has been demonstrated to be an effective method for understanding texts, *“allowing articles to ‘disappear in pairs’ by analyzing the similarities and differences in relation to the context”* (E1). The third level is multi-text-driven exploration, which necessitates that the system should support the comprehensive linking of all texts selected by the teacher. Therefore, this level requires *LitLinker* to conduct a thorough deconstruction of contexts, extract meaningful connections, and convey these connections to the teachers.

Finding 4: Teachers require detailed instructional activities based on the selected contexts. As E1 summarized after the meeting with her teaching team, the outcome plan of several lessons surrounding a context should include targeted reading materials and analysis in each lesson, an introduction facilitating students’

engagement, and related teaching activities. Also, in the evaluation study of *LitLinker* prototype, three teachers (I1, I2, I5) indicated that the system outputs should be more detailed and reduce human effort in modifying them for the later concrete plans for each lesson. *“The overall structure of the output is good, but I hope it can be more detailed; for example, providing more in-class and extracurricular activities related to the theme, so we can use them directly”* (I1). Therefore, we incorporated recommendations for literature and interdisciplinary course activities in the refined *LitLinker*.

Finding 5: Teachers are concerned the quality and reliability of the content purely generated by LLMs. In the evaluation study with six teachers (I1-6), our prototype generated traditional subject-related contexts using the LLM with specific templates and cognitive backgrounds of elementary students, without fine-tuning or retrieval-augmented generation (RAG). Three teachers (I4, I5, I6) expressed concerns about the quality of the LLM-generated content. *“The content generated for the art subject is quite repetitive”* (I5). *“We need to establish a dedicated article database for science as well, since many of our articles are highly relevant to science”* (I4).

Finding 6: Teachers suggest six metrics for evaluating the outcome of interdisciplinary literature lesson plan. As established by the team of E1-E7, the metrics are:

- **Appropriateness of Context**
 - *Content Alignment*: Does the context accurately cover the content of the selected materials?
 - *Internal Logic*: Is there a logical connection between the context and the selected materials?
- **Alignment with Educational Objectives**
 - *Curriculum Standards*: Does the content comply with national curriculum standards and teaching guidelines?
 - *Subject Goals*: Does it help achieve specific goals of language education, such as reading comprehension and writing skills?
- **Depth of Integration**
 - *Subject Integration*: Does it effectively integrate knowledge from different subjects?
 - *Knowledge Transfer*: Does it promote the application of language arts knowledge in other subject contexts?

3.3 Design Principles

Based on the findings from our design process and related literature, we derive four design principles of *LitLinker*.

DP1: *LitLinker* should provide step-by-step support that aligns with teachers’ habitual practices in interdisciplinary context ideation. Tailoring the assistance to users’ habitual practices (e.g., active students’ behaviors or teachers’ behaviors) is a commonly enacted principle in previous interactive systems in educational scenarios [17, 19, 36]. In the task of interdisciplinary context ideation, as revealed in Finding 1, *LitLinker* should support step-by-step context ideation through reverse thinking, which is a habitual practice of our teachers. In this practice, as E1 shared, teachers take various roles to analyze the potential contexts (we note this role as Context Analyst), analyze the texts in the reading materials, connect them to the contexts, and discuss the approaches (Text Analyst and Text Reviewer). After that, teachers try to summarize the contexts and associated reading materials into an actionable

lesson plan (Context Summarizer). *LitLinker* can prompt LLMs to play different roles when supporting teachers in each of these steps. We do not chase for generating one-step context ideation outcome with one LLM prompt, because teachers desire necessary human input in each step, and enabling multiple LLM agents to simulate human-human collaboration has been proven to improve output quality [14, 58].

DP2: *LitLinker* should provide teachers with detailed analyses of the contexts, reading materials, and their relationships. Our teachers reported that it was challenging to identify the connections between contexts and reading materials (Finding 2) and desired support during the analyses (Finding 3). Previous HCI works have demonstrated the strengths of LLMs in analyzing and connecting complex information [35, 68]. Similarly, to satisfy user requirements (Finding 3) in our task, *LitLinker* can leverage LLMs to recommend contexts, explain them in detail, identify relevant texts in the reading materials, and assess the relationship between the contexts and texts.

DP3: *LitLinker* should document the ideation outcomes in a lesson plan that aligns with the established educational practices in interdisciplinary literature teaching. A teacher without a lesson plan may struggle to effectively deliver the knowledge and objectives of the lesson [28]. Finding 4 suggests that the lesson plan should contain detailed instructional activities, including the related reading materials and in-class activities, based on the selected contexts. To make it further aligned with educational practices, *LitLinker* can adopt the six evaluation metrics of the outcome lesson plan (Finding 6) to guide the generation of ideation outcomes.

DP4: *LitLinker* should include database of interdisciplinary contexts and reading materials and provide flexible user control to achieve high-quality ideation outcomes. Prior research on the impact of LLMs in primary education indicates that generating false content is a disadvantage that may lead to “information pollution” for children [1, 42]. Finding 5 also indicates that LLMs sometimes were unable to create content that meets teachers’ needs when lacking access to educational resources. To generate high-quality outcomes, as inspired by previous works [31, 62], *LitLinker* could ground the content generation on diverse real-world contexts and reading materials. *LitLinker* should also support teachers to freely edit and question any content (e.g., texts in reading materials, outcome lesson plan) to make sure that they understand the content they are going to use in literature teaching.

4 Design of *LitLinker*

Based on the identified design principles, we introduce *LitLinker*, a human-AI collaborative system to support elementary school literature teachers in ideating interdisciplinary contexts. *LitLinker* not only facilitates teachers’ ability to explore various contexts and reading materials but also delivers structured outputs derived from this exploration process and the teachers’ thought processes. Inspired by prior research [68] and tailored to our interdisciplinary context exploration scenario, the primary interface of *LitLinker* consists of three parts: Contexts Exploration View, Texts Exploration View, and Collection View. After confirming the subjects and importing the reading materials, the system aligns with the user’s habitual

thinking process by selecting or specifying potentially related interdisciplinary contexts within Context Exploration View (DP1). Based on the selected context, the system enables multi-text-driven exploration (DP2). Specifically, *LitLinker* extracts connections between the context and literacy elements, displaying specific details in either Context Exploration View or Texts Exploration View, decided upon whether they are context-oriented or text-oriented. In Texts Exploration View, users can conduct in-depth analyses of individual texts and perform pairwise comparisons of any selected texts. In these two exploration processes, users have the opportunity to chat with the LLM at any time and modify the analyses provided by the system (DP4). Users can add analyzed themes and corresponding texts into the collection, allowing the system to generate structured ideation outcomes based on their selections (DP3). In *LitLinker*, all generated content relies on a context database, prompt templates, and generation guidelines (DP3, DP4). In the following subsections, we will describe how users can interact with *LitLinker* and the detailed design and implementation of *LitLinker*.

4.1 *LitLinker* Interface and Interaction Design

To initiate the exploration process, users are required to select the interdisciplinary subjects and reading materials they wish to explore (Figure 1 A). By clicking the “Collapse Config Panel” button (Figure 1 A turned into “Open config panel” at the top of the page after clicking on), users confirm their settings, which subsequently expands Contexts Exploration View. Upon clicking the “Generate Recommended Contexts” button, *LitLinker* displays a series of context cards (B), each containing the context title, the titles of the most relevant reading materials, the original context from Context Pool (see Section 4.2.2), and a detailed description of the context. Users may click the “Find” button on any context card of interest to further explore and edit the description of the context, allowing them to ask the LLM any questions (D) related to the theme and its connections (e.g., relate this context to additional musical works?) to aid in understanding and exploration. The description becomes editable, enabling users to reconstruct it based on the LLM’s responses. Users can also click “Star” to favorite a context and click the “Right Arrow” to specify the theme associated with texts in Texts Exploration View (B). If users are dissatisfied with a context, they have the option to click the “Delete” button to remove it. Users can click “Manually Add Context” to provide the title and background of a context in a pop-up window (C), allowing *LitLinker* to tailor a suitable context based on the information provided.

The middle part of the page is designated as Texts Exploration View (Figure 1), which facilitates a comprehensive examination of the interrelationships among various reading materials and specified contexts. Texts that are most relevant to the selected context are displayed at the top in Texts Exploration View, with each reading material accompanied by the analysis. This analysis encompasses the relationships among sentences, paragraphs, and the context (i.e., in-depth single-text analysis). To further support user exploration, inspired by [58] and the workflow of the teachers, additional ratings and critical recommendations are provided for LLM-generated analysis. Similar to Contexts View, users have the option to click the “Find” button on any card (E) to ask the LLM any questions about the contexts related to the reading material (e.g., Can you

