



Chatting with a Learning Analytics Dashboard: The Role of Generative AI Literacy on Learner Interaction with Conventional and Scaffolding Chatbots

Yueqiao Jin
Monash University
Clayton, Australia
ariel.jin@monash.edu

Kaixun Yang
Monash University
Clayton, Australia
Kaixun.Yang1@monash.edu

Lixiang Yan
Monash University
Clayton, Australia
lixiang.yan@monash.edu

Vanessa Echeverria
Monash University
Clayton, Australia
Escuela Superior Politécnica del
Litoral
Guayaquil, Ecuador
Vanessa.Echeverria@monash.edu

Linxuan Zhao
Monash University
Clayton, Australia
Linxuan.Zhao@monash.edu

Riordan Alfredo
Monash University
Clayton, Australia
riordan.alfredo@monash.edu

Mikaela Milesi
Monash University
Clayton, Australia
Mikaela.Milesi@monash.edu

Jie Xiang Fan
Monash University
Clayton, Australia
jie.fan@monash.edu

Xinyu Li
Monash University
Clayton, Australia
xinyu.li1@monash.edu

Dragan Gasevic
Monash University
Clayton, Australia
dgasevic@acm.org

Roberto Martinez-Maldonado
Monash University
Clayton, Australia
Roberto.MartinezMaldonado@monash.edu

Abstract

Learning analytics dashboards (LADs) simplify complex learner data into accessible visualisations, providing actionable insights for educators and students. However, their educational effectiveness has not always matched the sophistication of the technology behind them. Explanatory and interactive LADs, enhanced by generative AI (GenAI) chatbots, hold promise by enabling dynamic, dialogue-based interactions with data visualisations and offering personalised feedback through text. Yet, the effectiveness of these tools may be limited by learners' varying levels of GenAI literacy, a factor that remains underexplored in current research. This study investigates the role of GenAI literacy in learner interactions with conventional (reactive) versus scaffolding (proactive) chatbot-assisted LADs. Through a comparative analysis of 81 participants, we examine how GenAI literacy is associated with learners' ability to interpret complex visualisations and their cognitive processes during interactions with chatbot-assisted LADs. Results

show that while both chatbots significantly improved learner comprehension, those with higher GenAI literacy benefited the most, particularly with conventional chatbots, demonstrating diverse prompting strategies. Findings highlight the importance of considering learners' GenAI literacy when integrating GenAI chatbots in LADs and educational technologies. Incorporating scaffolding techniques within GenAI chatbots can be an effective strategy, offering a more guided experience that reduces reliance on learners' GenAI literacy.

CCS Concepts

• **Applied computing** → **Interactive learning environments.**

Keywords

learning analytics dashboard, generative AI literacy, generative AI chatbots, data visualisation



This work is licensed under a Creative Commons Attribution International 4.0 License.

LAK 2025, Dublin, Ireland

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0701-8/25/03

<https://doi.org/10.1145/3706468.3706545>

ACM Reference Format:

Yueqiao Jin, Kaixun Yang, Lixiang Yan, Vanessa Echeverria, Linxuan Zhao, Riordan Alfredo, Mikaela Milesi, Jie Xiang Fan, Xinyu Li, Dragan Gasevic, and Roberto Martinez-Maldonado. 2025. Chatting with a Learning Analytics Dashboard: The Role of Generative AI Literacy on Learner Interaction with Conventional and Scaffolding Chatbots. In *LAK25: The 15th International Learning Analytics and Knowledge Conference (LAK 2025), March 03–07, 2025, Dublin, Ireland*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3706468.3706545>

1 Introduction

As educational technology rapidly evolves, learning analytics dashboards (LADs) are gaining increasing attention for their potential to transform education by leveraging data to enhance learning and teaching practice [55, 62]. These dashboards generally aim to simplify complex learner data into more accessible visualisations, offering educators, students, and other stakeholders insights that can potentially be meaningful and actionable [31, 51]. By presenting information in a more visually digestible format, LADs aim to help learners and educators track progress [28], identify areas of strength and weakness [73], support informed decision-making [60], and assist with timely interventions [50].

Yet, despite significant advancements in learning analytics, and the development of LADs using state-of-the-art technologies, such as artificial intelligence (AI) and multimodal sensory data [46], their educational *effectiveness* has not always been matched with the sophistication of these technologies [28]. Indeed, recent concerns have been raised about whether the increasing complexity of educational data, driven by advancements in data collection and processing technologies, may overwhelm learners and educators [13, 15]. This complexity necessitates a growing need to ensure that visualisations remain comprehensible and do not contribute to cognitive overload [53].

A potential pathway to enhancing learners' comprehension of complex visualisations is the use of explanatory and interactive analytics that could serve to scaffold students' interactions with LADs [15, 69]. The rapid advancements in generative artificial intelligence (GenAI) tools, now equipped with multimodal capabilities, have made it increasingly feasible to utilise GenAI-powered conversational agents, commonly referred to as GenAI chatbots, to enrich LADs with interactive features and to transit from exploratory (focused on identifying patterns and generating hypotheses) to explanatory (aimed at explaining causes and providing deeper insights) analytics [16, 48, 55]. Studies have demonstrated that these chatbots can support learners' comprehension and help reduce their cognitive load by providing on-demand, real-time, and personalised explanations and feedback [33, 47]. In addition, by integrating retrieval-augmented generation (RAG) methodologies, LADs can evolve from a one-way information delivery system to a conversational platform [62, 69]. This transformation may allow learners to communicate and collaborate with the dashboard, access underlying resources and receive tailored guidance. Such interaction holds the potential to foster a more engaging and supportive learning environment, enabling learners to ask questions, seek clarifications, and receive immediate, contextually relevant responses [62]. Additionally, by incorporating an agentic system design in LADs [72], *conventional chatbots* (that passively answer learners' questions) can also be transformed into *scaffolding chatbots* (that actively guide learners through the key components of LADs), potentially enhancing their understanding and facilitating learning [21, 32].

While the integration of various types of GenAI chatbots into LADs holds great promise, the extent to which students can engage fairly and effectively with these tools largely depends on their level of GenAI literacy [3]. GenAI literacy refers to the ability to understand, interact with, and effectively use GenAI-driven tools

and technologies [4]. As these tools become more prevalent in educational settings, it is crucial to consider how varying levels of GenAI literacy among learners might influence their ability to benefit from GenAI-powered conversational agents [33]. For instance, learners with high GenAI literacy are likely to navigate and benefit from chatbot-assisted LADs more efficiently, making the most of the personalised feedback and real-time interactions these tools offer [4, 10]. Conversely, learners with lower levels of GenAI literacy may struggle to fully engage with *conventional chatbots* (those which reactively respond to questions, such as regular ChatGPT), potentially leading to frustration or underutilisation of the available resources if they are unsure how to formulate questions or unaware of the types of queries the agent can effectively respond to [3, 39]. In contrast, a *scaffolding chatbot* (which actively prompts learners with guiding questions) might lower the dependence on their GenAI literacy, considering the evident effectiveness of scaffolding techniques [32]. Consequently, it is essential to consider learners' GenAI literacy when integrating GenAI-powered learning interventions.

