SpatiaLearn: Exploring XR Learning Environments for Reflective Writing

Jinqiao Li
Bern University of Applied Sciences
Bern, Bern, Switzerland
jinqiao.li@bfh.ch

Seyed Parsa Neshaei
EPFL
Lausanne, Switzerland
seyed.neshaei@epfl.ch

Livia Müller
Bern University of Applied Sciences
Bern, Bern, Switzerland
livia.mueller@bfh.ch

Roman Rietsche
Bern University of Applied Sciences
Bern, Bern, Switzerland
roman.rietsche@bfh.ch

Richard Lee Davis
KTH Royal Institute of Technology
Stockholm, Sweden
rldavis@kth.se

Thiemo Wambsganss
Bern University of Applied Sciences
Bern, Bern, Switzerland
thiemo.wambsganss@bfh.ch

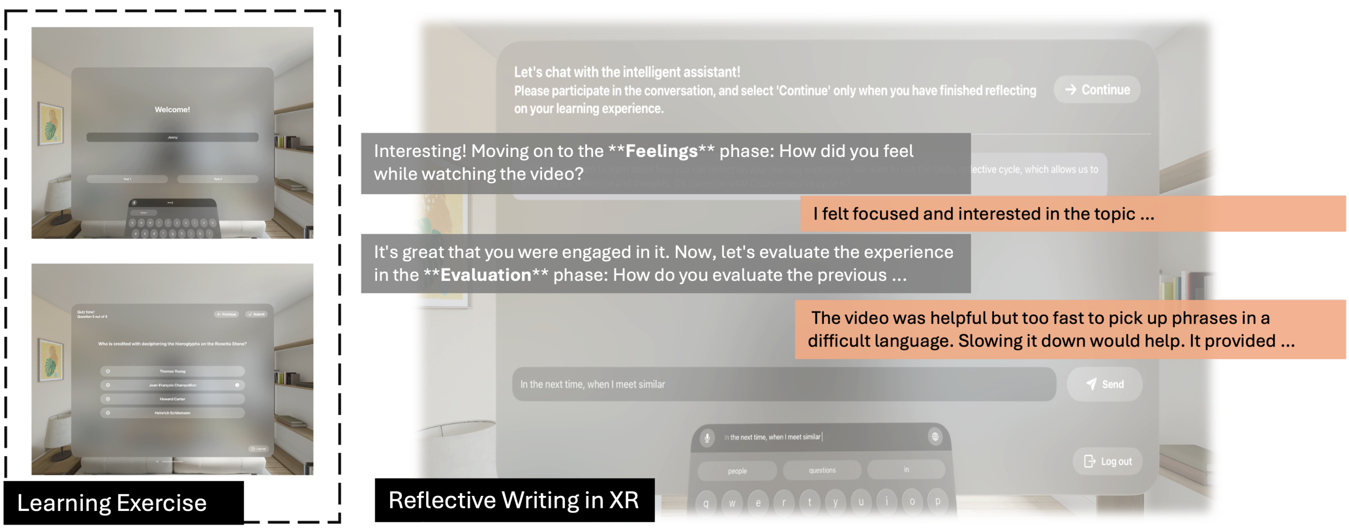

Figure 1: *SpatiaLearn*, a generative AI-supported, cross-platform application for reflective writing across diverse learning interfaces. Participants first engage with the learning exercise by watching a video and completing a quiz, followed by interaction with an LLM-powered conversational agent to conduct reflective writing guided by Gibbs' Reflective Cycle.

Abstract

Reflective writing promotes deeper learning by enhancing metacognitive awareness and critical thinking, but learners often struggle with structuring their reflections and maintaining focus. Generative AI and advances in spatial computing offer promising solutions. Extended reality (XR) environments create immersive, distraction-free settings, while conversational agents use dialog-based scaffolding guides to structure learners' thoughts. However, research on combining dialog-based scaffolding with XR for reflective writing remains limited. To address this, we introduce *SpatiaLearn*, an adaptive XR tool that enhances reflective writing through conversational guidance in both traditional and immersive environments. A within-subjects study ($N = 19$) compared participants' performance in traditional laptop and XR environments. Qualitative analysis shows the spatial interface enhances engagement but raises challenges like unfamiliar interactions and health concerns, requiring task adaptation for XR. This study advances the design of immersive tools for reflective writing, highlighting both the opportunities and challenges of spatial interfaces.

CCS Concepts

• **Human-centered computing** → **User studies**; **Mixed / augmented reality**; • **Computing methodologies** → **Natural language processing**; • **Applied computing** → **Distance learning**.

Keywords

Extended Reality (XR), Spatial Computing, Adaptive Education, Conversational Tutoring

CHI EA '25, Yokohama, Japan

ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3719742

ACM Reference Format:
Jinqiao Li, Seyed Parsa Neshaei, Livia Müller, Roman Rietsche, Richard Lee Davis, and Thiemo Wambsganss. 2025. SpatiaLearn: Exploring XR Learning Environments for Reflective Writing. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3706599.3719742

1 Introduction

Reflection, defined as the process of generating insights from one's experiences following a theoretical or practical learning session [34], is essential for fostering deeper understanding, critical thinking, and personal growth [8]. Reflective writing enables learners to evaluate past events critically [7] and plan for improved approaches in future learning tasks [10]. Prior research has demonstrated the benefits of reflective activities in many expertise training and education processes [3].

A common challenge in traditional reflection environments, such as utilizing laptop or mobile interfaces, is the difficulty of maintaining concentration due to the rich surrounding environment [1, 25, 32]. The frequent distractions from the surroundings conflict with the inherently high demand for sustained focus required for reflective writing [27, 30]. High-quality reflective writing requires sustained focus, deep engagement, and uninterrupted mental effort to analyze past experiences critically. However, traditional static laptop interfaces present various factors that may affect concentration, such as external noise, abundant contextual information, and peripheral views.