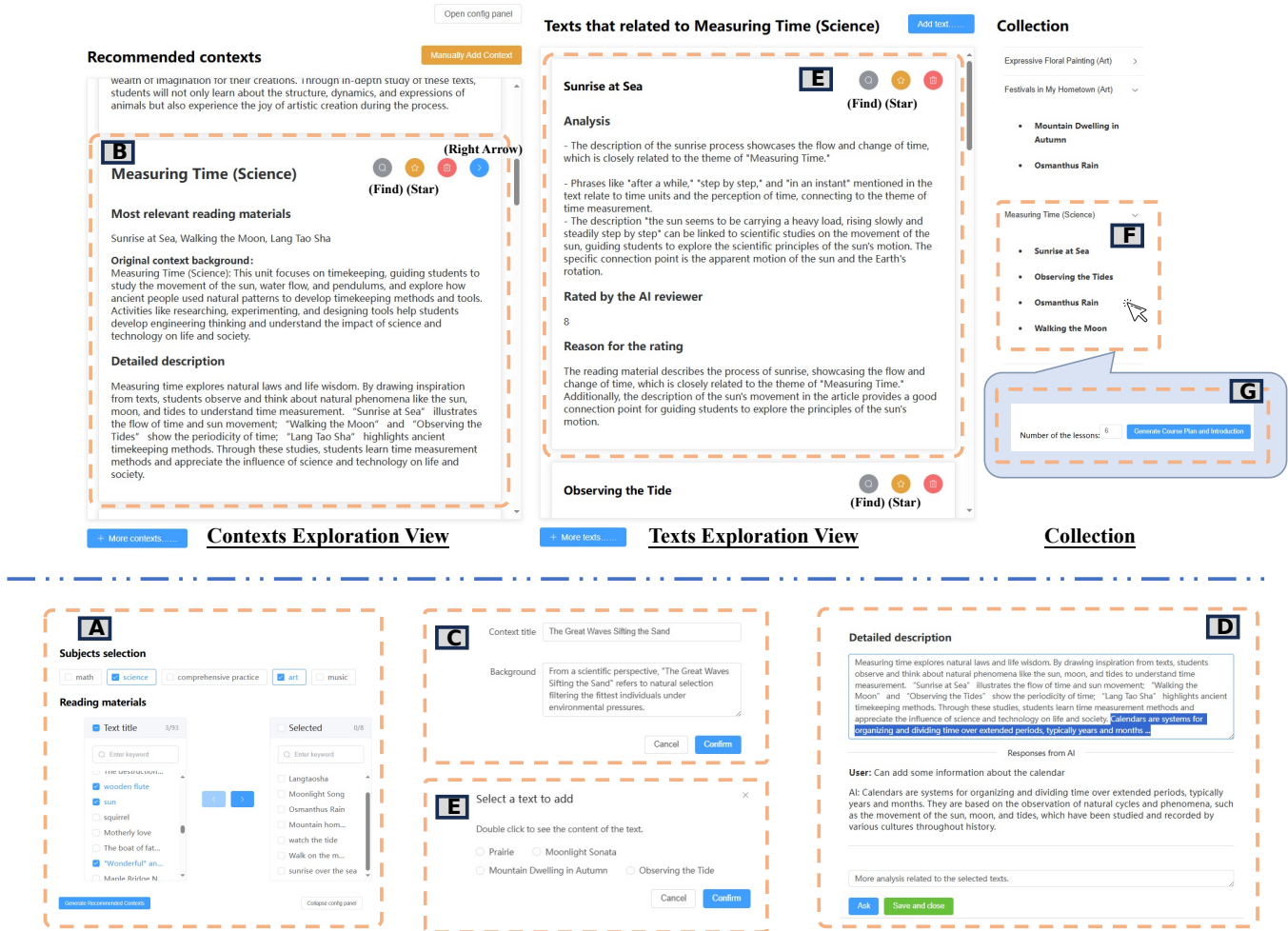


Figure 1: The user interface of *LitLinker*, translated from Chinese either by Google Translator or manually.

elaborate on the relationship between the descriptions of scenery in the text and the context?), and users can also modify the analysis. Users can click “Star” to favorite a reading material under the relevant context or click “Delete” if they no longer consider the text. If a user believes a reading material fits the context but was not recommended, they have the option to manually click “Add Text” to request the system to analyze it based on the context. To optimize generation speed, the system generates a limited number of contexts and analyzes a limited number of reading materials at a time (set as 8 during experiments). Users can click “More Contexts” and “More Contexts” to generate additional contexts and analyze more texts for further exploration.

When clicking a context within Collection View (Figure 1), users are directed to the Outcome Generation Panel (Figure 2). In this panel, users can examine the details of their chosen context as well as the analyses of the associated reading materials. Subsequently, users are prompted to enter the “Expected number of lessons” (Figure 1 G) and select the “Generate Introduction and Course Plan”

(Figure 2) to create a comprehensive course plan that includes explanations for each segment and an introduction aligned with the course plan. In this interface, all content can be edited by double-clicking to adjust the final outcomes. Clicking “Generate the Activities for Classroom” recommends teaching activities related to the literature and the interdisciplinary subjects. Users can click on the titles to delete unnecessary activities. At this point, a complete outcome based on a specific context has been generated. Clicking the download button at the bottom right of the panel allows users to download the content in txt and HTML formats for sharing or further editing.

4.2 Implementation of the System

Inspired by [27, 58], *LitLinker* aligns its behavior with teachers’ habitual practices in interdisciplinary ideation (DP1). We designed an LLM-based multi-agent system to simulate human collaboration. Following the multi-agent collaborative framework by Hong et al. [26], we decompose the task into multiple steps, allowing agents to use tools when necessary and interact with the user for feedback.

Course Plan

Section 1: Dawn's First Light • Dusk

"Sunrise at Sea"

Lesson 1: Understanding the sunrise process.

Lesson 2: Exploring the sun's apparent motion and Earth's rotation.

Section 2: Tides and Moon Phases

"Observing the Tides"

Lesson 3: Analyzing tidal patterns and time measurement.

"Walking the Moon"

Lesson 4: Understanding moon phases and ancient time measurement.

Section 3: Seasonal Changes • Fragrant Paths

"Osmanthus Rain"

Lesson 5: Exploring ancient time measurement through weather prediction.

Lesson 6: Comprehensive discussion class

Combining "Sunrise at Sea", "Observing the Tides", "Osmanthus Rain", and "Walking the Moon"

Discussing natural phenomena and time measurement, emphasizing its importance in daily life.

Generate the activities for classroom

Activity list

Tide Simulation Game Create a game that simulates tidal changes, allowing students to role-play and understand the relationship between tides and time.

Typhoon Prediction Experiment Conduct a small experiment simulating weather changes, where students attempt to predict the arrival of a "typhoon," experiencing the ancient wisdom of time measurement. (Click to delete)

Poetry Creation: Ode to Time Encourage students to write a poem about time, reflecting on their feelings about the passage of time, based on the content learned in this unit.

Time Travel Storytelling Have students work in groups to create a story about time travel, inspired by the natural phenomena discussed in this unit, and share their stories with the class.

Number of the lessons: 6

Regenerate

Download HTML

Download TXT

Figure 2: The user interface of LitLinker's Outcome Generation Panel, translated from Chinese either by Google Translator or manually. The content in this figure serves as an example of outcomes.

Also, the prompt design follows the Co-Star framework [43] to optimize outputs. The architecture of our workflow is illustrated in Figure 3, which will be discussed in Section 4.2.1. The effectiveness of the content generated through this architecture relies on a comprehensive Context Pool (DP4) and carefully designed prompts (DP2, DP3), which will be introduced in Section 4.2.2 and Section 4.2.3, respectively.

4.2.1 Architecture of LitLinker. In line with the teachers' habitual practices and DPs, we simulate four roles (DP1) that participate in the ideation of interdisciplinary contexts: Context Analyst (DP2), Text Analyst (DP2), Text Reviewer (DP2, DP4), and Context Summarizer (DP3). To facilitate their collaboration, similar to MetaGPT [26], we enable these agents to remember, think (*i.e.*, determine subsequent actions based on observations), act (*i.e.*, utilize tools and communicate with one another), and, crucially, observe user behavior through the interface to respond appropriately. Each role is equipped with the GLM4 model as its "brain." Before the application is implemented, all original contexts and reading materials are embedded using the Zhipu embedding-3 model⁵. The texts and embedding arrays are stored in the database. For contexts imported by users, LitLinker will complete this embedding process in real-time.

Context Analyst acts as a "teacher with expertise across various disciplines." When a user clicks "Generate Recommended Contexts," it retrieves background information of the original contexts that

Introduction

Exploring the mysteries of time, we become travelers through the ages in this unit. From "Sunrise at Sea," we feel the passage of time, witnessing the majestic rise of the sun step by step. In the tranquil nights of "Walking the Moon," we follow the moon's path and experience the gentle flow of time. The spectacular rise and fall of tides in "Observing the Tides" help us understand the cycles and rhythms of time. Just like the wisdom ancient people encapsulated in "Lang Tao Sha," we learn to capture the imprints of time from the breathing of nature, uncovering the scientific secrets of time measurement. This knowledge not only connects the stars in the sky and the tides of the sea but also reflects the progress of technology and the wisdom of daily life. This is a journey through time and space, allowing our minds to wander in the boundless sea of time.

are most relevant to the reading materials from the database. For each context, Context Analyst conducts an analysis based on all associated reading materials and generates a detailed description of the context. The Context Analyst's memory continuously retains the texts and analyzes contexts, similar to the role of a teacher. When a user initiates a request in "Find" mode, Context Analyst is tasked with providing detailed information related to the context.

Text Analyst acts as an "active-minded literature teacher." When a user requests an analysis of texts based on the selected context, Context Analyst provides Text Analyst with the details of the selected context. Initially, Text Analyst retrieves the most relevant texts. Subsequently, Text Analyst conducts a thorough analysis of each text individually, producing an in-depth single-text analysis. Furthermore, Text Analyst responds to users' inquiries related to the text.

Text Reviewer acts as a "conservative and experienced literature teacher." When Text Analyst generates an analysis, it seeks Text Reviewer's opinion, which entails Text Reviewer assessing the relevance and accuracy of the analysis based on the reading material and the selected context, providing evaluative information to the user. This additional review process provides users with critical information and corrects potential inaccuracies in the RAG.