Despite the emergence of many position pieces that emphasise the benefits and challenges of GenAI in learning analytics and LADs [14, 24, 30, 69], there is a notable gap in empirical research examining the actual effectiveness of different types of chatbot-assisted LADs (e.g., conventional versus scaffolding chatbots) in enhancing learners' comprehension of learning insights. Likewise, the impact of GenAI literacy on learners' comprehension when using chatbot-assisted LADs also remains underexplored. Additionally, there is limited knowledge about the cognitive processing patterns exhibited by learners with varying GenAI literacy levels when interacting with these chatbots.

To address these gaps, this paper presents a study involving 81 learners, comparing the effectiveness of two types of chatbot-assisted LADs—*conventional/reactive* and *scaffolding/proactive*—to enhance learners' comprehension of complex visualisations. We then investigated the association between learners' GenAI literacy and their comprehension when interacting with these GenAI chatbots. It also examines how learners' GenAI literacy relates to their comprehension and cognitive processing patterns during interactions with these chatbots. The research aims to provide empirical insights into the role of GenAI literacy in effectively using chatbot-assisted LADs, informing the design of more inclusive educational technologies for all learners.

2 Background

2.1 Learning Analytics Dashboard and Visualisations

LADs are instrumental in aiding educational stakeholders in the ongoing assessment of teaching and learning methodologies [55]. Comprising visual depictions and analyses of educational data, LADs encourage users to identify trends and address queries within the data [62]. Nevertheless, LADs face significant constraints, including the presentation of intricate visualisations, ineffective insight communication [13], and an incongruity with teachers' pedagogical requirements [29]. To mitigate these challenges, the focus of research has pivoted to the development of *explanatory LADs*.

These are tailored to visually guide users to crucial insights contextualised with information from the learning design, thereby aligning more closely with the pedagogical needs of stakeholders [15]. Embracing information visualisation techniques and data storytelling principles, these explanatory LADs utilise clear titles to communicate main messages, narratives with text explanations, and visual markers such as arrows and colour-coded information to highlight specific data points [15, 18]. Studies have demonstrated that these dashboards effectively direct attention to significant pedagogical insights [15, 41], promote profound reflective practice, and assist both teachers and students in comprehending their learning processes [15, 18]. Furthermore, they are particularly beneficial for teachers with limited visual literacy, facilitating an easier understanding of the information presented [52]. Such augmentations also effectively enhance learners' comprehension of the key insights compared to conventional visualisations [57]. Despite these advantages, a notable challenge remains: the automation of translating contextual information from learning activities into visual elements. This task frequently involves a co-design process where designers, researchers, and educators collaboratively define pedagogical intentions and align them with visual elements [16, 18, 41, 52]. Recent advancements have attempted to simplify this process by having educators specify pedagogical rules, although the actual translation into visual elements continues to rely on designers and researchers [19]. In parallel, other solutions are beginning to utilise unsupervised learning and natural language processing (NLP) to automatically generate student narratives, thus enabling academic advisors to better understand student success and risk behaviours [2, 36].

2.2 GenAI in Education

Recent strides in GenAI have markedly expanded the horizons for creating educational content and interacting with educational data [7, 68]. Large Language Models (LLMs) such as GPT and LLaMA have shown potential in autonomously generating instructional materials and practice quizzes, ushering in a new era in educational content creation [14, 30, 71]. Concurrently, innovations like OpenAI's Whisper have improved the accessibility of learning materials by enabling automatic transcription of lecture recordings, catering to a variety of preferences and needs [23]. Additionally, image-generating diffusion models such as Midjourney and DALL-E have enriched multimedia learning and fostered creativity in art-focused STEAM education by translating text into compelling visuals [35]. The advent of multimodal models like GPT-4o, which adeptly interpret both textual and visual information, represents a significant leap forward in providing contextually relevant explanations [1]. These advancements point to a future where the communication of insights to learners and educators is increasingly interactive, varied, and effective [69].

Incorporating multimodal GenAI into LADs holds promise for improving explanatory capabilities through interactive dialogues [19, 72]. Although chatbots have been used in various educational settings, such as intelligent tutoring systems and feedback mechanisms [47], their integration within LADs has faced challenges due to the difficulty of achieving high-quality, zero-shot understanding of visual language [54]. Effective explanations in LADs necessitate

contextual grounding within specific learning designs to support decision-making and encourage deep reflection [55, 62], which calls for advanced multimodal AI models capable of comprehending both natural and visual language. Moreover, the provision of accurate and contextually relevant explanations by GenAI in LADs relies on robust supporting infrastructures. Despite their potential, GenAI models often produce well-articulated but inaccurate outputs, a phenomenon known as "hallucination" [25]. To mitigate this, RAG techniques, which focus content generation on contextually pertinent materials [58], have gained traction. This involves encoding relevant information into vector embeddings that encapsulate semantic or relational meanings [44], allowing for dynamic retrieval during AI interactions through semantic searches using metrics like cosine similarity, thereby reducing hallucinations and enhancing domain-specific accuracy [59]. Furthermore, GenAI models can produce irrelevant responses if user prompts lack sufficient detail [64]. Since expecting all learners and educators to master the skill of effective prompt crafting is impractical, the deployment of AI agents, autonomous entities that refine prompts to achieve specific goals, emerges as a promising solution [65]. By integrating multimodal GenAI with RAG and autonomous AI agents, it becomes feasible to develop context-aware chatbots for LADs that deliver precise, contextually grounded explanations of visualisations, thereby markedly enriching the educational experience for both learners and educators.

LADs integrated with agentic chatbots embedded with scaffolding techniques can be promising to deliver a more structured and engaging learning experience [49, 65]. Scaffolding, an educational strategy with a strong theoretical grounding, entails decomposing complex information into manageable chunks, posing guiding questions, and providing feedback to enhance learner comprehension of a specific subject [21]. This can promote self-regulated learning by fostering metacognitive processes, encouraging monitoring behaviours, potentially improving learning outcomes [37, 63]. Empirical research underscores its effectiveness in aiding learners to achieve a deeper understanding and mastery by offering contextual guidance during the knowledge and skill development process [21, 32]. Therefore, the integration of scaffolding chatbots into LADs could potentially enhance learners' comprehension more efficaciously than conventional chatbots that simply respond to queries. While this method is promising and innovative system frameworks are currently under development [40, 72], empirical evidence on the efficacy of these chatbots in improving learners' understanding of key insights presented within LADs remains limited. This gap informs our first research question: **RQ1:** To what extent do conventional and scaffolding GenAI chatbots enhance learners' *comprehension* of key insights presented by visualisations in LADs?