To overcome these limitations, integrating XR with generative AI offers an opportunity to create immersive environments that minimize distractions and foster deeper engagement in reflective writing. XR encompasses Augmented Reality (AR), Virtual Reality (VR), and Mixed Reality (MR), seamlessly integrating real and virtual environments to create highly immersive and interactive experiences [22]. The emergence of advanced head-mounted devices, like Apple Vision Pro, offers an opportunity to create immersive, low-distraction spatial reflection environments tailored to reflective writing. Devices such as Meta Quest and Apple Vision Pro are shaping XR environments that promise enhanced productivity, focus, and engagement compared to conventional setups [26, 35]. Such new-generation devices equipped with high-resolution displays (e.g., Apple Vision Pro: 3660 x 3200 pixels per eye), now make text-based tasks more feasible within XR environments. Research suggests that XR's immersive capabilities can enhance cognitive engagement, improving memory, focus, and overall task performance [12, 21], as well as supporting self-reflection [33]. Additionally, XR's spatially immersive environments stimulate motivation and facilitate information retention through interactive and contextual learning experiences [11, 13].

Other than XR's strengths in fostering engagement and retention, advancements in generative AI, particularly large language models (LLMs), offer complementary capabilities. Many studies have demonstrated the effectiveness of supporting reflective writing with adaptive feedback from LLMs [14, 16, 17, 36]. For example, Suraworachet et al. [29] showed that providing comprehensive feedback significantly enhanced engagement, improving reflective writing performance. Integrating LLMs, such as GPT-4, into XR learning environments adds an interactive dimension, guiding students through complex tasks and delivering real-time, personalized feedback. Moreover, in the XR learning context, LLMs have significantly enhanced students' reflection motivation and engagement, ultimately improving learning outcomes [6]. Cheng et al. [6] developed a narrative-based conversational agent to engage younger learners in dialogue with a mythological character, encouraging reflection on their prior XR learning experiences. Their results demonstrated that the conversational agent effectively enriched participants' reflective writing processes.

Although studies have explored the effectiveness of using XR or LLMs to enhance reflective writing, no research has investigated the combined application of these two tools or evaluated their integrated impact. For example, Li et al. [19] explored LLMs' capabilities in generating reflective writing but did not consider integration with XR. Similarly, Bozkir et al. [4] discussed embedding LLMs in XR environments without addressing reflective writing. This gap underscores the need for research on the synergistic use of XR and LLMs to advance reflective writing.

To investigate how spatial learning environments provided by state-of-the-art XR devices in combination with LLMs can support reflective writing tasks, our study examines reflective writing activities across immersive XR and conventional learning environments. Specifically, we aim to understand the following research questions:

- RQ1: How does the XR environment influence students' engagement and performance in reflective writing tasks facilitated by an LLM-enabled conversational agent?
- RQ2: What are learners' perceptions and attitudes toward transitioning reflective writing activities into an XR-based spatial environment?

We designed an LLM-enabled XR reflective writing application (*SpatiaLearn*) to investigate these research questions. Specifically, we integrated a GPT4-based conversational agent to facilitate dialog-based reflective writing. This feature provides learners with personalized guidance throughout the reflective process. By combining XR's spatial interactivity with LLM's adaptive feedback, *SpatiaLearn* offers a novel approach to enhancing reflective writing experiences. It aims to understand XR's role in facilitating generative AI-supported educational activities, focusing on the interplay between spatial interfaces and LLM-driven feedback mechanisms.

To explore the impact of immersive technologies on reflective writing, we conducted a within-subjects study with 19 university students. Participants compared their reflective writing experiences in XR using Vision Pro to those with a traditional laptop-based interface. Results indicated that the XR interface promoted greater engagement and concentration, though challenges such as interaction modalities and potential health implications underscored the novelty of XR in educational settings. Our study highlights its potential to promote focus and cognitive engagement in high-demand educational activities like reflective writing. We contribute to the existing literature by:

- Providing new design knowledge by introducing *SpatiaLearn*, an LLM-enabled XR system design, to support dialog-based reflective writing tasks in both spatial and traditional 2D learning environments.

- Providing new insights through qualitative analysis of how XR environments influence engagement and performance in reflective writing tasks facilitated by an LLM-enabled conversational agent (RQ1).
- Offering insights into the opportunities and challenges of transitioning reflective writing into immersive spatial environments (RQ2).

2 Design of *SpatiaLearn*

2.1 System Design

SpatiaLearn is a novel XR+LLM learning system designed for dialog-based reflective writing tasks, applying a GPT-4 powered conversational agent to provide real-time, adaptive feedback. The system was deployed across two distinct platforms—Apple Vision Pro (visionOS) and laptops (macOS)—using the same general components to ensure comparability between spatial and conventional environments.

The front-end of *SpatiaLearn* was developed using Swift 5 and the SwiftUI framework with Xcode 15, enabling a single codebase to be deployed seamlessly on both visionOS and macOS. To minimize novelty effects and enhance user comfort, we designed *SpatiaLearn* by adhering to Apple's Human Interface Guidelines [2].

The backend, built with Python using the Flask framework, serves two key functions. First, it facilitates the conversational tutor by integrating OpenAI's GPT-4 model. Second, it records user interactions, like the conversational tutor message logs, enabling detailed data analysis. Two researchers iteratively refined the tutor's system prompt in a workshop, following structured prompting principles. As detailed in Appendix D, the final prompt design employed instruction-based and few-shot prompting techniques.

The system evaluates and compares participants' learning experiences through a dialogue-based reflective writing task, representing a prototypical activity within daily learning processes. The system incorporates a supplementary video-watching activity to contextualize this reflective task. This ensures participants have a shared knowledge base to inform and enhance the quality of their reflective writing [9].

Participants are provided with two input modalities for interacting with the conversational tutor across both environments: (1) text input via a physical keyboard and (2) voice input processed through voice-to-text functionality. Additionally, due to the specificity of the Vision Pro, participants also have the option to type using a virtual keyboard and perform tapping gestures within the spatial learning interface.

2.2 Reflection Activity Design

To make the reflection process more comparable and effective [16], we provide a learning exercise as the reflection context. This dialog-based reflective writing task is directly connected to the video content, providing an integrated evaluation of foundational knowledge and reflective insights. By leveraging XR's immersive and interactive features, the study explores how these environments can amplify the impact of LLM-driven feedback, engagement, and critical thinking. This focus allows us to assess the potential of XR to elevate generative AI-supported reflection experiences, using reflective writing as a practical and structured example.

Reflection Exercise. Participants engage in a learning exercise focused on factual knowledge to provide a foundational context for reflective writing. Participants watch a 5-minute video on specialized history or language learning topics selected to establish a consistent foundational understanding. To confirm comprehension, they complete five multiple-choice questions assessing recall and understanding of the content. This process ensures a shared baseline of knowledge, enhancing the meaningfulness and comparability of the following reflective writing activity.