Context Summarizer acts as a "literature teacher skilled in summarization," responsible for integrating the context with synthesizing contextual information with selected reading materials from the collection. It delivers structured outputs step-by-step in

⁵<https://bigmodel.cn/dev/api/vector/embedding-3>

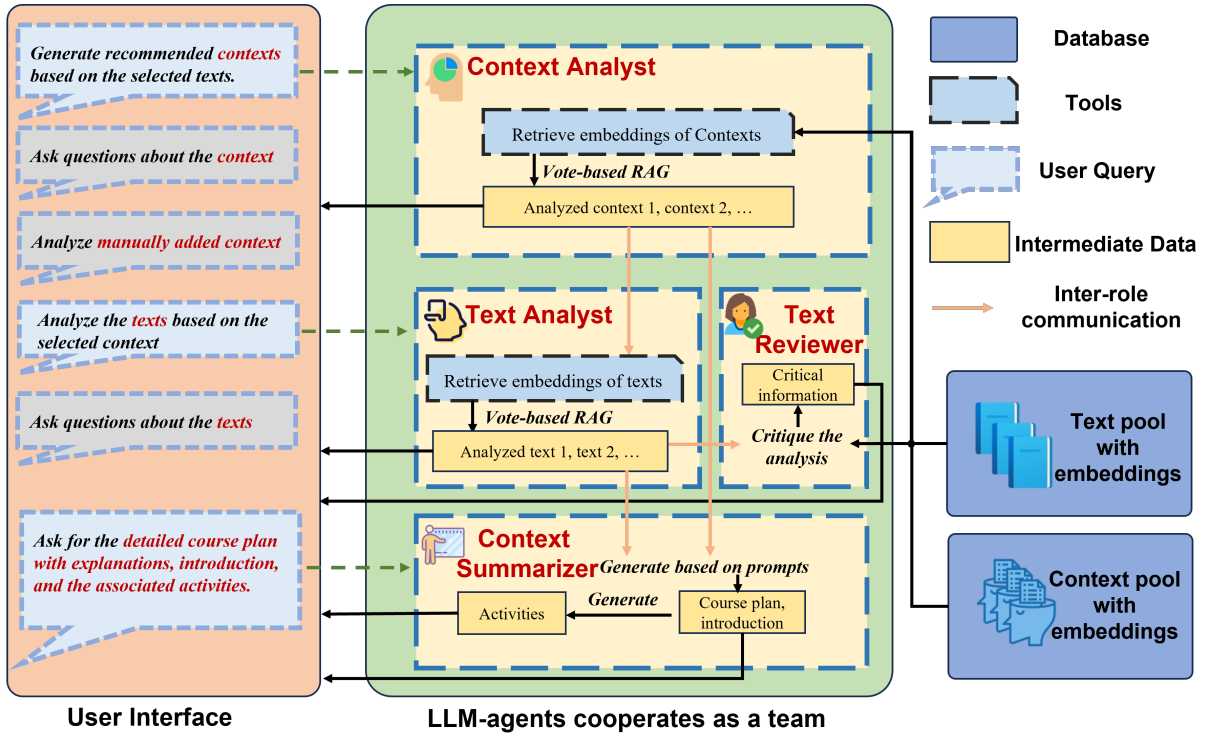


Figure 3: The architecture of LitLinker. The illustration of the main workflow of LitLinker. The responses to user’s queries with gray background are not shown in this figure.

the Outcome Generation Panel (*i.e.*, a detailed lesson plan with explanations for each segment, the introduction based on the lesson plan, and the associated activities).

In summary, various user actions within the interface trigger the actions of different roles. Upon reflection, these roles employ a range of tools and capabilities, which depend on carefully designed prompts, to accomplish their respective tasks.

4.2.2 Context Pool. To enhance the generation of reasonable and evidence-based contexts, we compiled a database of original contexts accompanied by background information across various subjects (including general contexts).

The Informal Education Database contains 113 original contexts derived from a famous Chinese journal Yao Wen Jiao Zi’s “Annual Popular Words (2008-2023)”⁶. These materials claim to be based on data from the National Language Resources Corpus, with additional content selected manually. One author read the 160 original contexts and included those comprehensible to elementary school students in the database. The context database for the subjects of art, science, mathematics, and music is extracted from the unit titles and descriptions of widely used textbooks for grades 3-6, including 144 original texts. These contexts were selected to fulfill teachers’ requirements for accuracy and diversity, providing a positive experience during the user study. Notably, teachers can easily add contexts in a batch; they only need to provide the subject, title, and

background information, similar to the operations performed on the interface.

4.2.3 Iterative Prompt Engineering. LitLinker aims to provide users with high-quality, structured outcomes based on specific evaluation criteria. We have incorporated valuable insights from experts gathered during the foundational study into the prompt design. The prompts are structured in accordance with the co-star [43] framework, which encompasses context, objective, style, tone, audience, and response in a single prompt. The style, tone, and audience specifications for each role are shown in the supplementary materials. The design details of the context, objective, and response are shown below.

The **context part** of the prompt not only explains the task (*e.g.*, find multiple points related to the context across various texts and generate a highly specific analysis based on the context) but also provides all the requisite knowledge for the task. Although the GLM-4 supports a 128k context length, overly long prompts may lead the LLM to neglect certain information. Therefore, we precisely designed the minimum information necessary for each role’s actions. For instance, Text Analyst is provided solely with the content of the text to be analyzed and the relevant context for each request, and this procedure is repeated for multiple texts to ensure the most precise analysis for each individual text. Meanwhile, Context Summarizer concentrates on generating relevant activities in accordance with the course plan, incorporating only the course

⁶<http://yaowenjiaozi.cn/>

plan and introduction into the prompt, without the content of the reading materials.

The **objectives part** in all prompts are consistently aligned with the six metrics established by the team (DP3) for evaluating the quality of contexts and outcomes. These clearly defined objective statements ensure that the outputs are in close accordance with the expectations of teachers. The response, which specifies the required output format and content, is embedded with templates developed by experts during the foundational study. Following the evaluation of the prototype, we provided feedback to experts E1-E6 and revised the templates to enhance their content richness. The template for Context Summarizer for generating a course plan is as follows:

Segment 1: Initial Encounter with the Scene · Uncontrollable Emotions

- “Prairie” [the title of the text] (The natural scenery and ethnic friendship in the prairie)
 - Lesson 1: Appreciate the beautiful sentences in the text and feel the beauty of the prairie.
 - Lesson 2: Feel the enthusiasm of the prairie people and the deep friendship between Mongolian and Han people.

Segment 2: Encounter with the Scene Again · Touching the Heartstrings

- “Lilac Knot” [the title of the text] (The feelings in the lilac knot)
 - Lesson 3: Understand the author’s way of expressing associations triggered by objects and explain your understanding.
 - Lesson 4: Understand the symbolic meaning of the lilac knot and appreciate the emotions embedded by the author.
- “Lodging by the River” [the title of the text] + “Visiting an Old Friend” [the title of the text] (The feelings in the mountains and rivers under the moonlight)
 - Lesson 5: Use “Lodging by the River” as an example to teach the method of learning ancient poetry and explore the imagery of the “moon.”
 - Lesson 6: Imagine the scenes in “Lodging by the River” and “Visiting an Old Friend” and understand the unique poetic feelings evoked by the “mountains and rivers.”

Segment 3: Another Encounter with the Scene · Endless Thoughts

- “Song of Flowers” [the title of the text] (Unity of man and nature, understanding philosophy)
 - Lesson 7: Feel the attitude towards life and inner ideals after transforming into a “flower,” and appreciate the author’s positive attitude towards life.

The template includes all the elements expected by teachers during the foundational study. The LLM can utilize the template to verify that the response includes all necessary content.

5 Experiment I: Within-subject study

We evaluate *LitLinker* via two experiments. Experiment I is a within-subjects study that aims at assessing the effectiveness of *LitLinker* in supporting novices of literature teaching to come up with interdisciplinary contexts. Experiment II is a qualitative study that examines *LitLinker*’s values in assisting expert Chinese language teachers in ideating interdisciplinary contexts for literature teaching. In this section, we present the design and results of Experiment I. The research questions are:

- RQ1. How would *LitLinker* affect the interdisciplinary exploration outcomes?
- RQ2. How would *LitLinker* affect the process of interdisciplinary exploration?
- RQ3. How would users perceive the usability and creative support of *LitLinker*?

5.1 Participants

We recruited 16 students (P1-P16, nine females, seven males; age: $Mean=23.44$, $SD=2.24$) via a post in a group chat and via word-of-mouth from two universities. This sample size follows the practices of previous work that evaluates systems for supporting ideation and education, such as CreativeConnect ($N = 16$) [10], AngleKindling ($N = 12$) [47], DiaryHelper ($N = 12$) [33], and LessonPlanner ($N = 12$) [17]. We also calculated the required sample size with G* Power to conduct Wilcoxon Signed-Rank tests that compare measures in *LitLinker* and baseline conditions. With Tails = Two (can not tell in advance which of the two measures is larger), Parent distribution = Normal (default), Effect size = 0.80 (calculated as Cohen’s d [11], large effect), α err prob = 0.05 (default), and power ($1 - \beta$ err prob) = 0.8 (an acceptable threshold), the output recommended smallest sample size is 15. Five participants, including two third-year undergraduate and three graduate students, major in the fields related to computer science (CS). They could provide feedback from those who know about the techniques used in *LitLinker*, and they were capable of our ideation tasks as they all learned Chinese literature in elementary schools. The other eleven participants are all graduate students majoring in education-related fields. All of them are teachers in training, while P2, P4, and P9 have teaching experience. The others are pre-service teachers.. All participants expressed their interest in utilizing web resources and AI-assisted tools to design interdisciplinary contexts ($Mean = 4.62$, $SD = 0.48$; 1 - no interest at all, 5 - very interest in). All participants reported having experience with large language models (LLMs, e.g., ChatGPT) ($Mean = 4.0$, $SD = 0.94$; 1 - no experience, 5 - use daily).

5.2 Experimental Design

Experiment I is a within-subject design. Each participant completed one interdisciplinary exploration task with *LitLinker* and the other with a baseline setup.

5.2.1 Baseline Condition and LitLinker Condition. The baseline system is the LLM web application, which employs the Zhipu LLM model, the same model implemented in *LitLinker*. Participants could upload and read document files, thereby enabling the web app to access selected reading materials. Also, the participants had the option to utilize the Google search engine and Microsoft Office

Table 2: Participants involved in the within-subjects study

ID	Gender	Year	Age	Major	Freq. of AI Usage
P1	F	Graduate	23	Curriculum and Instruction – English Language Education	Daily
P2	M	Graduate	23	Curriculum and Instruction – English Language Education	Weekly
P3	F	Graduate	23	Curriculum and Instruction – English Language Education	Weekly
P4	F	Graduate	25	Curriculum and Instruction – English Language Education	Weekly
P5	F	Graduate	24	Curriculum and Instruction – English Language Education	Daily
P6	F	Graduate	23	Chinese Literature	Have Tried
P7	F	Graduate	22	Early Childhood Education and Child Development	Weekly
P8	F	Graduate	22	Early Childhood Education and Child Development	Infrequently
P9	F	Graduate	28	Educational Psychology	Daily
P10	F	Graduate	23	TESOL(Teaching English to Speakers of Other Languages)	Have Tried
P11	M	Graduate	29	Computer Science and Technology	Weekly
P12	M	Graduate	22	Computer Science and Technology	Daily
P13	M	Graduate	23	Electronics and Information Engineering	Weekly
P14	M	Undergraduate	20	Artificial Intelligence	Weekly
P15	M	Undergraduate	21	Artificial Intelligence	Weekly
P16	M	Graduate	24	Education	Daily

Word. In the *LitLinker* condition, participants were allowed to use Google, Microsoft Office Word, and *LitLinker*, but were not allowed to use the LLM web app. In both conditions, participants could freely choose whether, when, and how to utilize the provided tools.