2.3 GenAI Literacy

GenAI literacy is crucial to understanding how students interact with and benefit from different types of chatbots in educational settings. While AI literacy broadly encapsulates competencies such as critically evaluating AI technologies, communicating and collaborating effectively with AI, and using AI tools ethically in various

contexts [38, 45], GenAI literacy addresses the specific skills required for engaging with generative models like ChatGPT [75]. It is important to distinguish between general AI literacy and GenAI literacy, as existing frameworks often fail to cover the specific competencies needed to effectively utilise GenAI tools [4]. For instance, Zhao et al. [75] suggests that GenAI literacy should include pragmatic, safety, reflective, socio-ethical, and contextual understandings. Bozkurt [10] further highlights that GenAI literacy should adopt an adaptable approach, allowing individuals to understand generative AI based on their needs and enabling them to grow alongside its rapid evolution.

Current research indicates that although GenAI tools can significantly improve student learning outcomes, the quality of student-generated prompts often falls short, pointing to a pressing need for enhanced GenAI literacy [39]. Additionally, despite the attempts made to integrate GenAI into educational frameworks, there is a notable lack of research examining how students' GenAI literacy impacts their interaction with GenAI-powered tools within LADs [12]. This research gap is particularly significant given the potential of GenAI chatbots to offer rich, contextually relevant explanations that could substantially improve students' comprehension of complex visualisations in LADs. Understanding these interactions could reveal valuable insights into designing more effective GenAI chatbots for LADs. This would ensure that all students, regardless of their initial literacy levels, can fully benefit from these advanced educational technologies, therefore enhancing learners' comprehension of key insights conveyed through LADs. By understanding how varying levels of GenAI literacy impact interaction with GenAI chatbots, educators and researchers can create more equitable learning experiences and ensure that complex learning analytic visualisations are accessible and understandable to all students. This motivates our second and third research questions: **RQ2**: To what extent is learners' GenAI literacy associated with their *comprehension* of complex analytical visualisations when using a conventional GenAI chatbot compared to a scaffolding GenAI chatbot? **RQ3**: To what extent do learners' *interactions* vary based on their level of GenAI literacy when interacting with the conventional and scaffolding chatbots?

3 Methods

3.1 Learning Context

The context and the dataset used to generate visualisations were collected from a healthcare simulation where students, divided into teams of four, played the roles of two primary (blue and red) and two secondary (green and yellow) nurses. The high-fidelity simulation involved a ward set-up with hospital beds, medical equipment, and advanced patient manikins controlled by teachers to simulate varied heart rates and pulses. The simulation aimed to enhance teamwork, communication, and prioritisation skills in clinical emergencies. Multimodal data collected included positioning (x-y coordinates and body orientation via an indoor positioning system), audio (captured by wireless headset microphones), and heart rate (recorded by FitBit Sense wristbands).

The multimodal LAD, featuring three visualisations, was developed and validated in previous studies to aid students' post-simulation reflection [42]. In prior evaluations, students reported

difficulties in navigating and understanding the complex visualisations [67]. Consequently, these visualisations offered a chance to explore how GenAI chatbots could enhance comprehension in this study. The three visualisations, as shown in Figure 1, vary in complexity: one modality (positioning) for the bar chart, two (positioning and audio) for the communication network, and three (positioning, audio, and heart rate) for the ward map.

The **bar chart** depicted prioritisation strategies based on students' positional data, allowing easy comparison of time spent on different behaviours and understanding of resource allocation during the simulation [70]. The **communication network** or sociogram, mapped interaction patterns using positional and audio data, revealing communication frequencies and directions, including interactions with the patient, doctor, and relative, thus highlighting participant roles and engagement levels [74]. The advanced **ward map** combined students' physical positions, verbal communication duration, and peak heart rate locations. Inspired by sports analytics [22], it used a heatmap to show the frequency and distribution of verbal communications and mapped students' spatial distribution, providing a comprehensive view of physical and behavioural engagement. It also indicated where students experienced the highest physiological arousal by displaying peak heart rate locations and values.

3.2 Design of GenAI chatbots

To explore how different chatbot designs can support learners with varying levels of GenAI literacy, we developed both conventional and scaffolding GenAI chatbots. These chatbots were developed based on the open-source prototype of VizChat [72]. The conventional and scaffolding chatbots were distinguished by their core functionalities and interaction styles (Figure 2). *Conventional* chatbots were reactive, delivering precise and contextually relevant information in response to user-initiated queries [40, 72], without initiating unsolicited interactions. In contrast, *scaffolding* chatbots were more proactive [49], guiding users through visualisations with structured narratives and scaffolding questions crafted by human experts (available in *repository*).

Prompt integration is crucial for both types: conventional chatbots processed user queries to generate relevant responses through the interpretation of natural language input and matching it with visualisation and knowledge databases using RAG. Scaffolding chatbots further synthesised prompts based on user interactions and a pre-defined narrative, facilitating exploration with guiding questions and real-time feedback. Both utilised a **knowledge database** built using LangChain and Chroma, populated with task-specific materials converted into vector embeddings via OpenAI's embedding model (text-embedding-ada-002) [20]. **Response generation** leverages GPT-4o's multimodal capabilities: conventional chatbots generated precise answers from semantic searches of the knowledge database, while scaffolding chatbots provided feedback on prior responses, further guiding users through the process step-by-step. For instance, to aid learners in grasping the intricacies of a social network, a scaffolding chatbot could initially pose questions regarding the size of the nodes. Once it had verified that users had a clear understanding through their responses, the chatbot would proceed to query about the edges connecting the nodes. This scaffolding approach, exemplified by guided questions on visualisation

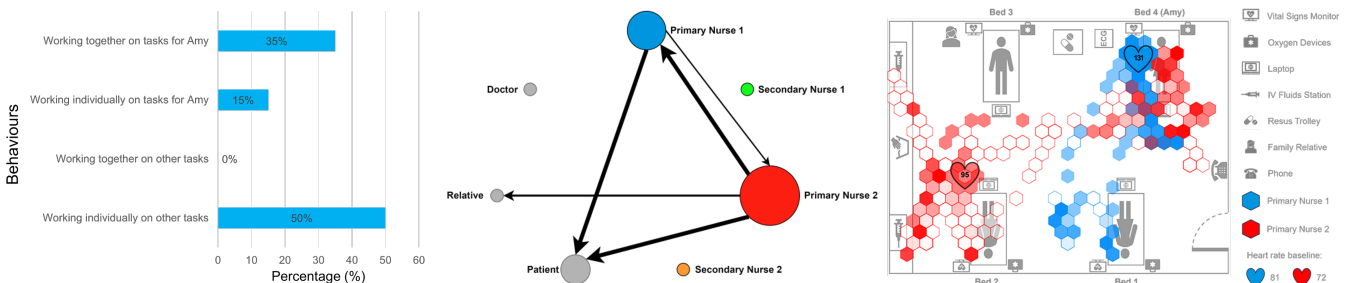


Figure 1: Three LAD visualisations, representing the activity of two nurses in the first phase of their simulation, used in the current study (from left to right): a bar chart, representing the extent of time dedicated to particular tasks; a communication network, representing the amount of conversation among team members; and a ward map showing location, speech and highest heart rate. Enlarged figures are available in our repository.