Dialog-Based Reflective Writing Task. The reflective writing task constitutes the core reflection activity, supported by an LLM-enabled conversational agent. Delivered across different immersive environments, the task uses Gibbs' Reflective Cycle [10] as a structured framework for guiding participants through a meaningful reflective process. The cycle comprises six stages: description, feelings, evaluation, analysis, conclusion, and action plan.

The GPT-4-powered tutor engages participants in a dialogue format, facilitating the reflective writing exercise by (1) introducing the user to the concept of reflective thinking, (2) guiding them through the reflective steps, and (3) providing adaptive responses based on participant inputs. This reflective exercise is directly connected with the video content from the preparatory task. Participants apply Gibbs' reflective framework to their learning experience from the video, analyzing and evaluating their engagement and comprehension within the corresponding learning environment.

3 Study Methods

To examine the spatial interface's impact on dialog-based reflective writing tasks, we conducted a within-subject user study employing a qualitative analysis. This allowed for the direct comparison of participants' performance across different interface conditions while controlling for individual differences. To mitigate potential sequence effects, participants were randomly assigned to begin the study in either the immersive or traditional 2D environment. Nineteen participants from a Western European university were recruited (13 females, 6 males; $M = 26.89$, $SD = 5.74$, age range: 21–40). In the end, each participant took part in a semi-structured interview conducted in English. The experiment lasted 90–120 minutes, and participants were compensated $60 USD.

3.1 Study Design

Figure 2 illustrates the whole study design for a student participating in our experiment. Before starting the experiment, participants were asked to complete a single-item self-assessment to measure their familiarity with the learning material [18] and prior experience with XR technologies [24].

According to the pre-test questionnaire, only a few participants reported having limited experience with XR learning. Three participants had prior experience with Microsoft HoloLens, and three others had limited exposure to VR devices in museums or libraries. The experiment began with an introductory guidance session for the XR device to mitigate the novelty effect. Each participant experienced the learning and reflective writing tasks on both interaction environments—the XR interface and the conventional laptop interface. To maintain consistency, participants reflected on the preparatory learning experience specific to each environment: reflections in

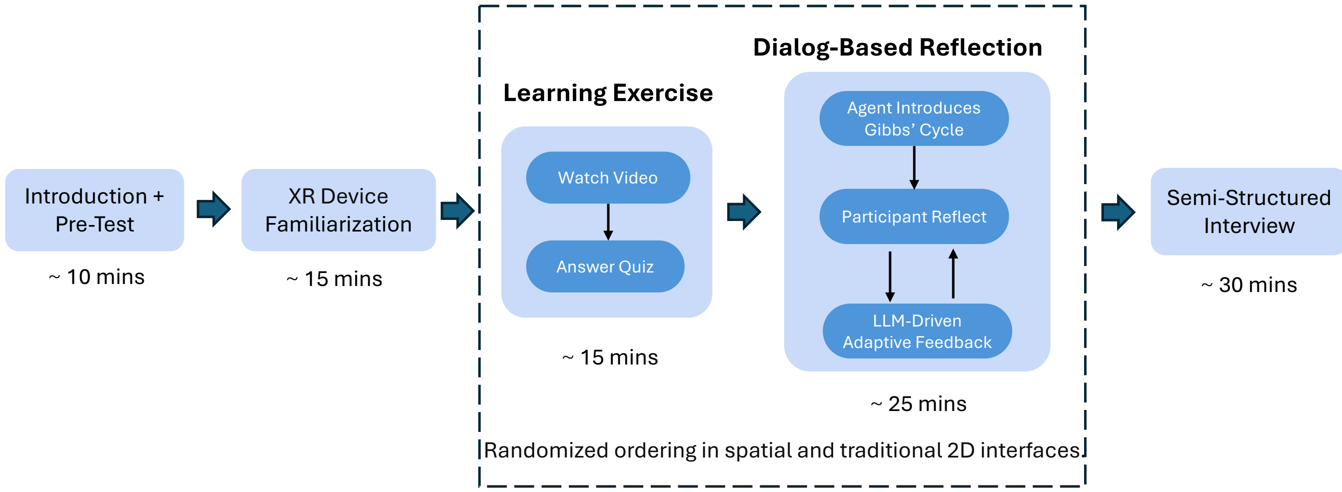

Figure 2: The overall study design is shown. Participants complete reflective writing tasks with the conversational agent across both spatial and conventional interfaces, using different learning videos as context.

the XR interface focused on the video-watching experience conducted in XR, while reflections in the laptop interface focused on the corresponding 2D experience.

The experiment concluded with a semi-structured interview to gather participants' viewpoints. The primary objective was to gain deeper insights into the influence of the spatial interface on learning experiences and task performance. The interview questions were grounded in theory, covering five key areas: engagement and overall user experience [6], learning outcomes and effectiveness [13], cognitive load and adaptation [5], and individual concerns and perceptions of future potential. Table 1 outlines the qualitative thematic structure for analyzing the interview transcriptions.

3.2 Measures and Analysis

Participants' perceptions were captured using a questionnaire with a five-point Likert scale [20], assessing satisfaction and engagement levels. For satisfaction, the questions were: *"How satisfied are you with your learning experience on the laptop?"* and *"How satisfied are you with your learning experience on the Apple Vision Pro?"* Engagement was assessed similarly: *"How engaging was your learning experience on the laptop?"* and *"How engaging was your learning experience on the Apple Vision Pro?"* These questions were adapted from [31], a framework specifically designed to assess user experience in immersive virtual environments.

Post-test interviews were analyzed using qualitative content analysis [23], chosen for its structured methodology [28]. Coding was conducted by three researchers, starting with a deductive approach based on predefined categories aligned with research questions and literature. Iterative discussions refined the framework, resolved discrepancies, and identified nuances [15]. An inductive approach was then applied to uncover emergent themes, capturing both expected and unexpected findings. The analysis identified four primary categories: *"Impact on Learning Process"*, *"Impact on Cognitive Process and Learning Outcome"*, *"Device Usability and User Experience"*, and *"Future Considerations and Broader Implications"*, with 20 subcategories. Table 1 provides example quotes illustrating participants' perspectives on the spatial learning interface.