5.2.2 Tasks-systems Assignment. We invited the experienced Chinese language teacher E1 (Table 1), who participated in our design process, to help us prepare the task materials and grade the task outcome. Specifically, E1 selected 16 reading materials from six textbooks, covering 15 different units, and divided them into two groups - Text Set 1 and Text Set 2. Each participant completed tasks with both groups of texts. The task assignments were as follows:

- Baseline (Text Set 1) + *LitLinker* (Text Set 2)
- *LitLinker* (Text Set 2) + Baseline (Text Set 1)
- *LitLinker* (Text Set 1) + Baseline (Text Set 2)
- Baseline (Text Set 2) + *LitLinker* (Text Set 1)

5.2.3 Procedure. Each participant was assigned two lesson-planning tasks. The task prompt was:

You are a sixth-grade Chinese language teacher, aiming to implement an interdisciplinary approach over a period of 3-5 days, establishing logical and thematic connections between in-class and extracurricular reading materials. You are facing challenges in organizing these reading materials. Your goal is to analyze the internal relationships among the texts, identify an ‘interdisciplinary context’, integrate the selected texts within this context, develop a course plan in a document explaining the relationship between the reading materials and the selected context, and propose potential literature or interdisciplinary instructional activities. You may refer to the provided documents of example lesson plans based on interdisciplinary contexts. You need to choose at least three from eight reading materials and integrate them with the context.

Participants received the original reading materials one day before the task and were instructed to spend 10 minutes reading them. On the day of the task, the four participants assigned to the same task-system group came to our lab. We introduced the background knowledge about literature and interdisciplinary instruction, and the components necessary for creating an effective context (3.3 DP2). To encourage their best performance, we informed the participants that their outputs would be evaluated by an expert based on the appropriateness of Context, alignment with the educational objectives, and depth of integration. The top three participants would be awarded a supplementary reward of 100 RMB.

For each task, we first introduced the task and demonstrated the assigned system (*LitLinker* or baseline). Subsequently, we assisted participants in setting up their task environment on the provided computers or their own devices, including opening the baseline system or *LitLinker*, Google, and Microsoft Office Word. Each participant independently completed the lesson planning task. We allocated 30 minutes for each task and informed them that they could complete the task ahead of time if they wished. Following the completion of each task, participants were required to fill out a questionnaire. A 10-minute break was provided between the two tasks. After completing both tasks, we conducted a final semi-structured interview. Each participant spent approximately 90 minutes in the experiment and received 100 RMB for compensation.

5.3 Measurements

RQ1. Ideation outcomes. To assess the quality of the outcomes, we invited E1 to rate all outcomes in a randomized order. E1 did not know which condition each output lesson plan came from and reported that she spent four to six minutes evaluating each outcome, and the assessment was based on three criteria (derived in Finding 6 in Section 3.3): Content Alignment, Alignment with Educational Objectives, and Depth of Integration, using a 7-point Likert scale (1 - not satisfied at all, 7 - fully satisfied) to indicate the extent to which each criterion was satisfied. Additionally, to make sense of

the ratings, we required E1 to provide comments for each lesson plan, highlighting good and bad aspects.

RQ2. Ideation process. Based on NASA-TLX, we formulated six questions to measure workload during the lesson planning process, including mental demand, physical demand, temporal demand, performance, effort, and frustration - the higher a TLX score suggests the higher the perceived workload.

RQ3. Perception of *LitLinker*. We adapted ten questions from the System Usability Scale (SUS) [6] to study effectiveness, efficiency, and satisfaction. To understand how users perceive *LitLinker*'s generated contexts and analysis, we adapted six questions from the Creativity Support Index (CSI) [9] to evaluate Exploration, Expressiveness, and Immersion. The Creativity Support Index (CSI) includes three key dimensions: Exploratory, which measures how well the system helps users explore diverse ideas; Expressiveness, which reflects the system's ability to support clear expression of creative ideas; and Immersion, which indicates how well the system allows users to stay focused on creative tasks.

The measured items for task workload, SUS, and CSI are rated using a standard 7-point Likert Scale (1 - strongly disagree, 7 - strongly agree).

5.4 Results and Analyses

For quantitative data, we used the Wilcoxon Signed-Rank test. We used G* Power software to compute sensitivity, given Parent distribution = Normal, $\alpha = 0.05$, Power = 0.8, and Total Sample size = 16, which outputs a required effect size of 0.77. For qualitative data, we conducted thematic analysis on the semi-structured interview under the same settings in the design process (Section 3.2) and presented comments from the outcomes as supporting evidence.

5.4.1 RQ1: Ideation Outcomes. Figure 4 shows the quality of outcome lesson plans from the within-subjects study. The results indicate high variance in outcome quality for both the baseline and *LitLinker* conditions, suggesting significant differences in user performance. *LitLinker* showed improvements in Completion (*LitLinker*: $M = 5.56$, $Mdn = 6$, $SD = 0.61$; baseline: $M = 4.19$, $Mdn = 4.5$, $SD = 2.01$; $p = 0.02$, Wilcoxon effect size (r) = 0.93), providing more guidance in organizing lesson components. In Subject Integration (*LitLinker*: $M = 4.63$, $Mdn = 6$, $SD = 2.26$; baseline: $M = 3.56$, $Mdn = 3$, $SD = 2.00$; $p = 0.34$, $r = 0.50$), *LitLinker* assisted in combining literature knowledge with disciplinary contexts like science and arts. For Knowledge Transfer (*LitLinker*: $M = 4.44$, $Mdn = 5.5$, $SD = 2.21$; baseline: $M = 3.56$, $Mdn = 3$, $SD = 2.00$; $p = 0.38$, $r = 0.42$), outcomes created with *LitLinker* were more likely to promote students' application of disciplinary knowledge in their classroom environment. These results suggest that *LitLinker* performed well in supporting users to connect the interdisciplinary contexts to the reading materials.

However, regarding Content Alignment (*LitLinker*: $M = 4.25$, $Mdn = 5$, $SD = 2.80$; baseline: $M = 5.69$, $Mdn = 7$, $SD = 2.23$; $p = 0.22$, $r = 0.57$), Internal Logic (*LitLinker*: $M = 4.00$, $Mdn = 3.5$, $SD = 2.65$; baseline: $M = 5.00$, $Mdn = 5.5$, $SD = 2.06$; $p = 0.29$, $r = 0.42$), Subject Objectives (*LitLinker*: $M = 3.44$, $Mdn = 3$, $SD = 2.24$; baseline: $M = 4.25$, $Mdn = 5$, $SD = 1.89$; $p = 0.32$, $r = 0.39$), and Curriculum Standards (*LitLinker*: $M = 3.44$, $Mdn = 3$, $SD = 2.65$; baseline:

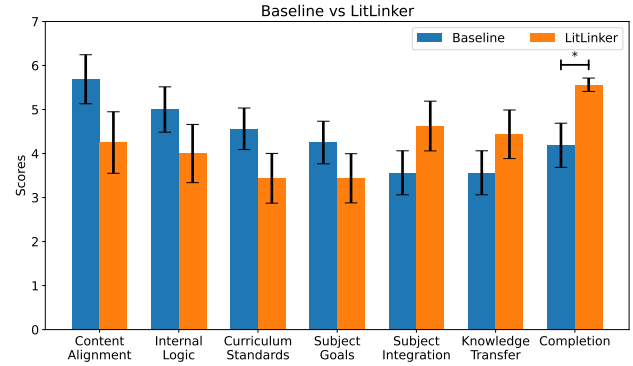


Figure 4: RQ1 results regarding the outcomes evaluated by E1 in seven different aspects. *: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, +: $p < 0.1$**

$M = 4.56$, $Mdn = 6$, $SD = 1.93$; $p = 0.16$, $r = 0.53$), the outcome lesson plans with the baseline system were rated significantly higher than those with *LitLinker*. This indicates that the outputs from *LitLinker* were less appropriate in terms of aligning the contexts with educational objectives compared to the baseline. We identify two possible reasons for these results based on the experimental setup and E1's comments on the outcome lesson plans. First, few participants had teaching experience, and none of them had taught Chinese literature before. In other words, participants were unfamiliar with the specific educational objectives associated with literature. During the introduction of the system and the task, we overemphasized the concept of "interdisciplinary", which led participants to focus more on other subjects rather than the educational objectives of the literature itself. In contrast, the experienced teacher E1 who evaluated the outcomes is highly sensitive to curriculum standards and literature objectives. E1 commented on an outcome that received a Curriculum Standards score of 1 (P08 - *LitLinker*), "The forced interdisciplinary integration is counterproductive, as it shifts the focus from literature to art or science, losing sight of the primary subject." Second, the system occasionally provided excessive or inaccurate interpretations of the texts, leading to lower appropriateness of context scores. "I feel that the system's trying to fit the analysis into a specific context, but it does not always feel very relevant." (P7). Most participants were not familiar with the reading materials, making it difficult for them to identify these issues. In contrast, E1 could easily identify such deviation. E1 commented on an outcome that received an appropriateness of context score of 1 (P11 - *LitLinker*), "The analysis does not match my understanding of the text."

In summary, while *LitLinker* improved Completion, Subject Integration, and Knowledge Transfer of the outcome lesson plans, it resulted in lower ratings concerning the appropriateness of context and alignment with educational objectives. This might be due to the experimental setup and the system's misinterpretation of the texts. We will focus on these issues in expert interviews to evaluate the quality of the outcomes.