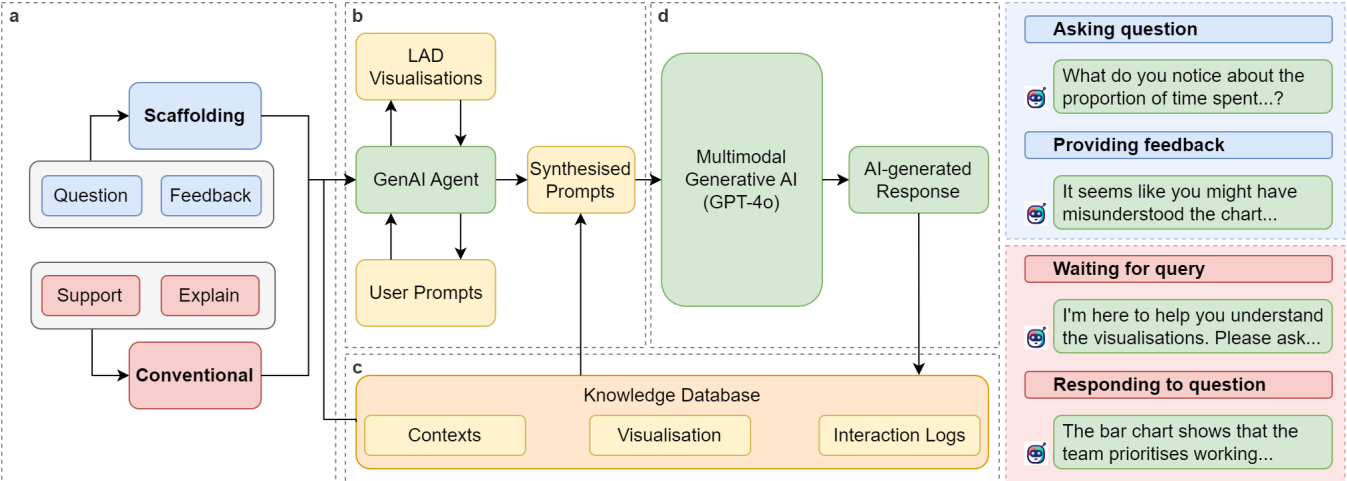


Figure 2: System design of conventional and scaffolding Generative AI (GenAI) chatbots, highlighting four main components: a) distinctive traits separating the conventional chatbot from the scaffolding one, b) interaction dynamics between user prompts, LAD visualisations, and the GenAI agent leveraging retrieval-augmented generation for prompt synthesis, c) a knowledge database with essential contextual information for learning tasks, d) generation of contextually relevant responses using multimodal GenAI. Examples of behaviours: the conventional chatbot (red) focuses on support and explanations, while the scaffolding chatbot (blue) offers guided questions and feedback.

elements [66], aimed to enhance learners' comprehension of key insights presented by visualisations in LADs.

3.3 Study design

We conducted a 2x2 mixed-method experiment to evaluate the role of GenAI literacy on learner comprehension across baseline and intervention phases. This design also allowed us to analyse learners' cognitive processes when interacting with two types of GenAI chatbots: conventional and scaffolding. The experiment was administered via Qualtrics¹, and participants were recruited through Prolific². Our sample consisted of current and graduated medical or nursing students with no prior knowledge of the learning context, ensuring improvements were due to the interventions rather than pre-existing familiarity. The study was designed to be completed within an hour, and participants were compensated £8 for

¹<https://www.qualtrics.com/>

²<https://www.prolific.co/>

their time. Ethics approval was obtained from Monash University (Project Number: 37307), with informed consent from each participant. The study consisted of three main components. The details of each component are elaborated upon as follows:

3.3.1 Part 1: Demographics and background questions. First, participants were asked to provide demographic information, including (i) self-reported gender, (ii) age range, (iii) region, and (iv) highest level of education. Following this, they rated their experience in data analysis and GenAI tools on a 5-point Likert scale, ranging from "None" (1) to "Expert" (5).

3.3.2 Part 2: Generative AI Literacy Test. We used a 20-item instrument, the Generative AI Literacy Assessment Test (GLAT) [26], to measure GenAI literacy. Experts reviewed and revised the instrument to ensure face validity, and its content validity was assessed with 204 individuals for relevance, comprehensiveness, and comprehensibility. Structural validity and internal consistency were

confirmed using classical test theory and item response theory, resulting in a reliable 2PL model (Cronbach's Alpha = 0.76) and a robust factor structure ($\chi^2(133) = 135.06, p > .05, \text{RMSEA} = 0.007, 90\% \text{ CI } [0, 0.035]$). We applied the correction-for-guessing formula [61] to account for guessing, as the 2PL model considers only item difficulty and discrimination. The test can be accessed in the *repository*.

3.3.3 Part 3: Comparative Study. Each participant completed three activities in order: i) reading contextual information about the visualisations, ii) completing a *baseline* analytical writing task and six evaluation questions, and iii) being randomly assigned to one of two *intervention* groups—conventional GenAI chatbots or scaffolding GenAI chatbots—to complete an intervention analytical writing task and six evaluation questions. Each participant viewed two dashboards for the baseline and intervention. The dashboards were identical except for the addition of the chatbot component in the intervention phase.

The types of visualisations (e.g., bar charts, communication networks, ward maps) remained consistent across both phases, although the specific data and insights varied. The six evaluation questions also maintained a consistent format but were customised to match the insights from each set of visualisations. Additional details on the contextual information, writing tasks, and evaluation questions are provided below.

Contextual information. Participants were provided with a detailed description of a healthcare simulation scenario. In the scenario, two nursing students managed four manikin patients (Beds 1-4) during their shift, focusing primarily on the deteriorating patient in Bed 4 (Amy) while the needs of the other patients also required prioritisation. An actor, playing a relative of the patient in Bed 3, frequently attempted to distract the nurses. The simulation aimed to develop students' skills in teamwork, communication, and prioritisation. This contextual information equipped the participants with necessary background information to understand the visualisations and perform the subsequent writing tasks.

Analytical writing tasks. After understanding the contextual information, participants were directed to a custom-built website for the analytical writing tasks, featuring a Learning Analytics Dashboard (LAD). They were asked to analyse three visualisations and wrote a 100-150 word response on *how the two nurses managed the primary patient (Amy) while attending to other beds, focusing on their task prioritisation, verbal communication, and stress levels*. The dashboard contained three components: a **Display Component**, showing one visualisation at a time to avoid cognitive overload; a **Writing Space** for articulating analysis; and an **Instructions Component** with standardised task instructions for the baseline. During the intervention phase, this component was replaced by a **GenAI Interaction** component. This allowed participants to interact with either conventional or scaffolding GenAI chatbots, depending on their assigned conditions. The task aimed to immerse learners in visual analytics, allow them to comprehend insights, and interact with the chatbot intervention.