4 Results

This section presents the results of the semi-structured interviews and self-reports via questionnaires. To maintain participant anonymity, individuals are referred to using identifiers ranging from P1 to P19.

4.1 RQ1: Higher Engagement and Satisfaction in Spatial Learning Interface

For the reflective writing task, participants generally reported higher engagement and satisfaction with the adaptive spatial learning interface compared to the traditional 2D laptop interface. Participants attributed increased engagement to enhanced focus enabled by the immersive interface. For instance, P9 noted, *"I just feel like I'm less distracted maybe with the (Vision) Pro."* P2, who has ADHD, stated, *"For the first time in my life, I could take notes while watching a learning video,"* highlighting the interface's potential to support attention-challenged learners. Participants also expressed a greater inclination to use the speech-to-text function, as it felt more intuitive within the XR environment. P10 remarked, *"It feels more natural to use voice in this (XR environment)."* Overall, participants reported comparable satisfaction levels when reflecting on the spatial interface. Additionally, 89.4% of participants found the spatial environment engaging for reflective activities, representing a 5.3% increase compared to the conventional laptop interface.

Another noteworthy finding is the increased engagement and satisfaction observed during the learning exercise, where participants watched a factual knowledge video and completed five multiple-choice quiz questions. Participants achieved a higher average score on the spatial learning interface (M = 3.79) than on the laptop interface (M = 3.63). This outcome improvement may be attributed to the enhanced focus and attention facilitated by the spatial learning interface. Several participants (P3, P10, P16) described the experience as *"sitting in a big cinema"*, emphasizing that the large screen naturally captured their attention. Given that video-watching and reflective writing correspond to lower- and higher-order levels of Bloom's Taxonomy, these findings suggest that the spatial interface can enhance engagement and potentially support a range of cognitive tasks by providing a more immersive learning environment.

4.2 RQ2: Attitudes and Perception towards LLM-Powered Reflective Writing in Spatial Interface

We gained valuable insights into participants' attitudes toward reflection in the spatial interface through individual interviews. On the positive side, participants adapted to the spatial interface briefly, indicating its ease of usability. However, specific challenges emerged that may influence users' willingness to adopt the spatial interface. Initial unfamiliarity affected their perceived efficiency, while health and ethical concerns were also highlighted as important considerations.

Increased Adaptation Over Time. Figure 4 shows participants' engagement and satisfaction levels across two tasks, revealing changes within the short time gap between tasks (10 to 40 minutes). Participants' initial unfamiliarity with the Apple Vision Pro controls appeared to decrease over time, as indicated by improved focus and comfort during the second task (e.g., P9: *"the second time was easier because I already knew how to use it"*). Additionally, participants reported finding it more natural to reflect by speaking in the spatial interface than using the laptop. As participant P13 remarked, "*Speaking with the Vision Pro feels more natural and futuristic, whereas talking with a laptop feels awkward.*"

Discrepancy Between Perceived and Measured Performance. Some participants subjectively reported feeling less efficient when typing in the spatial interface (e.g., P9: *"it was definitely faster to chat (by typing) with it on the laptop"*). However, our quantitative analysis revealed a different outcome. The Efficiency Score (normalized using min-max scaling: $\frac{\text{Messages Sent}}{\text{Time on Device}}$) was comparable across both interfaces. The scores for the laptop ($M = 0.34$) and the HMD ($M = 0.35$) were nearly identical, suggesting a discrepancy between participants' subjective perceptions and their actual performance. Appendix Table 2 presents detailed results, including the mean number of messages sent, message length, total time spent on tasks, and overall efficiency scores across the spatial and conventional laptop interfaces.

Health Concerns. Some participants reported physical discomfort with prolonged HMD use, including neck and eye strain (e.g., P9, P10, P12) due to the device's weight and screen proximity. P12 expressed health concerns, stating, *"Without a health report, I won't use it for long and wouldn't recommend it to young teenagers."*[1] Additionally, red marks were observed on some participants' cheeks after 10 minutes of use, highlighting the need for advanced weight reduction and ergonomic improvements to accommodate diverse head shapes and sizes.

Ethical Concerns. P4 expressed concerns about the potential ethical implications of the spatial interface, particularly regarding privacy and the ability to supervise children's screen time. P4 noted that the interface might hinder parents' ability to monitor their children's activities, raising concerns about young children's susceptibility to becoming overly immersed in the digital environment without sufficient parental intervention. Additionally, Participant 6 expressed concern about society's increasing reliance on such devices.[2]

5 Discussion and Future Work

5.1 Promising Insights of Reflective Writing in XR Interface

By integrating adaptive feedback from an LLM-powered conversational agent into a spatial learning interface, *SpatiaLearn* advances the potential for personalized and immersive learning experiences. This aligns with prior research emphasizing the role of XR in promoting engagement and reducing distractions for educational activities [12, 21]. Our study contributes to the growing literature on XR-enhanced learning by examining its impact on reflective writing tasks, an area relatively underexplored in existing studies. In our study, reflective writing tasks supported by the LLM-powered agent demonstrated promising results in XR environments. The XR interface enhanced participants' immersion and engagement during dialog-based reflective writing. Participants reported feeling more focused and creative in the XR environment, facilitating deeper reflection, like *"easier to think critically."* These results align with studies demonstrating the potential of XR to support higher-order cognitive processes [11].

We propose that this improvement is driven by the isolated nature of the XR environment, which facilitates more natural and less-restricted interactions with the conversational agent. This finding complements existing research on XR's capacity to support reflective and dialog-based learning [14, 36] while also providing new insights into how these environments can amplify the benefits of LLM-driven adaptive feedback. By combining immersive interfaces and AI feedback, *SpatiaLearn* contributes a novel approach to enhancing engagement and depth in reflective writing, bridging gaps between XR, LLMs, and reflective pedagogy. Moreover, the increased adaptation to the new learning interface shows a high potential for seamless integration of XR tools into educational workflows. It suggests that some initial usability challenges participants report may diminish as users become more familiar with the interface.

5.2 Challenges of Reflective Writing in XR Interface

Despite these promising findings, several challenges persist that align with and expand upon existing research. First, participants hold polarized attitudes towards reflective writing in the spatial interface. While some participants found it *"amazing"* (P3), others

[1]Apple advises against use for individuals under 13 years old https://support.apple.com/guide/apple-vision-pro/important-safety-information-c0c84db82a44/visionos.