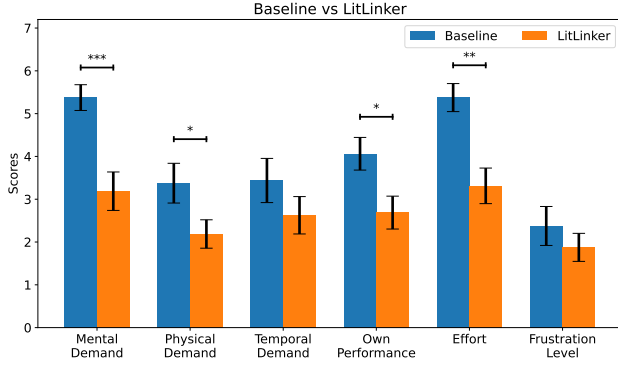


Figure 5: RQ2 results regarding evaluating how *LitLinker* affects the workload during the task. For each metric, a higher NASA-TLX score suggests a higher perceived workload. *: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, +: $p < 0.1$**

5.4.2 RQ2: Ideation Process. Figure 5 shows the participants' ratings on six NASA-TLX dimensions between the baseline system and *LitLinker* in the user study. Users rated mental demand and physical demand significantly lower with *LitLinker* (*LitLinker*: $M = 3.19$, $Mdn = 3$, $SD = 1.74$; baseline: $M = 5.38$, $Mdn = 5$, $SD = 1.17$, $p = 0.0009$, $r = 1.48$; *LitLinker*: $M = 2.19$, $Mdn = 2$, $SD = 1.29$; baseline: $M = 3.38$, $Mdn = 2.5$, $SD = 1.80$, $p = 0.01$, $r = 0.76$). Users felt significantly more satisfied with their performance using *LitLinker* (*LitLinker*: $M = 2.69$, $Mdn = 2$, $SD = 1.49$; baseline: $M = 4.06$, $Mdn = 4.5$, $SD = 1.48$; $p = 0.03$, $r = 0.93$) and reported significantly lower Effort (*LitLinker*: $M = 3.31$, $Mdn = 3.5$, $SD = 1.61$; baseline: $M = 5.38$, $Mdn = 5.5$, $SD = 1.27$; $p = 0.003$, $r = 1.42$). Frustration Level scores were also significantly lower with *LitLinker* (*LitLinker*: $M = 1.88$, $Mdn = 1.5$, $SD = 1.27$; baseline: $M = 2.38$, $Mdn = 1$, $SD = 1.76$; $p = 0.23$, $r = 0.33$), suggesting that users felt more confident during the exploration process. Additionally, users reported lower average Temporal demand with *LitLinker* (*LitLinker*: $M = 2.63$, $Mdn = 2$, $SD = 1.69$; baseline: $M = 3.44$, $Mdn = 3$, $SD = 2.00$; $p = 0.27$, $r = 0.44$). In summary, the participants indicate that *LitLinker* alleviates workload compared to the baseline system across all dimensions.

5.4.3 RQ3: Perception of *LitLinker*. Figure 6 presents the statistical results comparing the baseline system and *LitLinker* based on effectiveness, efficiency, and satisfaction scores. Participants rated *LitLinker* higher in all three aspects. Satisfaction scores were significantly higher for *LitLinker* (*LitLinker*: $M = 5.13$, $Mdn = 5.5$, $SD = 1.42$; baseline: $M = 3.88$, $Mdn = 3.83$, $SD = 1.10$; $p = 0.03$, $r = 0.99$), indicating greater user satisfaction with *LitLinker*. Efficiency scores also showed a significant increase with *LitLinker* (*LitLinker*: $M = 5.58$, $Mdn = 5.63$, $SD = 0.93$; baseline: $M = 4.48$, $Mdn = 4.75$, $SD = 1.36$; $p = 0.02$, $r = 0.94$), suggesting that participants found *LitLinker* more helpful and relevant to their tasks. For effectiveness, *LitLinker* received a higher mean score (*LitLinker*: $M = 4.69$, $Mdn = 4.83$, $SD = 1.48$; baseline: $M = 4.02$, $Mdn = 4$, $SD = 1.14$; $p = 0.06$, $r = 0.50$), implying that participants felt more effective using *LitLinker*.

For the three dimensions of the Creativity Support Index (Exploratory, Expressiveness, and Immersion), our system achieved higher average scores in all aspects. Users rated *LitLinker* significantly higher on Exploratory dimension (*LitLinker*: $M = 5.56$, $Mdn = 5.75$, $SD = 1.14$; baseline: $M = 4.44$, $Mdn = 4.5$, $SD = 1.31$; $p = 0.01$, $r = 0.91$), suggesting better support for exploratory creativity. On the Immersion dimension, users also gave *LitLinker* significantly higher ratings (*LitLinker*: $M = 4.91$, $Mdn = 5$, $SD = 1.28$; baseline: $M = 3.90$, $Mdn = 4.5$, $SD = 1.30$; $p = 0.02$, $r = 0.75$), indicating an enhanced ability to focus during creative tasks. For the Expressiveness dimension, users rated *LitLinker* higher (*LitLinker*: $M = 4.19$, $Mdn = 4.5$, $SD = 1.55$; baseline: $M = 3.88$, $Mdn = 3.75$, $SD = 1.14$; $p = 0.36$, $r = 0.23$), indicating a moderate improvement in expressiveness. Despite the overall strengths of *LitLinker* in creative support, six participants in the semi-structured interviews pointed out ways to further improve its creative support. First, in terms of exploratory dimensions, the system's interpretation of contexts and the quality of the AI-generated responses did not align with the expectations of some users, thereby diminishing their motivation to engage in further exploration. "(For a science context,) I expected it to provide more information related to students' everyday lives, but it mostly analyzed from abstract directions like aesthetics, which didn't match my expectations." (P6). "Its answers didn't address the questions I wanted to explore; sometimes it felt like it was using fancy language to evade the core issues." (P13). Regarding Expressiveness, four participants noted that *LitLinker* tended to "persuade" them to accept the contexts established by the system, rather than supporting their own critical thinking and expression. "It felt like scrolling through TikTok, reading the AI-generated decent outcomes without much thought." (P12). We observed that the eleven participants with education-related backgrounds ($M = 3.77$, $SD = 1.59$) tended to give lower scores on Expressiveness than the five students ($M = 5.10$, $SD = 1.34$) with CS-related backgrounds. P7, a student from an educational major, stated that "I find the system too acquiescent; it often responds with 'You are right' or 'I agree with you'. However, sometimes my thoughts were not necessarily correct, and I would prefer it to provide a more critical perspective." This suggests the system should further customize the support to users with education majors to construct and refine their initial ideas.

6 Experiment II: Expert Interviews

In Experiment I, we evaluated novice teachers' perceptions of our system, including the ideation process and their overall experience using *LitLinker*. Additionally, an experienced teacher, E1, evaluated the outcomes of the novice teachers. In Experiment II, we shifted our focus to the perspectives of experts with varying levels of literature teaching experiences to gain more feedback. We conducted expert interviews using a think-aloud protocol and a semi-structured interview with nine teachers to collect insights related to our research questions.

6.1 Participants

The study involved nine Chinese language teachers (5 female, 4 male, E1-5, E8-11 in Table 1) in a local elementary school, including two novice teachers with less than five years of teaching experience

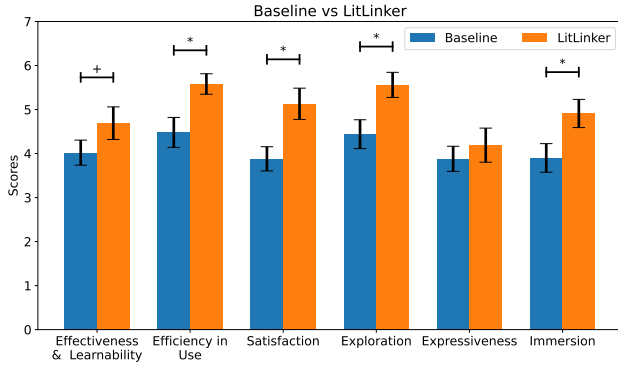


Figure 6: RQ3 results regarding users' perception of *LitLinker*, based on 3 aspects from SUS, and 3 aspects from CSI. *: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, +: $p < 0.1$**

and six expert teachers with at least five years of teaching experience [4]. Detailed information about the participants is presented in Table 3.

6.2 Method

We conducted interviews offline with E8-E11 and I1-I6 in lab sessions and with E1-E5 online after they freely used it for three days. This setup could help us to gain diverse insights, e.g., learnability of *LitLinker* in lab and field environments. I1-I6 participated in the evaluation of *LitLinker*'s prototype, and their feedback on the unchanged features in *LitLinker* was also presented in this section.

For E8 - E11, who were not involved in the iterative design process, we conducted offline interviews. The process began with a 5-minute introduction to the research background (i.e., the background of interdisciplinary literature instruction). This was followed by a 10-minute tutorial on how to use *LitLinker*. Participants were then allocated 30 minutes to complete exploration tasks while engaging in a think-aloud protocol. Subsequently, a 15-minute semi-structured interview was conducted.

For I1 - I6, who participated in the evaluation of the prototype process, the procedure was consistent with the same methodology above, except that during the semi-structured interview, we asked for more suggestions regarding the prototype.

For E1 - E5, who engaged in the iterative design process, we conducted online interviews. The *LitLinker* was made available for a duration of 3 days, and a tutorial video was provided. Participants were asked to freely use *LitLinker* during this period to complete exploration tasks. Finally, we conducted a 15-minute semi-structured interview with each expert.

Surrounding our RQs, the interview questions (Appendix A) are about ideation outcomes, ideation process and perception of *LitLinker*, while the participants were assigned the following exploration tasks:

- Task 1: Freely choose 4-15 reading materials, explore two contexts of interest, and add them to the collection.
- Task 2: Generate an introduction, course plan, and activities.