Evaluation questions. After each analytical writing task, participants answered six multiple-choice questions to assess comprehension of the visualisations. Each visualisation was paired with two questions, designed based on Bloom's taxonomy Levels

1 (knowledge) and 2 (comprehension) [8]. Knowledge questions assess information retrieval by identifying specific data points or patterns, e.g., *Which behaviour did the two nurses spend the least time on?* Comprehension questions evaluated interpreting multiple insights and identifying inconsistencies, e.g., *How did the nurses spend their time working on tasks for Amy compared to other tasks?* Higher taxonomy levels were excluded due to limited contextual knowledge. Two researchers designed the questions, while a third validated them for applicability across visualisation sets. Discrepancies were resolved through discussion. The question structure remained constant, with randomised answer choices to minimise bias [5], and included an 'I am not sure' option to reduce guessing [34]. An example of the first set of evaluation questions for the baseline condition is available at the provided [link](#).

3.4 Participants

A total of 81 participants took part in the study, meeting the required sample size (72; 36 per condition) based on a priori power analysis for a medium effect size (0.25) with 80% power at a 0.05 significance level. Participants were divided into two groups: 41 with a conventional chatbot and 40 with a scaffolding chatbot. All participants were current or graduated students with a medical or nursing background. Participants were from six regions: North/Central America (35), Europe (22), Africa (15), Australia (4), other regions (3), and South America (2). They identified as female (46) and male (35). Ages ranged mostly between 25-34 (32) and 18-24 (29), followed by 35-44 (11), 45-54 (6), 55-64 (2), and 65+ (1). Their education levels were mainly Bachelor's degrees (37) and high school or equivalent (19), with some holding Master's degrees (11), vocational training or diplomas (7), Doctorates (4), or other qualifications (3). Regarding data analysis experience, participants were mostly intermediate (30) or beginners (29), with fewer advanced users (12), no-experience participants (8), and experts (2). Familiarity with GenAI tools was mostly intermediate (44), followed by beginners (25), advanced users (9), experts (2), and one participant with no prior experience.

3.5 Data Analysis

First, we defined and calculated three metrics to investigate the association between GenAI literacy and learners' comprehension of complex visualisations using two types of chatbots: a conventional GenAI chatbot and a scaffolding GenAI chatbot. The first metric, **Comprehension_score**, represents the total number of correct answers out of six evaluation questions, ranging from 0 to 6, with each participant completing two sets of questions to yield *Baseline_score* and *Intervention_score*. The **Improvement_score** quantifies the change in comprehension by calculating the difference between the *Intervention_score* and the *Baseline_score*. The **GenAI_literacy** metric is the sum of correct responses on the GenAI literacy test, ranging from 0 to 20. For both the comprehension score and GenAI literacy score, the correction-for-guessing formula [61] was applied to adjust for guessing in these multiple-choice questions with four response options. All calculations and analyses were conducted in Python using libraries such as NumPy, SciPy, and Statsmodels. The next sections provide analysis details for each RQ.

3.5.1 Preliminary Analysis. A preliminary analysis was conducted to ensure the comparability of the two intervention conditions in

terms of participants' GenAI literacy, data analysis skills, and GenAI expertise (rated on a scale from 1–None to 5–Expert). This step was crucial to guarantee the robustness and reliability of the subsequent findings [11]. To evaluate whether there were statistically significant differences across the groups, we used the Mann-Whitney test. Additionally, we conducted an epistemic network analysis (ENA) to examine the differences in interactions between participants and GenAI chatbots across the two intervention conditions. This analysis helps determine whether the two groups can be combined into a single dataset for subsequent ENA to address RQ3. Further methodological details of the coding scheme and the ENA procedure are elaborated in section 3.5.4.

3.5.2 RQ1 – Within-subject Analysis. To address RQ1, we conducted within-subject analyses to examine changes in learners' comprehension scores for each intervention. We used Wilcoxon signed-rank tests to compare the statistical differences between learners' comprehension scores at baseline and during the intervention, separately for the conventional and scaffolding chatbots. A Mann-Whitney *U* test was conducted to compare the *Improvement_score* of the two chatbots. Effect sizes were calculated using Rank-Biserial correlation.

3.5.3 RQ2 – Regression Analysis (GenAI literacy). For RQ2, we used ordinary least squares (OLS) regression to examine the relationship between learners' *GenAI_literacy* (independent variable; IV) and their *Improvement_score* (dependent variable; DV). To control for initial comprehension levels, we included the *Baseline_score* as an additional IV.

Formula: $Improvement_score = \beta_0 + \beta_1 \times Baseline_score + \beta_2 \times GenAI_literacy + \beta_3 (Baseline_score \times GenAI_literacy)$
This formula included an intercept (β_0), main effects for *Baseline_score* (β_1) and *GenAI_literacy* (β_2), and an interaction term (β_3) between *Baseline_score* and *GenAI_literacy*. This interaction term helps understand how learners with different levels of GenAI literacy benefit from each intervention. We conducted an ANOVA to decide whether to include the interaction term by comparing models with and without it. The regression analyses were conducted separately for the conventional and scaffolding chatbots to better understand the effects of GenAI literacy on each intervention. The models were fitted using the OLS function from the Statsmodels library, and the assumptions were validated through various methods. To assess linearity, we plotted the predicted values against the observed values. We checked the normality of residuals with the Shapiro-Wilk test and QQ plots. Homoscedasticity was evaluated using the Breusch-Pagan test, and the Durbin-Watson test was used to assess the independence of residuals. All assumptions were met.

3.5.4 RQ3 – Epistemic Network Analysis. Coding Scheme. To analyse user-chatbot interactions within the context of information and cognitive processing, we developed a coding scheme based on Information Processing Theory (IPT) [6]. IPT aids in understanding how learners process, store, and retrieve information, allowing systematic categorisation and analysis of GenAI chatbot interactions. This approach helps identify how effectively chatbots manage cognitive load, enhance comprehension, and facilitate knowledge transfer. The coding scheme encompassed five constructs: *Information Seeking*, which captures instances where learners or chatbots seek to fill

knowledge gaps (input stage); *Clarification*, involving further explanation or rephrasing the information to facilitate comprehension (encoding); *Confirmation*, check the accuracy of their understanding or the information they have stored (verification and validation); *Integration*, integrating new information with existing knowledge (information storage and organisation); and *Reflection*, where individuals reflect on and evaluate processed information (retrieval and metacognitive processes). Each construct is divided into two codes: one for messages from learners (*User*) and one for messages from chatbots (*Chatbot*). Additionally, the code *User.Command* covers other learner commands, such as greetings, signing off, or simple navigation and control commands. Messages were coded at the utterance level, allowing for multiple codes per utterance. Two researchers initially coded 50% of the data. Cohen's kappa was used to measure inter-rater reliability, with a threshold of 0.6 for acceptance [43]. Each code achieved a kappa greater than 0.7, indicating strong agreement between coders. One researcher then completed the remaining coding. Users with no chatbot interaction or fewer than two utterances were excluded due to insufficient co-occurrence information. The code *Chatbot.Confirmation* was also removed due to minimal occurrences, resulting in a final set of 10 codes.