[2]However, it should be noted that Apple Vision Pro supports AirPlay on other devices, allowing guardians to supervise what their child is seeing on the headset at each point in time. The participants were unfamiliar with this feature; thus, this suggests the need for improved visibility, e.g., by presenting a notification informing the parents of this feature when Family Sharing is activated. Moreover, the participants were not informed of and did not have the chance to experiment with the EyeSight feature, which mirrors the wearer's eyes on the screen outside the Vision Pro, improving social connections while one is wearing the headset.

described it as *"scary"* (P6). Some adjusted over time and reported increased engagement (P13, P14), while others preferred the familiarity of the laptop interface (P1, P18). This variability aligns with prior studies highlighting the importance of tailoring XR environments to individual preferences to enhance user experience [12, 13].

Second, perception bias and psychological preferences remain. While participants reported lower typing efficiency in the spatial interface than the laptop, efficiency scores showed no significant difference. This subjective performance evaluation in the XR environment mirrors findings in [14]. Health and ethical concerns were also noted, consistent with broader research on XR adoption [11].

5.3 Limitations and Future Work

We acknowledge several limitations in our study and suggest directions for future research. First, the small sample size ($N = 19$) limits the applicability of the findings. Experiments with more participants are needed to validate the results and explore XR's full potential for educational activities requiring high cognitive engagement. In this study, we identified XR environment's potential in supporting attention-challenged learners. With a broader participant base, it would be valuable to investigate this possibility more systematically. Second, participants' exposure to the XR interface was limited to approximately two hours. This short duration may not capture long-term adaptation, changes in engagement, or potential physical and cognitive fatigue. A longer-term experiment would help address these concerns, validate adaptation over time, and explore strategies to enhance overall usability.

Lastly, the learning environment and headset device settings play a crucial role in shaping the learning experience. The environment determines the context in which the activity takes place, while the configuration of the headset can significantly impact user comfort. In this study, all participants used the same headset in a controlled environment, specifically a university seminar room, which did not account for real-world distractions. Future research could explore how different learning environments influence students' learning processes and performance. Learners may have varying preferences for study locations, such as a library, home, or public spaces, which could affect their concentration levels and overall learning outcomes. Investigating these variations would provide valuable insights into the relationship between environmental context and learning effectiveness. Furthermore, in this study, all participants used a headset with a fixed-length strap. The lack of customization may have contributed to discomfort and did not accommodate differences in face and head shapes. Future research should examine how various environmental factors, including noise levels, spatial constraints, and personalized device settings, affect usability, comfort, and learning effectiveness in both controlled and real-world settings.

Addressing these limitations will offer deeper insights into XR's usability and effectiveness for reflective writing tasks and other learning activities requiring high concentration levels.

6 Conclusion

In this study, we introduce *SpatiaLearn*, a tool designed to support reflective writing in an XR environment by incorporating adaptive elements to create personalized and immersive learning experiences. By leveraging the immersive capabilities of XR and the adaptive feedback provided by an LLM-powered conversational agent, our research demonstrates how reflective writing, a cognitively demanding activity that benefits from timely feedback, can be significantly enhanced. The findings indicate that the XR environment fosters greater engagement, creativity, and critical thinking by offering a focused and minimally distracting space for reflection. Additionally, integrating LLM-driven feedback helps learners navigate complex cognitive tasks, providing guidance that improves the depth and quality of their reflective writing. This study contributes empirical evidence to the growing field of XR-enhanced learning. It underscores the potential for adaptive XR tools to bridge the gap between advanced AI technologies and traditional educational practices. Future research could explore the long-term impacts of such tools, including their scalability and applicability across diverse learning contexts.

References

[1] Kamer-Ainur Aivaz and Daniel Teodorescu. 2023. Advantages and Disadvantages of Using Devices During In-Person Courses. Case Study at a Public University in Romania. *Studies in Business and Economics* 18, 2 (2023), 24–36.

[2] Apple Inc. 2025. *Human Interface Guidelines.* https://developer.apple.com/design/human-interface-guidelines/

[3] Giovanna Artioli, Laura Deiana, Francesco De Vincenzo, Margherita Raucci, Giovanna Amaducci, Maria Chiara Bassi, Silvia Di Leo, Mark Hayter, and Luca Ghirotto. 2021. Health professionals and students' experiences of reflective writing in learning: A qualitative meta-synthesis. *BMC medical education* 21 (2021), 1–14.

[4] Efe Bozkir, Süleyman Özdel, Ka Hei Carrie Lau, Mengdi Wang, Hong Gao, and Enkelejda Kasneci. 2024. Embedding large language models into extended reality: Opportunities and challenges for inclusion, engagement, and privacy. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces.* 1–7.

[5] Josef Buchner, Katja Buntins, and Michael Kerres. 2022. The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning* 38, 1 (2022), 285–303.

[6] Alan Y Cheng, Meng Guo, Melissa Ran, Arpit Ranasaria, Arjun Sharma, Anthony Xie, Khuyen N Le, Bala Vinaithirthan, Shihe Luan, David Thomas Henry Wright, et al. 2024. Scientific and Fantastical: Creating Immersive, Culturally Relevant Learning Experiences with Augmented Reality and Large Language Models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* 1–23.

[7] Binta M Colley, Andrea R Bilics, and Carol M Lerch. 2012. Reflection: A key component to thinking critically. *The Canadian Journal for the Scholarship of Teaching and Learning* 3, 1 (2012).

[8] Jordi Colomer, Teresa Serra, Dolors Cañabate, and Remigijus Bubnys. 2020. Reflective learning in higher education: Active methodologies for transformative practices. , 3827 pages.

[9] Janet Eyler and Dwight E Giles Jr. 1999. *Where's the Learning in Service-Learning? Jossey-Bass Higher and Adult Education Series.* ERIC.

[10] Graham Gibbs. 1988. Learning by doing: A guide to teaching and learning methods. *Further Education Unit* (1988).

[11] Patrick Guilbaud, T Christa Guilbaud, and Dane Jennings. 2021. Extended reality, pedagogy, and career readiness: A review of literature. In *International Conference on Human-Computer Interaction.* Springer, 595–613.