6.3 Findings

6.3.1 RQ1: Ideation Outcomes. Overall, seven of nine teachers confirmed that the outcomes generated by *LitLinker* effectively support literature instruction within the classroom environment. E10 emphasized the comprehensiveness of the text analysis, where she could find the content she wanted within a document more strategically, instead of sifting through countless reference books: “I can abandon varied reference books, (because) the system conducts a comprehensive analysis of the specific content of the texts. It includes themes, content, and key points.” Additionally, E11 noted that the activities could be easily implemented in the classroom: “yes, these activities (such as music appreciation) can be incorporated into upcoming lessons to increase student engagement.” I3 mentioned that the structured outcomes could potentially “make my teaching process more systematic.”

We observed the different views between novice and expert teachers on how they embrace the integration of generated activities into real classrooms, though almost all teachers consider *LitLinker*'s outcomes are beneficial for instructions. Experienced teachers (E9, E10, E11) preferred their self-centered teaching approach and were more conservative in using the generated activities. “To be honest, I have never used these (recommended) activities before, and in the future, I might only add a bit of them to my existing lesson plan to make students more interested.” (E11) On the contrary, newer teachers were more open and tended to reconstruct their established curriculum based on the recommended activities. E8 stated, “I plan to adjust my instruction method according to them (activities); some recommended activities are excellent in the current context and can be adapted for use in other contexts.”

Despite these positive findings, five of nine teachers pointed out that some outcomes did not align well with the objectives of literature instruction: “the outcomes are still somewhat disconnected from practical application. The content generated based on the provided template does not fully meet the current teaching needs, especially considering the recently revised curriculum standards” (E2). Teachers suggested a potential solution: “Import the curriculum standards and teaching objectives for each text, and to consider these goals when constructing contexts. For instance, what are the learning objectives and abilities required for each grade level?” (E5)

Despite modifications made to the prototype, 4 out of 9 experts indicated a need for additional outcomes that could directly enhance student learning in literature: “the documents generated are very useful for lesson preparation; however, they cannot be directly provided to students for learning purposes. I hope it can produce some homework questions” (E1).

In summary, the comprehensive and structured outcomes facilitate educational activities; however, additional focus is required to ensure the alignment of these outcomes with the literature objectives.

6.3.2 RQ2: Ideation process. Experts have reached a consensus that the task load is low when using *LitLinker* to construct contexts for the classroom, attributable to its well-structured layout and functional settings. E11 said, “the system does not burden my memory due to the collection feature”. However, E5, who used *LitLinker* in the wild for three days, suggested improvements to its UI design

and features to further decrease mental demand: “More specific instructions could be incorporated into the page to clearly indicate the available actions. Additionally, I would like to have a feature that synchronizes generated records through user login for long-term usage.” Experts in the lab sessions commented more on *LitLinker*’s error cases which may increase their effort in the ideation process. For example, E8 input “What activities can be designed around ‘Stepping Stones’ [the title of one text]”, and *LitLinker* responded “Students can observe stone bridges and steps in their daily lives, explore their design principles and practical uses. They can participate in group projects to build stone bridge models using materials like rocks, experiencing the joy of collaboration.” E8 commented, “This is indeed related to the article, but only the content, not the core idea or intended message”. E8 did not directly incorporate these outputs into the final outcome; instead, he edited manually. Additionally, teachers may feel frustrated due to LLM’s hallucinations. I3 input “Give me more sentences and analysis related to ‘Osmanthus Rain’ [the title of one text] and the context about poetic life”, and LLM responded “‘What I like is osmanthus. The osmanthus tree looks clumsy, unlike the plum tree, which has a graceful posture.’ This contrast highlights the author’s unique affection for osmanthus and reflects the author’s ability to discover poetic elements in life through a lens of beauty”. I3 remarked that “it did give me one more sentence but not really fit in the context.” In summary, *LitLinker* is generally user-friendly but requires more UI and feature refinements to optimize the ideation process.

6.3.3 RQ3: Perception of *LitLinker*. Both users in the lab session, who learned *LitLinker* through verbal instructions from the developer, and users in the wild, who studied it via a video tutorial, agreed that our system is easy to learn. With the exception of E2, all other teachers affirmed that *LitLinker* supported their exploration of contexts and the interrelationships between texts and contexts. *Expanding thinking* and *inspiring creativity* were mentioned as advantages for supporting context exploration. “The system provides a broader range of relationships between context and text, gradually expanding the context. I believe AI should work in this manner, incrementally broadening the scope to help me explore more possibilities, rather than rigidly offering a single answer” (E11). Furthermore, E5 added: “It can provide inspiration, demonstrating how to develop the class based on this context.” Besides expanding the breadth of thinking, E10 expressed appreciation for *LitLinker*’s summarization capabilities: “after selecting a substantial number of texts, I was astonished that it could truly organize them and produce a comprehensive design, which is impossible in my typical lesson preparation.” These comments indicate that *LitLinker* facilitates users in opening and focusing their cognitive processes during interdisciplinary exploration. However, E1 and E3 raised an issue about the repetition of suggested contexts when they tried different reading materials in their three-day usage of *LitLinker*. E1 noted, “I found that the same contexts reappearing despite my selection of entirely different reading materials”. Despite receiving feedback and modifications implemented after the prototype evaluation, this issue still troubled users and will be further discussed in 7.

In comparison to other generative tools and web search engines, users have reported a high level of trust in the output of our system, despite occasional acceptable errors. I3 stated, “I feel it is more

closely aligned with reading materials compared to previous tools, although a few analyses of the context are not very accurate. Overall, I still trust this system.” E4, I4 emphasized their reliance on personal experience and subjective judgment during the exploration process: “I always trust my own design more. When my ideas are limited, I use this system to evaluate its outputs and determine which content is usable” (E4).

In summary, users perceive *LitLinker* as highly usable and effective, and they generally express trust in its outputs.

7 Discussion

7.1 Design Considerations for AI-Teacher Collaboration in Interdisciplinary Contexts

Based on the findings from two phases of experiments, we propose three design considerations for interdisciplinary context ideation.

7.1.1 Encouraging Teacher Reflection Rather Than Full Automation. One of the important features of *LitLinker* is its high level of automation. During the exploration process, users can generate interdisciplinary contexts with detailed analyses and outcomes without the need for manual input unless asking the LLMs for answers. This presents a significant advantage for experienced experts who possess a clear understanding of their requirements for classroom practice, as it alleviates their workload (e.g., reducing the need for typing and searching), allowing them to concentrate fully on exploration. However, findings from the within-subjects study indicated that novice teachers’ behaviors suggest a potential risk associated with excessive automation. They might rely too heavily on the system, potentially neglecting their own cognitive process and allowing *LitLinker* to dominate the exploration, which could result in content that is disproportionately focused on disciplinary or literature-specific outcomes. Therefore, *LitLinker* should encourage teachers to make reflections to improve their teaching. Such a focus is already central to other teaching tools, such as simulations that enhance student engagement [55] and help teachers understand students’ learning states in online classrooms [38].

7.1.2 Improving the Diversity and Suitability of Informal Educational Context Pool. Based on DP4, we have collected 113 informal educational contexts from annual popular words in China and 144 subject-related contexts from textbooks to enable *LitLinker* to suggest contexts. In the within-subjects study, participants expressed appreciation for the diversity of contexts available in the laboratory setting. However, experts reported the lack of diverse and suitable contexts, especially the informal ones, when they would like to apply them in the literature courses. Besides, in the experiments, we observed *LitLinker* retrieves similar contexts in its RAG process to make recommendations. While implementing advanced RAG strategies might alleviate this, we argue for the need for more diverse datasets. The contexts sourced from Yao Wen Jiao Zi include popular words that Chinese teachers and students are likely to hear about, which could ease teachers’ load in understanding the contexts. However, these popular words (e.g., “My youth is back!”, “Prime Spot Debut”) may not always align seamlessly with pedagogical goals [53] and may lean toward entertaining purposes. Therefore, future work could consider identifying suitable types of

informal contexts with teachers and collecting more such contexts from diverse sources.

7.1.3 Meeting the Expectations of Different Users. In accordance with DP1, we carefully designed *LitLinker*'s workflow to align with teachers' practices and developed prompts using templates provided by E1-E7. However, our findings reveal that different user groups have varying expectations of the systems. For example, users without AI knowledge may need more expressive support that encourages them to criticize their own thoughts as well as the LLM output. Pre-service or new teachers could also be more depending on *LitLinker* in the ideation process. To support and motivate specific users, collecting various templates from them and replace the original ones in *LitLinker* can be beneficial (e.g., prompting Text Reviewer to offer more incisive feedback can motivate teachers to think deeply). Also, providing visualization of agent interactions and outputs can help users better understand the meaning of each step, to promote their critical thinking.

7.2 Reflections on AI-Teacher Collaboration

Previous researchers on human-computer interaction and education have proposed various intelligent tutoring agents, e.g., QuizBot [49], Sara [57], and DesignQuizzer [46]. These agents act like teachers to prompt questions and give adaptive feedback to learners, which do supplement the lack of accessible teachers in online learning scenarios. Nevertheless, we argue that human educators are still irreplaceable, especially in offline classrooms. Teaching is not just about transferring knowledge but also about shaping students' cognitive, social, and emotional development [1, 8, 22], which normally requires human teachers' interaction and empathy with students. Yet, these are qualities that AI currently lacks due to its generalization and lack of contextual awareness.

Our design, development, and evaluation of *LitLinker* provide a concrete example and implications of AI-teacher collaboration in such interdisciplinary teaching scenarios. In the design process, we identified with teachers the places where LLMs can help in their context ideation process, such as detailed analyses of the contexts, reading materials, and their relationships (DP2 Section 3.3) as well as documented outcome (DP3). This human support helps AI improve its outcome lesson plans and promotion of knowledge transfer to that subject (Figure 4). In addition, *LitLinker* reduced users' workload in ideating contexts for literature teaching (Figure 6), supporting that AI can serve as an auxiliary tool or a collaborator to improve the overall efficiency of teachers' tasks [25]. We implemented *LitLinker* that aligns its LLM output as closely as possible with teachers' practices (DP1), but the novices' outcome lesson plans with *LitLinker* did not align effectively with the teaching goals (Figure 4), and experts also pointed out that some recommended contexts were unreasonable. It indicates that the iterative design of AI-teacher collaborative systems can be longitudinal, wherein the diverse teaching experiences can be utilized as data to refine AI-generated content through techniques such as fine-tuning or multi-modal fusion.