Epistemic Network Analysis. Epistemic Network Analysis (ENA) was conducted to examine the co-occurrence of interaction patterns between learners and GenAI chatbots, segmented by learners' GenAI literacy levels. Learners were categorised into low and high GenAI literacy groups using the Quantile function in Python's Pandas library, with the upper 50% labelled as high literacy and the remaining 50% as low literacy. This dichotomous grouping method follows practices in previous ENA studies to model differences in learning behaviours [17, 74]. For the conventional chatbot condition, the analysis included 16 learners with low literacy and 11 with high literacy, while the scaffolding chatbot condition comprised 19 learners with low literacy and 21 with high literacy. In our study, the unit of analysis was defined as the individual interaction episodes between learners and GenAI chatbots, with the utterances from both learners and chatbots treated as lines. We used a stanza window size of two lines to accumulate connections, as recommended for dialogue data analysis [56]. To maximise differences between the mean of the two groups along the x-axis, we used Means Rotation (MR) for dimensional reduction [9], facilitating the interpretation of resulting networks. Differences were illustrated by subtracting the mean networks of the two groups in comparison plots. We performed Mann-Whitney *U* tests to examine statistical differences in interaction patterns along the x and y axes, applying Bonferroni correction for multiple comparisons with an initial alpha value of 0.05.

4 Results

4.1 Preliminary Findings

The Mann-Whitney *U* test for learners' GenAI literacy showed a median score of 11.0 (IQR = 7.0) for the conventional group (*n* = 41) and 12.0 (IQR = 5.0) for the scaffolding group (*n* = 40), with no statistically significant difference, $U = 827, p = .95, r = .01$. Similarly, for learners' data expertise, both groups had a median score of 3.0 (IQR = 1.0), with no significant difference, $U = 881, p = .55, r = .06$. For learners' GenAI expertise, both groups also had median scores of 3.0 (IQR = 1.0), with no significant difference,

$U = 817, p = .98, r = .01$. These findings indicate that learners in both groups had comparable levels of GenAI literacy, data expertise, and GenAI expertise, supporting both within and between group analyses for RQ1. However, ENA revealed significant differences in learners' interaction patterns between the two groups (see Figure 3, left–below in Section 4.4). Specifically, a Mann-Whitney test showed that the conventional group ($Mdn = -0.23, N = 27$) was statistically significantly different from the scaffolding group ($Mdn = 0.21, N = 40, U = 90, p < .001, r = 0.83$), along the x-axis (MR1; 16.3%) but not the y-axis (SVD2; 21.1%). This finding suggests that it is important to analyse these two groups separately to understand specific differences between learners with low and high GenAI literacy (e.g., RQ2 & 3).

4.2 RQ1 – Comprehension Improvement

The Mann-Whitney U test for the conventional group indicated significant improvements from the *Baseline_score* (Median = 3.0, IQR = 1.0) to the *Intervention_score* (Median = 4.0, IQR = 1.0), $W = 71, p < .001, r = 0.87$. This large effect size suggests a substantial increase in comprehension scores due to the conventional GenAI chatbot. Likewise, the Mann-Whitney U test for the scaffolding group also showed significant improvements from the *Baseline_score* (Median = 3.0, IQR = 2.0) to the *Intervention_score* (Median = 5.0, IQR = 1.0), $W = 32, p < .001, r = 0.94$. This large effect size indicates a substantial improvement in comprehension scores attributed to the scaffolding GenAI chatbot. Additionally, there was no significant difference in the *Improvement_score* between the two chatbots ($U = 703, p = .14$), indicating that both chatbots improve learners' comprehension. These findings underscore the effectiveness of both the conventional and scaffolding chatbot interventions in enhancing learners' comprehension of complex visualisations.

4.3 RQ2 – GenAI Literacy

For the conventional group, the OLS regression model included the interaction term between *Baseline_score* and *GenAI_literacy*, which significantly improved the model's explanatory power ($\Delta F(1, 38) = 6.27, p = 0.017$). The model explained 44.5% of the variance in the *Improvement_score* ($F(3, 38) = 10.14, p < .001$). The intercept was significant ($\beta = 2.16, SE = 0.48, t = 4.51, p < .001$), indicating that, on average, learners exhibited an improvement score of 2.16 points when both their *Baseline_score* and *GenAI_literacy* were zero. The main effect of *Baseline_score* was also significant ($\beta = -0.36, SE = 0.17, t = -2.16, p = 0.037$), showing that a one-point increase in learners' *Baseline_score* resulted in a 0.36-point decrease in their *Improvement_score*. This negative relationship was expected due to the ceiling effect, given the maximum score is six points. The main effect of *GenAI_literacy* was significant as well ($\beta = 0.19, SE = 0.07, t = 2.84, p = .007$), indicating that a one-point increase in learners' *GenAI_literacy* resulted in a 0.19-point increase in their *Improvement_score*. This is substantial considering the different scales: the *GenAI_literacy* score ranged up to 20 points, while the *Improvement_score* had a mean of 1.1 points and a standard deviation of 1.2 points. Although the interaction term between *Baseline_score* and *GenAI_literacy* was significant ($\beta = -0.05, SE = 0.02, t = -2.51, p = 0.017$), the effect size was quite small. This indicates a statistically significant but minor negative

interaction between *Baseline_score* and *GenAI_literacy*. Specifically, for each additional point in *Baseline_score*, the positive effect of *GenAI_literacy* on the *Improvement_score* decreased by 0.05 points, which could also be attributed to the ceiling effect.

For the scaffolding group, the OLS regression model did not include the interaction term, as it did not significantly improve the model's explanatory power ($\Delta F(1, 37) = 1.02, p = 0.32$). The model explained 50.2% of the variance in the *Improvement_score* ($F(2, 38) = 19.12, p < .001$). The intercept was significant ($\beta = 3.52, SE = 0.46, t = 7.61, p < .001$), indicating that, on average, learners exhibited an improvement score of 3.52 points when both their *Baseline_score* and *GenAI_literacy* are zero. The main effect of *Baseline_score* was significant ($\beta = -0.88, SE = 0.14, t = -6.12, p < .001$), indicating that a one-point increase in learners' *Baseline_score* resulted in a 0.88-point decrease in their *Improvement_score*. This negative relationship is consistent with the ceiling effect, given the maximum score is six points. The main effect of *GenAI_literacy* was also significant, albeit with a relatively small effect size compared to the conventional group ($\beta = 0.08, SE = 0.03, t = 2.53, p = 0.016$). This indicates that a one-point increase in learners' *GenAI_literacy* resulted in a 0.08-point increase in their *Improvement_score*.

These findings suggest that higher GenAI literacy positively influences learners' improvement scores in both the conventional and scaffolding chatbot interventions, with a more substantial effect observed in the conventional group.