[12] Shuang He, Wen Sun, and Cuiqin Lan. 2024. Design and exploration of virtual multi-scene learning space based on XR technology. In *Proceedings of the 2024 9th International Conference on Distance Education and Learning.* 223–228.

[13] Farzam Kharvari and Lorenz Ewald Kaiser. 2022. Impact of extended reality on architectural education and the design process. *Automation in Construction* 141 (2022), 104393.

[14] Sun Kyung Kim, Youngho Lee, Hyoseok Yoon, and Jongmyung Choi. 2021. Adaptation of extended reality smart glasses for core nursing skill training among undergraduate nursing students: Usability and feasibility study. *Journal of medical Internet research* 23, 3 (2021), e24313.

[15] Udo Kuckartz. 2014. *Qualitative text analysis: A guide to methods, practice and using software.* Sage.

[16] Harsh Kumar, Ruiwei Xiao, Benjamin Lawson, Ilya Musabirov, Jiakai Shi, Xinyuan Wang, Huayin Luo, Joseph Jay Williams, Anna N Rafferty, John Stamper, et al. 2024. Supporting self-reflection at scale with large language models: insights from randomized field experiments in classrooms. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale.* 86–97.

[17] Gareth Lewis, Matt B Johns, Lydia S Vamvakeridou-Lyroudia, Albert S Chen, Slobodan Djordjević, and Dragan A Savić. 2024. Empowering Water Engineers to Develop XR Learning Applications with the WATERLINE Project. *Engineering Proceedings* 69, 1 (2024), 21.

[18] Minzi Li and Xian Zhang. 2021. A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing* 38, 2 (2021), 189–218.

[19] Yuheng Li, Lele Sha, Lixiang Yan, Jionghao Lin, Mladen Raković, Kirsten Galbraith, Kayley Lyons, Dragan Gašević, and Guanliang Chen. 2023. Can large language models write reflectively. *Computers and Education: Artificial Intelligence* 4 (2023), 100140.

[20] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of Psychology* 22, 140 (1932), 1–55.

[21] Guido Makransky and Lau Lilleholt. 2018. A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development* 66, 5 (2018), 1141–1164.

[22] Steve Mann, Tom Furness, Yu Yuan, Jay Iorio, and Zixin Wang. 2018. All reality: Virtual, augmented, mixed (x), mediated (x, y), and multimediated reality. *arXiv preprint arXiv:1804.08386* (2018).

[23] Philipp Mayring and Thomas Fenzl. 2019. *Qualitative inhaltsanalyse*. Springer.

[24] Michelle D Miller, Giovanni Castillo, Norman Medoff, and Alexis Hardy. 2021. Immersive VR for organic chemistry: Impacts on performance and grades for first-generation and continuing-generation university students. *Innovative Higher Education* 46 (2021), 565–589.

[25] María Ángeles Pérez-Juárez, David González-Ortega, and Javier Manuel Aguiar-Pérez. 2023. Digital distractions from the point of view of higher education students. *Sustainability* 15, 7 (2023), 6044.

[26] Georg Regal, Helmut Schrom-Feiertag, Quynh Nguyen, Marco Aust, Markus Murtinger, Dorothé Smit, Manfred Tscheligi, and Mark Billinghurst. 2022. VR [We Are] training-workshop on collaborative virtual training for challenging contexts. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–6.

[27] Michelle M Robbins, Grace O Onodipe, and Alan Marks. 2020. Reflective writing and self-regulated learning in multidisciplinary flipped classrooms. *Journal of the Scholarship of Teaching and Learning* 20, 3 (2020).

[28] M Schreier. 2012. Qualitative content analysis in practice. *Jacobs University Bremen* (2012).

[29] Wannapon Suraworachet, Qi Zhou, and Mutlu Cukurova. 2023. Impact of combining human and analytics feedback on students' engagement with, and performance in, reflective writing tasks. *International Journal of Educational Technology in Higher Education* 20, 1 (2023), 1–24.

[30] Kara Taczak and Liane Robertson. 2017. Metacognition and the reflective writing practitioner: An integrated knowledge approach. *Contemporary perspectives on cognition and writing* (2017), 211–229.

[31] Katy Tcha-Tokey, Olivier Christmann, Emilie Loup-Escande, and Simon Richir. 2016. Proposition and validation of a questionnaire to measure the user experience in immersive virtual environments. *International Journal of Virtual Reality* 16, 1 (2016), 33–48.

[32] Chih-Hsuan Wang, Jill D Salisbury-Glennon, Yan Dai, Sangah Lee, and Jianwei Dong. 2022. Empowering College Students to Decrease Digital Distraction through the Use of Self-Regulated Learning Strategies. *Contemporary Educational Technology* 14, 4 (2022).

[33] Yixin Wang, Yun Suen Pai, and Kouta Minamizawa. 2022. It's Me: VR-based Journaling for Improved Cognitive Self-Regulation. In *SIGGRAPH Asia 2022 Posters*. 1–2.

[34] Kate Williams, Mary Woolliams, and Jane Spiro. 2020. *Reflective writing*. Bloomsbury Publishing.

[35] Sinuo Wu, Karen Stendal, and Devinder Thapa. 2024. Emerging Trends in XR-mediated Virtual Team Collaboration in Digital Workspaces: A Systematic Literature Review. *Advances in Information Systems Development: Information Systems Development, Organizational Aspects, and Societal Trends* (2024), 85–108.

[36] Yudan Wu, Shanhe You, Zixuan Guo, Xiangyang Li, Guyue Zhou, and Jiangtao Gong. 2023. MR. Brick: designing a remote mixed-reality educational game system for promoting children's social & collaborative skills. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

A Screenshots During the Experiment

Figure 3 illustrates images captured during various stages of the experiment across different learning interfaces.