We look forward to building an extensible version of *LitLinker* in which teachers without programming skills can easily customize

the system based on their needs to achieve more effective collaboration with AI. However, the ultimate responsibility for decisions of enacted teaching activities must remain with human educators.

7.3 Generalizability

LitLinker is open-source, allowing elementary school Chinese literature teachers to easily deploy and use it. *LitLinker* also has great potential to be generalized to support interdisciplinary teaching in elementary subjects apart from literature, in higher education, in other cultural contexts, and in self-learning tasks. For example, many elementary schools have explored STEM (science, technology, engineering, and mathematics) as an interdisciplinary approach to help students develop critical thinking, creativity, and problem-solving skills [16, 54], and some universities have opened interdisciplinary programs in which students can take courses in various domains. *LitLinker* can facilitate these educators with detailed analyses of the contexts and teaching materials, but to ensure it aligns with teachers' habitual practices, teachers' involvement in the design process is essential. To generalize *LitLinker* and our findings to other cultural or language contexts, we should further be aware of the biases embedded in the context pool. Many of *LitLinker*'s suggested contexts are sourced from annual popular words in China, which may potentially lead to misunderstanding or miscommunication in a multicultural environment. Lastly, while we position *LitLinker* as a support tool for teachers, its analyses of reading materials and contexts could be useful for self-learners. For instance, learners can select the literature they are interested in, get recommended contexts from *LitLinker*, and read them within a context in their after-school time.

7.4 Limitations and Future Work

This study has several limitations. First, we only met with the head teacher E1 of the interdisciplinary literature teaching team (E1-E7) in the foundational study. It introduced limitations such as a one-sided perspective or inaccurate representation of information, which might challenge the robustness of the derived knowledge. It would be more effective and insightful if we could directly reach them, e.g., via a brainstorming session or design workshop, and involve more teaching teams. Second, the participants in the within-subjects were students assigned to work as teachers, which means they do not possess adequate knowledge of teaching standards and the objectives of literature lessons. Therefore, the data obtained may not accurately reflect the perspectives of actual teachers. Also, the within-subjects study design with a small sample size ($N = 16$) significantly limits the robustness of our quantitative findings. Participants in our user study directly compare the systems rather than evaluate each system independently, which may lead to biased judgment. Future studies should consider a between-subjects study design with a larger sample size. Third, though we had an experienced teacher (E1) to score and comment on the outcome of ideation in Experiment I following the given criteria, it would further reduce the level of subjectivity involved in this rating task by inviting multiple experienced raters. Fourth, we evaluated *LitLinker* in a short period and did not systematically study participants' usage patterns. A future field study is needed to observe teachers' usage of *LitLinker* in their classrooms over a long period, e.g., the

adopted contexts and associated instructional methods they actually use. Fifth, our system is limited to textual content generation. Research indicates that the integration of multi-modal information (e.g., images, and videos) can enhance educational activities. Future work could explore the incorporation of these elements into the ideation process within interdisciplinary contexts. Finally, the algorithm implemented in *LitLinker* is not yet the most advanced LLM technology. Incorporating more advanced RAG algorithms and multi-agent LLM architectures has the potential to enhance the quality of the system's output.

8 Conclusion

In this paper, we developed an interactive system, *LitLinker*, through an iterative design process involving 13 Chinese language teachers. *LitLinker* is to support the ideation of interdisciplinary contexts for teaching literature, with the help of LLM-generated content. *LitLinker* includes three parts *i.e.*, Context Exploration View, Text Exploration View, and Collection. It provides space that users can explore and curate contexts deemed suitable for literature teaching, and then find out the reading materials that can be integrated into the teaching process. *LitLinker* provides detailed and precise analyses of the reading materials during exploration, subsequently providing structured outcomes for teachers. The within-subjects experiment involving 16 participants demonstrates that *LitLinker* facilitate the depth of interdisciplinary integration and eases their workload during the exploration process. Additionally, expert interviews were conducted with 9 teachers, who highlighted they trust in *LitLinker* and believe it expands their thinking. We discuss design considerations and insights derived from the user study for human-AI collaboration in education.

Acknowledgments

This work is supported by the Young Scientists Fund of the National Natural Science Foundation of China with Grant No.: 62202509 and the General Projects Fund of the Natural Science Foundation of Guangdong Province in China with Grant No. 2024A151012226.

References

- [1] Ibrahim Adeshola and Adeola Praise Adepoju. 2023. The opportunities and challenges of ChatGPT in education. *Interactive Learning Environments* (2023), 1–14.
- [2] Stephanie V Bestmeyer, Monica M Elser, Katie V Spellman, Elena B Sparrow, Stephanie S Haan-Amato, and Anna Keener. 2015. Collaboration, interdisciplinary thinking, and communication: new approaches to K–12 ecology education. *Frontiers in Ecology and the Environment* 13, 1 (2015), 37–43.
- [3] Veronica Boix Mansilla and Yves Lenoir. 2010. Interdisciplinarity in United States Schools: Past, Present, and Future. *Issues in integrative studies* 28 (2010), 1–27.
- [4] Josephine Booth, Mike Coldwell, Lisa-Maria Müller, Emily Perry, and James Zucollo. 2021. Mid-career teachers: A mixed methods scoping study of professional development, career progression and retention. *Education Sciences* 11, 6 (2021), 299.
- [5] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis*. American Psychological Association.
- [6] John Brooke. 2013. SUS: a retrospective. *Journal of usability studies* 8, 2 (2013).
- [7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc94967418bfb8ac142f64a-Paper.pdf
- [8] Cecilia Ka Yuk Chan and Louisa HY Tsi. 2024. Will generative AI replace teachers in Higher Education? A study of teacher and student perceptions. *Studies in Educational Evaluation* 83 (2024), 101395.
- [9] Erin Cherry and Celine Latulipe. 2014. Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)* 21, 4 (2014), 1–25.
- [10] DaEun Choi, Sumin Hong, Jeongeun Park, John Joon Young Chung, and Juho Kim. 2024. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–25.
- [11] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.
- [12] Charlene M Czerniak and Carla C Johnson. 2014. Interdisciplinary science teaching. In *Handbook of Research on Science Education, Volume II*. Routledge, 395–411.
- [13] Samantha L Dias-Lacy and Ruth V Guirguis. 2017. Challenges for New Teachers and Ways of Coping with Them. *Journal of Education and Learning* 6, 3 (2017), 265–272.
- [14] Zhuoyun Du, Chen Qian, Wei Liu, Zihao Xie, Yifei Wang, Yufan Dang, Weize Chen, and Cheng Yang. 2024. Multi-Agent Software Development through Cross-Team Collaboration. *arXiv preprint arXiv:2406.08979* (2024).
- [15] Lyn D English. 2016. STEM education K-12: Perspectives on integration. *International Journal of STEM education* 3 (2016), 1–8.
- [16] Lyn D English and Donna T King. 2015. STEM learning through engineering design: Fourth-grade students' investigations in aerospace. *International journal of stem education* 2 (2015), 1–18.
- [17] Haoxiang Fan, Guanzheng Chen, Xingbo Wang, and Zhenhui Peng. 2024. Lesson-Planner: Assisting Novice Teachers to Prepare Pedagogy-Driven Lesson Plans with Large Language Models. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–20.
- [18] Preciosa Fernandes, Carlinda Leite, Ana Mouraz, and Carla Figueiredo. 2013. Curricular contextualization: Tracking the meanings of a concept. *The Asia-Pacific Education Researcher* 22 (2013), 417–425.
- [19] Raymond Fok, Joseph Chee Chang, Tal August, Amy X Zhang, and Daniel S Weld. 2024. Qlarify: Recursively Expandable Abstracts for Dynamic Information Retrieval over Scientific Papers. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–21.
- [20] Jonas Frich, Michael Mose Biskjaer, and Peter Dalsgaard. 2018. Why HCI and creativity research must collaborate to develop new creativity support tools. In *Proceedings of the Technology, Mind, and Society*. 1–6.
- [21] Katy Ilonka Gero, Vivian Liu, and Lydia Chilton. 2022. Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference*. 1002–1019.
- [22] Alex Guilherme. 2019. AI and education: the importance of teacher and student relations. *AI & society* 34 (2019), 47–54.
- [23] Patricia L Hardré, Chen Ling, Randa L Shehab, Mark A Nanny, Matthias U Nollert, Hazem Refai, Christopher Ramseyer, Jason Herron, and Ebisa D Wollega. 2013. Teachers in an interdisciplinary learning community: Engaging, integrating, and strengthening K-12 education. *Journal of Teacher Education* 64, 5 (2013), 409–425.
- [24] Amin Heyrani Nobari, Muhammad Fathy Rashad, and Faez Ahmed. 2021. Creativegan: Editing generative adversarial networks for creative design synthesis. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Vol. 85383. American Society of Mechanical Engineers, V03AT03A002.
- [25] Steffen Holter and Mennatallah El-Assady. 2024. Deconstructing Human-AI Collaboration: Agency, Interaction, and Adaptation. In *Computer Graphics Forum*, Vol. 43. Wiley Online Library, e15107.
- [26] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiwu Zheng, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. 2024. MetaGPT: Meta Programming for A Multi-Agent Collaborative Framework. In *The Twelfth International Conference on Learning Representations*. <https://openreview.net/forum?id=VtmBAGCN7o>
- [27] Yihan Hou, Manling Yang, Hao Cui, Lei Wang, Jie Xu, and Wei Zeng. 2024. C2Ideas: Supporting Creative Interior Color Design Ideation with a Large Language Model. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.
- [28] Md Hafiz Iqbal, Shamsun Akhter Siddiqie, and Md Abdul Mazid. 2021. Rethinking theories of lesson plan for effective teaching and learning. *Social Sciences & Humanities Open* 4, 1 (2021), 100172.
- [29] Hyoungho Jin and Juho Kim. 2024. CodeTree: A System for Learnersourcing Subgoal Hierarchies in Code Examples. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–37.
- [30] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103