4.4 RQ3 – Cognitive Processing Differences

Within the **conventional GenAI chatbot** condition, ENA revealed significant differences in interaction patterns between learners with low and high GenAI literacy. A Mann-Whitney test showed that learners with low literacy ($Mdn = -0.16, N = 16$) were statistically significantly different from those with high literacy ($Mdn = 0.16, N = 11, U = 45.00, p = 0.04, r = 0.49$) along the X-axis (MR1; 16.8%) with an alpha of 0.05. However, no substantial difference was found along the Y-axis (SVD2; 22.3%), with the Mann-Whitney test indicating that low GenAI literacy learners ($Mdn = 0.08, N = 16$) were not statistically significantly different from high GenAI literacy learners ($Mdn = 0.01, N = 11; U = 81.00, p = 0.75, r = 0.08$). As shown in Figure 3 (mid), learners with low GenAI literacy (red) exhibited a strong emphasis on seeking or offering clarification, as evidenced by the thicker edge between *User.Clarification* and *Chatbot.Information*, reflecting a need for rephrasing and explaining information. Conversely, learners with high GenAI literacy (blue) demonstrated more complex interaction patterns, actively engaging in information exchange fields, such as *User.Integration* and *Chatbot.Information*. They also displayed intricate networks involving *User.Reflection* and *Chatbot.Integration*, indicating a greater propensity to combine new information with prior knowledge and reflect on their understanding.

Within the **scaffolding GenAI chatbot** condition, ENA also revealed notable differences in interaction patterns between learners with varying levels of GenAI literacy. A Mann-Whitney test indicated that learners with low GenAI literacy ($Mdn = -0.09, N = 19$) were statistically significantly different from those with high GenAI literacy ($Mdn = 0.12, N = 21; U = 308.00, p = 0.00, r = 0.54$) along the X-axis (MR1; 18.1%) with an alpha of 0.05. However,

similar to the conventional chatbot condition, no significant difference was observed along the Y-axis (SVD2; 26.6%). The Mann-Whitney test showed that low GenAI literacy learners ($Mdn = 0.17$, $N = 19$) were not statistically significantly different from high GenAI literacy learners ($Mdn = 0.13$, $N = 21$; $U = 217.00$, $p = 0.64$, $r = -0.09$). From Figure 3 (right), learners with high GenAI literacy (blue) showed the most prominent connection between *Chatbot.Information* and *User.Information*, followed by the connection between *Chatbot.Information* and *User.Integration*. These connections suggested extensive bidirectional information exchanges and the ability of high-literacy learners to integrate information into their cognitive framework effectively. In contrast, low GenAI literacy learners (red) showed more diverse interactions with the chatbot, exemplified by connections between *User.Clarification*, *User.Command*, *User.Reflection*, and *Chatbot.Information*. These learners often sought clarification, prompted simple commands, and engaged in reflective processes to understand the information before articulating it in writing. As a result, their interactions were less structured and more exploratory, highlighting potential challenges in effectively absorbing and integrating information.

5 Discussion and Conclusion

This study investigated the effectiveness of conventional and scaffolding chatbots in enhancing learners' comprehension of complex visualisations in LADs and the role of GenAI literacy in these interactions. For **RQ1**, we found that both conventional and scaffolding chatbots significantly improved learners' understanding of insights from complex visualisations in LADs. This finding aligns with and provides empirical support for the proposed benefits of GenAI chatbots [14, 30, 69]. It also underscores the value of transitioning from exploratory to explanatory analytics [15], particularly in education, where learners and educators may struggle to extract insights from complex visualisations and therefore, need support and guidance during the process [13, 18, 29]. The current findings advocate for using conventional and scaffolding GenAI chatbots to deliver such support based on their evident effectiveness. Additionally, while both types of chatbots resulted in the same extent of comprehension improvement, the scaffolding chatbots might be a better option for learners with lower GenAI literacy, further discussed in response to RQ2 below.

Regarding **RQ2**, we found that improvements in learners' comprehension are related to their GenAI literacy, with higher literacy leading to greater improvements. This finding aligns with previous research that highlights the importance of considering learners' GenAI literacy in educational studies incorporating GenAI technologies, as the degree of benefit may depend on how effectively learners can utilise the support provided by these technologies [4, 10, 75]. Additionally, the stronger positive relationship between GenAI literacy and comprehension improvement with conventional chatbots compared to scaffolding chatbots underscores the necessity of considering GenAI literacy, especially when learners are exposed to GenAI technologies without guidance. This finding resonates with prior studies identifying a low level of prompting skills among students [12, 39]. Conversely, the results suggest that integrating GenAI chatbots with scaffolding questions can potentially reduce the dependence on learners' GenAI literacy, making these chatbots broadly applicable regardless of a learner's prior GenAI

literacy. This aligns with the documented benefits of scaffolding techniques in supporting better learning [21, 32] and contributes to achieving more equitable and inclusive education in the age of GenAI [27, 71].

In terms of **RQ3**, with the conventional GenAI chatbot, we found that high GenAI literacy learners engaged in more complex cognitive processing activities compared to their lower literacy counterparts. This finding supports the assumption that GenAI literacy might affect a learner's efficiency and effectiveness when engaging with GenAI tools, highlighting the need to measure and foster GenAI literacy as an essential skill [68]. For the scaffolding GenAI chatbot condition, high GenAI literacy learners exhibited more structured interactions, whereas low literacy learners displayed more exploratory interactions, indicating they needed additional steps to comprehend and synthesise information into written responses. This aligns with previous studies underscoring the importance of tailored guidance and scaffolding in supporting lower proficiency learners [21]. Additionally, the finding that low GenAI literacy learners could engage in iterative processes, navigating and learning through exploration and reflection, suggests that the scaffolding chatbot can effectively support these learners by providing dynamic, adaptive assistance to explore and understand complex content [32].

The current study has several **implications**. Specifically, it highlights the need for educational technologies to consider learners' varying levels of GenAI literacy when designing and implementing GenAI-powered tools. Future research should explore strategies to improve GenAI literacy among students [10, 39, 68], thus enabling them to gain greater benefits from these technologies. Additionally, the findings suggest a need for the integration of scaffolding techniques within GenAI chatbots to provide more equitable support across diverse learner populations [21, 32]. Practitioners should consider incorporating scaffolding questions and guided interactions to help learners with lower GenAI literacy effectively navigate complex data visualisations. Furthermore, the study's findings suggest a promising direction for enhancing LADs by combining explanatory analytics with GenAI-driven personalised support [19, 72], ultimately contributing to more inclusive and effective educational environments. Continued investigation into the long-term impacts of these technologies on learning outcomes and their potential in various educational contexts will be critical to maximising their utility and ensuring they foster a supportive and equitable learning experience [68].