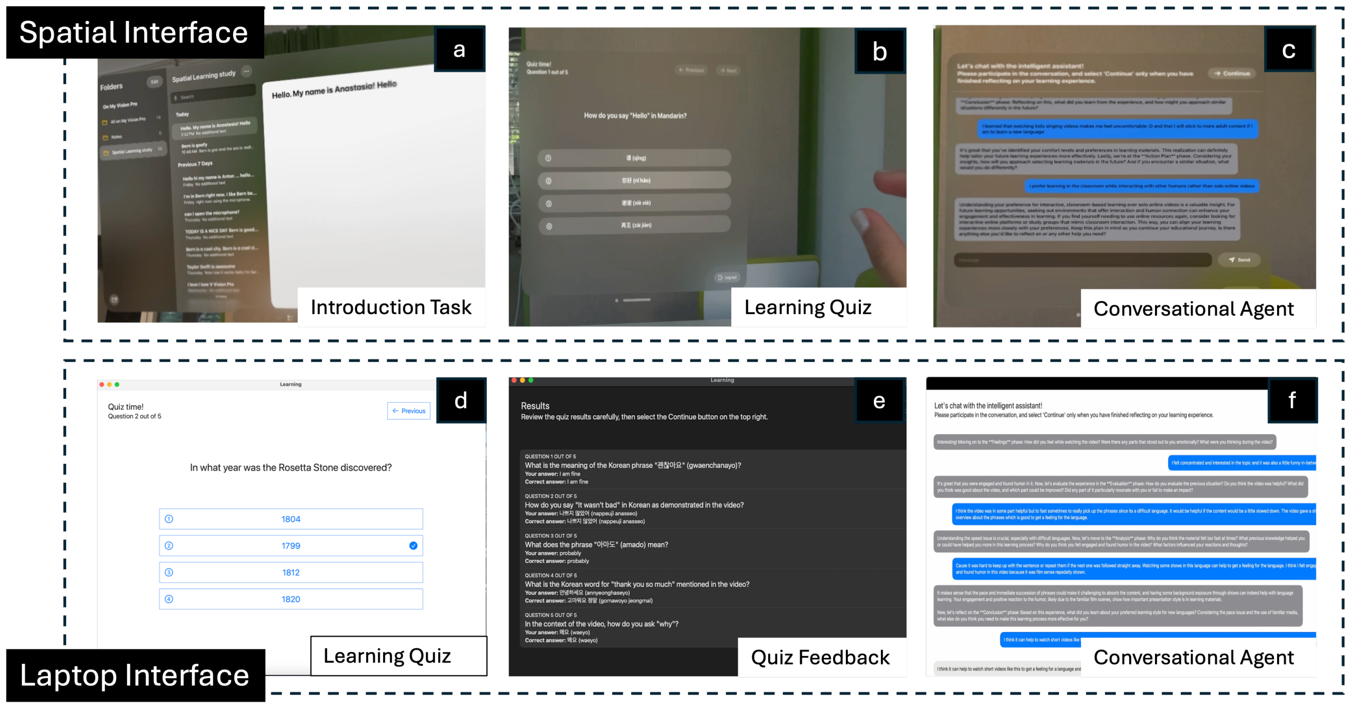

Figure 3: Screenshots during the practical study sessions in the lab on two different devices. The screenshots represent the two learning tasks shown on different interfaces: spatial interface (a-c) and laptop interface (d-f). (a) Introduction Task: The introduction to spatial interface operations before the learning process. (b) Learning Quiz: The quiz after watching a learning video on the spatial interface. (c) Conversational Agent: A screenshot of user interaction with a conversational AI for reflective writing based on previous learning experience on spatial interface. For the laptop interface: (d) Learning Quiz: The quiz after watching a learning video on the laptop. (e) Quiz Feedback: The laptop interface provides quiz results and feedback. (f) Conversational Agent: A screenshot of user interaction with a conversational AI for reflective writing based on previous learning experience on the laptop.

B Thematic Analysis

Table 1: Examples of Participants' Attitudes Toward Spatial Learning Categorized by Theme. The table presents participants' attitudes toward spatial learning, organized into four main categories. Each category is further divided into subcategories, with corresponding quotes provided for each subcategory.

Category	Subcategory	Example quote
Impact on learning process	Impact on memorization	*"I guessed some quiz on Vision Pro and got them correct. I think that's because of unconscious learning." (P7)*
	Impact on reflection	*"I think Apple Vision Pro is better for reflection because it's more quiet and creative. It helps me think more." (P18)*
	Focus and attention	*"I was more focused there, maybe the new impression and more it's interesting." (P14)*
Impact on cognitive process and learning outcome	Quiz results	*"I don't think the results depend on the technology." (P1)*
	Content absorption	*"(I) prefer video over text content (in spatial interface)." (P5)*
	Memorization and recall	*"It's much more enhanced with the audio." (P15)*
Devices usability and user experience	Comfort and physical usability	*"Vision Pro is easy and faster for task 1 than laptop, but in task 2 the virtual keyboard hard to use. And laptop is easier to take notes." (P18)*
	Initial setup experience	*"I thought it would be more complicated, but it's like it seems like easy to just like to look at something and click it with your finger." (P18)*
	Navigation and interaction	*"Wish more friendly to normal people, no high degree required to use." (P6) "It (the navigation) was easier than I thought." (P12)*
	Engagement and immersion	*"the Apple Vision Pro was better for me because I felt like I was exactly almost inside the video." (P8)*
	Enjoyment & Satisfaction	*"with the Apple Vision Pro, you feel like you're participating and it's more enjoyable." (P8)*
	Adaptation over time	*"the second time was easier because I already knew how to use it, like the calibration." (P9)*
	Input modality	*"I think on the Apple Vision Pro, it's a bit harder to type. it was definitely easier to chat with it on the laptop." (P8)*
	Interface design and aesthetics	"*I think you could probably do a lot with the Apple Vision Pro compared to the computer because I mean, you can use all the space rather than just a screen." (P9)*
	Perceived message efficiency	"*It was definitely faster to chat (by typing) with it on the laptop.*" (P9)
Future considerations and broader implications	Suggested improvements	*"If they make it like sunglasses or just some really like just glasses that you put on that could be cool. " (P12)*
	Health concerns	*"Vision will be gone much faster than they used books and laptops." (P12)*
	Ethical concerns	*"On an iPad or on a TV, I see what my child is watching, but on the Vision Pro, I don't know." (P4)*
	Barriers to adaption	*"It's heavy to wear for a long time. And it's very strange to see all the images and stuff." (P16)*
	Potential use cases	*"I could potentially see myself watching a video and taking notes next to it. ... It's hard to use several screens on laptop because of the small size." (P9)*