- (2023), 102274.
- [31] Nabin Khanal, Chun Meng Yu, Jui-Cheng Chiu, Anav Chaudhary, Ziyue Zhang, Kakani Katija, and Angus G Forbes. 2024. FathomGPT: A natural language interface for interactively exploring ocean science data. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–15.
 - [32] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.
 - [33] Junze Li, Changyang He, Jiaxiong Hu, Boyang Jia, Alon Y Halevy, and Xiaojuan Ma. 2024. DiaryHelper: Exploring the Use of an Automatic Contextual Information Recording Agent for Elicitation Diary Study. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.
 - [34] Katrine Lindvig and Lars Ulriksen. 2019. Different, difficult, and local: A review of interdisciplinary teaching activities. *The Review of Higher Education* 43, 2 (2019), 697–725.
 - [35] Michael Xieyang Liu, Tongshuang Wu, Tianying Chen, Franklin Mingzhe Li, Aniket Kittur, and Brad A Myers. 2024. Selenite: Scaffolding Online Sensemaking with Comprehensive Overviews Elicited from Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 837, 26 pages. <https://doi.org/10.1145/3613904.3642149>
 - [36] Ziyi Liu, Zhengzhe Zhu, Lijun Zhu, Enze Jiang, Xiyun Hu, Kylie A Peppler, and Karthik Ramani. 2024. ClassMeta: Designing Interactive Virtual Classmate to Promote VR Classroom Participation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
 - [37] Ryan Louie, Andy Coenen, Cheng Zhi Huang, Michael Terry, and Carrie J Cai. 2020. Novice-AI music co-creation via AI-steering tools for deep generative models. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
 - [38] Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An adaptable system for instructors to grasp student learning status in synchronous online classes. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–25.
 - [39] Chancharik Mitra, Mihran Miroyan, Rishi Jain, Vedant Kumud, Gireeja Ranade, and Narges Norouzi. 2024. RetLLM-E: Retrieval-Prompt Strategy for Question-Answering on Student Discussion Forums. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23215–23223.
 - [40] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, et al. 2024. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100027.
 - [41] Reza Hadi Mogavi, Chao Deng, Justin Juho Kim, Pengyuan Zhou, Young D Kwon, Ahmed Hosny Saleh Metwally, Ahmed Tlili, Simone Bassanelli, Antonio Bucchiarone, Sujit Gujar, et al. 2024. ChatGPT in education: A blessing or a curse? A qualitative study exploring early adopters' utilization and perceptions. *Computers in Human Behavior: Artificial Humans* 2, 1 (2024), 100027.
 - [42] Emiliana Murgia, Zahra Abbasiantaeb, Mohammad Aliannejadi, Theo Huibers, Monica Landoni, and Maria Soledad Pera. 2023. ChatGPT in the classroom: A preliminary exploration on the feasibility of adapting ChatGPT to support children's information discovery. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. 22–27.
 - [43] Emanuele Antonio Napoli, Fadi Barbàra, Valentina Gatteschi, and Claudio Schifanella. 2024. Leveraging Large Language Models for Automatic Smart Contract Generation. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 701–710.
 - [44] Elena Navrotskaya. 2024. Reading Club in English as an Interdisciplinary Course in University: A Didactic Proposal. *International Journal of Humanities Education* 22, 1 (2024).
 - [45] Carole L Palmer and Katrina Fenlon. 2010. Information research on interdisciplinarity. *The Oxford handbook of interdisciplinarity* (2010), 174–188.
 - [46] Zhenhui Peng, Qiaoyi Chen, Zhiyu Shen, Xiaojuan Ma, and Antti Oulasvirta. 2024. DesignQuizzer: A Community-Powered Conversational Agent for Learning Visual Design. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW1 (2024), 1–40.
 - [47] Savvas Petridis, Nicholas Diakopoulos, Kevin Crowston, Mark Hansen, Keren Henderson, Stan Jastrzebski, Jeffrey V Nickerson, and Lydia B Chilton. 2023. Anglekinding: Supporting journalistic angle ideation with large language models. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–16.
 - [48] Diana Rhoten, V Boix Mansilla, Marc Chun, and Julie Thompson Klein. 2006. Interdisciplinary education at liberal arts institutions. *Teagle Foundation white paper* 13 (2006), 2007.
 - [49] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L Murnane, Emma Brunskill, and James A Landay. 2019. Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–13.
 - [50] Julie Sarama, Kimberly Brenneman, Douglas H Clements, Nell K Duke, and Mary Louise Hemmeter. 2017. Interdisciplinary teaching across multiple domains: The C4L (Connect4Learning) curriculum. In *Implementing a standards-based curriculum in the early childhood classroom*. Routledge, 1–53.
 - [51] Orit Shaer, Angelora Cooper, Osnat Mokryn, Andrew L Kun, and Hagit Ben Shoshan. 2024. AI-Augmented Brainwriting: Investigating the use of LLMs in group ideation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.
 - [52] Muhammad Shidiq. 2023. The use of artificial intelligence-based chat-gpt and its challenges for the world of education; from the viewpoint of the development of creative writing skills. In *Proceeding of international conference on education, society and humanity*, Vol. 1. 353–357.
 - [53] Tatjana Sidekierskienė and Robertas Damaševičius. 2024. Pedagogical memes: a creative and effective tool for teaching STEM subjects. *International Journal of Mathematical Education in Science and Technology* (2024), 1–31.
 - [54] Gale M Sinatra, Anyana Mukhopadhyay, Taylor N Allbright, Julie A Marsh, and Morgan S Polikoff. 2017. Speedometry: A vehicle for promoting interest and engagement through integrated STEM instruction. *The Journal of Educational Research* 110, 3 (2017), 308–316.
 - [55] Xiaohan Tang, Sam Wong, Kevin Pu, Xi Chen, Yalong Yang, and Yan Chen. 2024. VizGroup: An AI-assisted Event-driven System for Collaborative Programming Learning Analytics. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–22.
 - [56] Massood Towhidnejad, Salamah Salamah, and Thomas B Hilburn. 2011. Introducing case study based teaching in computing curricula: Pre-conference workshop. *Journal of Computing Sciences in Colleges* 26, 6 (2011), 13–15.
 - [57] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
 - [58] Minghao Wu, Jiahao Xu, and Longyue Wang. 2024. TransAgents: Build Your Translation Company with Language Agents. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 131–141.
 - [59] Wenhan Xiong, Jingyu Liu, Igor Molybog, Hejia Zhang, Prajwal Bhargava, Rui Hou, Louis Martin, Rashi Rungta, Karthik Abinav Sankararaman, Barlas Oguz, et al. 2024. Effective Long-Context Scaling of Foundation Models. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 4643–4663.
 - [60] Zihan Yan, Yanhong Wu, Danli Luo, Chao Zhang, Qihang Jin, Wei Chen, Yingcai Wu, Xiang'Anthony' Chen, Guanyun Wang, and Haipeng Mi. 2023. NaCanva: Exploring and Enabling the Nature-Inspired Creativity for Children. *Proceedings of the ACM on Human-Computer Interaction* 7, MHCI (2023), 1–25.
 - [61] Morgan B Yarker and Soonhye Park. 2012. Analysis of teaching resources for implementing an interdisciplinary approach in the K-12 classroom. *Eurasia Journal of Mathematics, Science and Technology Education* 8, 4 (2012), 223–232.
 - [62] Aybars Yazici, Paola Meija-Domenzain, Jibril Frej, and Tanja Käser. 2024. GELEX: Generative AI-Hybrid System for Example-Based Learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–10.
 - [63] Ramazan Yilmaz and Fatma Gizem Karaoglan Yilmaz. 2023. Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans* 1, 2 (2023), 100005.
 - [64] Hye Sun You. 2017. Why teach science with an interdisciplinary approach: History, trends, and conceptual frameworks. *Journal of Education and Learning* 6, 4 (2017), 66–77.
 - [65] Kangyu Yuan, Hehai Lin, Shilei Cao, Zhenhui Peng, Qingyu Guo, and Xiaojuan Ma. 2023. CriTrainer: An Adaptive Training Tool for Critical Paper Reading. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–17.
 - [66] Chao Zhang, Cheng Yao, Jiayi Wu, Weijia Lin, Lijuan Liu, Ge Yan, and Fangtian Ying. 2022. StoryDrawer: a child-AI collaborative drawing system to support children's creative visual storytelling. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. 1–15.
 - [67] Chengbo Zheng, Kangyu Yuan, Bingcan Guo, Reza Hadi Mogavi, Zhenhui Peng, Shuai Ma, and Xiaojuan Ma. 2024. Charting the Future of AI in Project-Based Learning: A Co-Design Exploration with Students. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
 - [68] Chengbo Zheng, Yuanhao Zhang, Zeyu Huang, Chuhan Shi, Minrui Xu, and Xiaojuan Ma. 2024. DiscipLink: Unfolding Interdisciplinary Information Seeking Process via Human-AI Co-Exploration. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–20.

A Interview Questions in Expert Interviews

The fixed questions for the semi-structured interviews in Experiment II (expert interviews) are shown below.

RQ1: Ideation outcomes

- What do you think about the quality of the outcomes generated by *LitLinker*?
- Do you think the results generated by our system can assist you in preparing lessons or designing a new interdisciplinary context in the classroom setting?

RQ2: Ideation process

- Did you feel the task load is high while using *LitLinker*? Specifically, did you feel any increased cognitive load or mental demands?

RQ3: Perception of *LitLinker*

- Do you think our system can assist you in ideating interdisciplinary topics more efficiently?
- Do you think our system can help you better explore connections between texts and the contexts?
- If you were to prepare an interdisciplinary reading case, compared to your usual preparation methods (e.g., searching, meeting with other subject teachers, or using ChatGPT), do you think *LitLinker* could support you better?
- Do you trust the output of our system compared to your previous experiences with web searching or using ChatGPT?