C Overview Behavioral Details

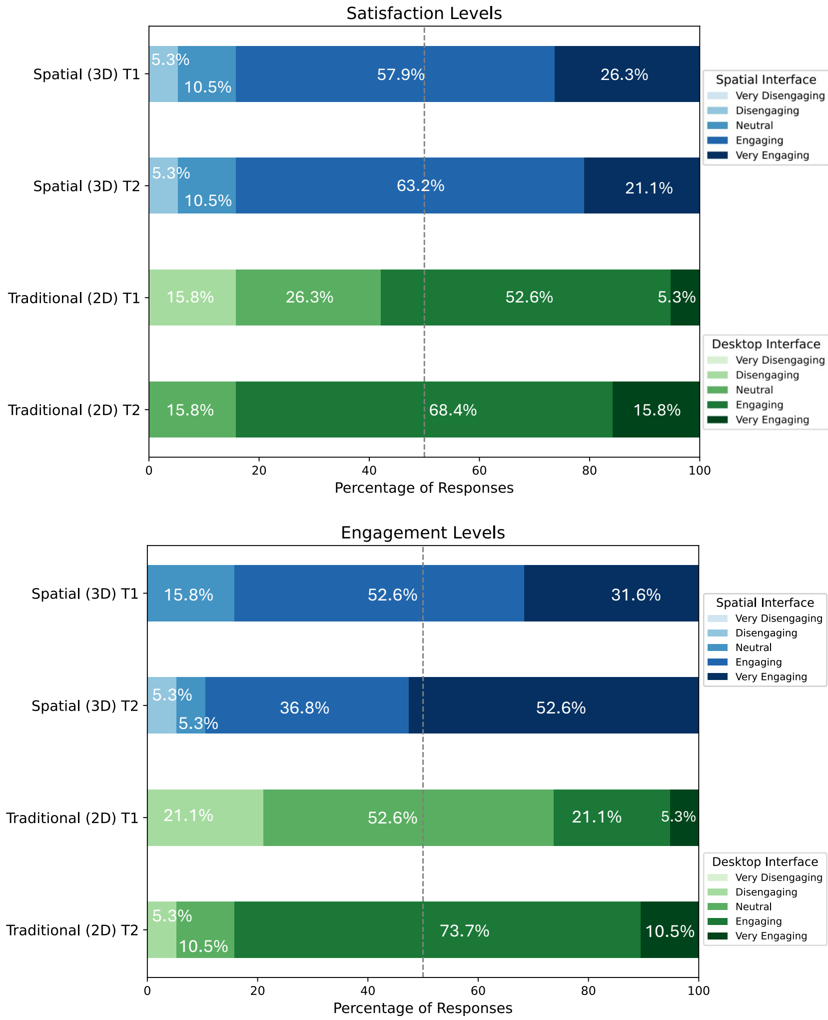

Figure 4: Comparison of satisfaction and engagement scores between the two environment. Blue bars represent the percentage distributions for the Spatial 3D interface, while green bars represent those for the conventional 2D desktop interface. "T1" denotes the video-watching learning task, and "T2" represents the reflective writing task.

Table 2: Mean values of message count, average text length, total time, and efficiency for laptop interface and spatial interface (AVP). AVP refers to Apple Vision Pro. The table format presents the data as Mean (standard deviation).

	Message Count	Average Text Length	Total Time	Normalized_Efficiency
Laptop	9.37 (2.87)	30.14 (22.06)	752.42 (261.42)	0.34 (0.26)
AVP	7.47 (1.81)	28.66 (17.11)	591.74 (204.26)	0.35 (0.22)

D Prompt Used for GPT-4

chat_system_prompt = """You are an instructor helping your student with reflecting on their previous learning experience (which was watching an instructional video), using the Gibbs reflective cycle. The Gibbs cycle is a model of reflection that encourages students to think systematically about the experiences they had during a specific situation, event, or activity. The model has six stages: Description, Feelings, Evaluation, Analysis, Conclusion, and Action Plan. Details of the classes, as well as example questions for each class, are as follows:

- ****Description**** of the experience: Kick off by setting the scene. What happened? Think of it like you're telling a friend about a movie plot but keep it brief. - Example questions:

(1) Can you briefly describe your experience in task 1?
(2) Can you describe what happened in the video you watched?
(3) What were the main points or events?

- ****Feelings**** and thoughts of the experience: Now, dive into how you felt about the situation. This isn't just "I felt sad" or "I was happy." Try to explore those feelings a bit more—what sparked them? - Example questions:

(1) How did you feel while watching the video?
(2) Were there any parts that stood out to you emotionally?
(3) What were you thinking during the video?

- ****Evaluation**** of the experience, both good and bad: Here, you weigh the pros and cons. What went well? What didn't? It's a bit like reviewing a restaurant; not everything might be perfect, but there could be some highlights! - Example questions:

(1) How do you evaluate the previous situation?
(2) Do you think it's helpful or not?
(3) Have you learned something from this phase?
(4) Which part could be improved?
(5) What did you think was good or bad about the video?
(6) Did any part of it particularly resonate with you or fail to make an impact?

- ****Analysis**** to make sense of the situation: Get your detective hat on! Why do you think things turned out the way they did? Look for the cause-and-effect relationships. - Example questions:

(1) Why is the material good? Or why not?
(2) What previous knowledge helped you in this learning process?
(3) Why do you think you felt that way about certain parts of the video?
(4) What factors influenced your reactions and thoughts?

- ****Conclusion**** about what you learned and what you could have done differently: Reflect on what you learned. Could you have done something differently? What skills did you pick up? - Example questions:

(1) What did you learn from this learning experience?
(2) What else do you need to make this learning process more effective?
(3) How might these changes improve the experience or the message?

- ****Action Plan**** about how you would deal with the similar situations in the future, or general changes you might find appropriate: Finally, think about the future. If this situation pops up again, what will you do? It's like planning your moves in a chess game. - Example questions:

(1) When you face a similar learning situation next time, what could you do differently?
(2) How will you develop your skills to make sure you could do differently next time?
(3) If you were to watch a similar video again, what would you do differently?
(4) How will you apply what you've learned from this reflection in future similar situations?

You should start by teaching the student about the Gibbs cycle, its benefits, the meaning of each component, and examples and best practices for each component. Then, whenever the student says they are ready to start reflecting, ask them sequential questions from the example questions provided above, to help them reflect on their experiences and guide them through the Gibbs cycle, component by component, until they form their reflection. Each of your messages should be very concise and short, maximum one paragraph long and not more than 150 words."""