This study has several **limitations** that should be acknowledged and that pave the way for further research. Firstly, the sample size, while statistically sufficient, was relatively small and may not represent the broader student population, limiting the generalisability of the findings. Future studies should involve larger and more diverse samples to validate these results across different educational contexts. Secondly, the study primarily focused on short-term interactions with GenAI chatbots, and it remains unclear how sustained use over longer periods might affect learners' comprehension and engagement. Longitudinal studies are needed to explore the enduring impacts of GenAI literacy and chatbot integration on learning outcomes. Additionally, while the study underscores the importance of GenAI literacy, it did not address specific instructional strategies for improving this literacy among students. Future research could

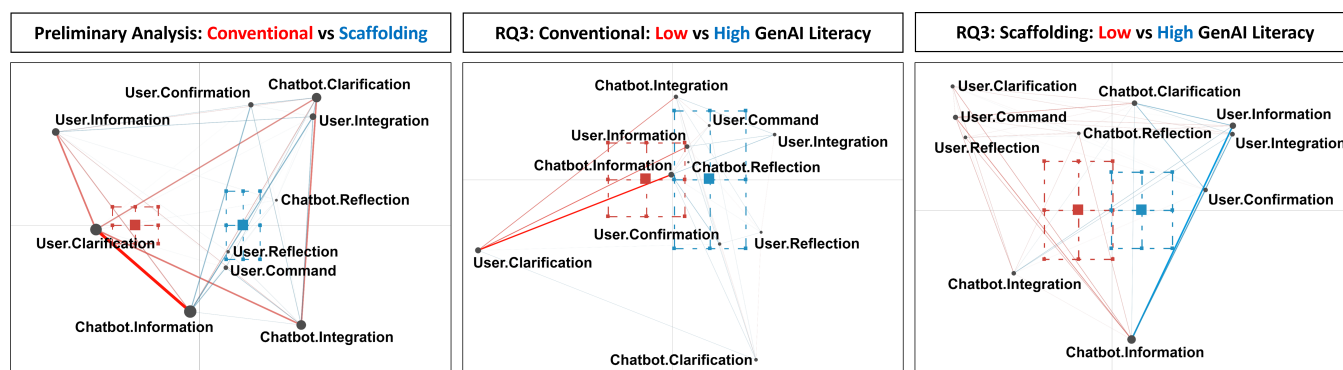


Figure 3: Comparison plots for the preliminary analysis (left), RQ3 conventional (mid), and scaffolding chatbots (right).

develop and evaluate targeted interventions aimed at enhancing GenAI literacy. Finally, the complexities of different disciplines and how GenAI chatbots might uniquely support learning in varied subject areas were not examined. Subsequent studies could investigate the applicability and effectiveness of conventional and scaffolding chatbots across different educational domains, offering a nuanced understanding of their role in diverse learning environments.

Acknowledgments

This research was in part supported by the Australian Research Council (DP220101209, DP240100069). L.Y.'s work is fully funded by the Digital Health CRC (Cooperative Research Centre). D.G.'s work was, in part, supported by the DHCRC and Defense Advanced Research Projects Agency (DARPA) through the Knowledge Management at Speed and Scale (KMASS) program (HR0011-22-2-0047). The DHCRC is established and supported under the Australian Government's Cooperative Research Centres Program. The U.S. Government is authorised to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA or the U.S. Government.

References

- [1] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altmenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* (2023).
- [2] Ahmad Al-Doulat, Nasheen Nur, Alireza Karduni, Aileen Benedict, Erfan Al-Hossami, Mary Lou Maher, Wenwen Dou, Mohsen Dorodchi, and Xi Niu. 2020. Making Sense of Student Success and Risk Through Unsupervised Machine Learning and Interactive Storytelling. In *Artificial Intelligence in Education*. Springer International Publishing, Cham, 3–15.
- [3] Ali Abbas Falah Alzubi. 2024. Generative Artificial Intelligence in the EFL Writing Context: Students' Literacy in Perspective. *Qubahan Academic Journal* 4, 2 (May 2024), 59–69. <https://doi.org/10.48161/qaj.v4n2a506>
- [4] Ravinithesh Annappureddy, Alessandro Fornaroli, and Daniel Gatica-Perez. 2024. Generative AI Literacy: Twelve Defining Competencies. *Digital Government: Research and Practice* (Aug. 2024). <https://doi.org/10.1145/3685680>
- [5] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. 2014. *Standards for Educational and Psychological Testing*. AERA, Washington, DC.
- [6] Richard C Atkinson. 1968. Human memory: A proposed system and its control processes. *The psychology of learning and motivation* 2 (1968).
- [7] Zied Bahroun, Chiraz Anane, Vian Ahmed, and Andrew Zacca. 2023. Transforming education: A comprehensive review of generative artificial intelligence in educational settings through bibliometric and content analysis. *Sustainability* 15, 17 (2023), 12983.
- [8] Benjamin S Bloom, David R Krathwohl, Bertram B Masia, et al. 1984. Bloom taxonomy of educational objectives. In *Allyn and Bacon*. Pearson Education London.
- [9] Dale Bowman, Zachari Swiecki, Zhiqiang Cai, Yeyu Wang, Brendan Eagan, Jeff Linderoth, and David Williamson Shaffer. 2021. The Mathematical Foundations of Epistemic Network Analysis. In *Advances in Quantitative Ethnography*, Andrew R. Ruis and Seung B. Lee (Eds.). Springer International Publishing, Cham, 91–105.
- [10] Aras Bozkurt. 2024. Why Generative AI Literacy, Why Now and Why it Matters in the Educational Landscape? Kings, Queens and GenAI Dragons. *Open Praxis* 16, 3 (2024), 283–290.
- [11] Gary Charness, Uri Gneezy, and Michael A Kuhn. 2012. Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization* 81, 1 (2012), 1–8.
- [12] Thomas K.F. Chiu. 2024. Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence* 6 (June 2024), 100197.
- [13] Linda Corrin. 2018. Evaluating students' interpretation of feedback in interactive dashboards. *Score reporting research and applications* (2018), 145–159.
- [14] Mutlu Cukurova. 2024. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. *BJET* (2024).
- [15] Vanessa Echeverria, Roberto Martinez-Maldonado, Roger Granda, Katherine Chiliza, Cristina Conati, and Simon Buckingham Shum. 2018. Driving data storytelling from learning design. In *Proceedings of the 8th international conference on learning analytics and knowledge*. 131–140.
- [16] Vanessa Echeverria, Roberto Martinez-Maldonado, Simon Buckingham Shum, Katherine Chiliza, Roger Granda, and Cristina Conati. 2018. Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *JLA* 5, 3 (2018), 73–97.
- [17] Yizhou Fan, Yuanru Tan, Mladen Raković, Yeyu Wang, Zhiqiang Cai, David Williamson Shaffer, and Dragan Gašević. 2022. Dissecting Learning Tactics in MOOC using Ordered Network Analysis. *Journal of Computer Assisted Learning* (2022), in press.
- [18] Gloria Milena Fernandez Nieto, Kirsty Kitto, Simon Buckingham Shum, and Roberto Martinez-Maldonado. 2022. Beyond the learning analytics dashboard: Alternative ways to communicate student data insights combining visualisation, narrative and storytelling. In *LAK '22*. 219–229.
- [19] Gloria Milena Fernandez-Nieto, Roberto Martinez-Maldonado, Vanessa Echeverria, Kirsty Kitto, Dragan Gašević, and Simon Buckingham Shum. 2024. Data storytelling editor: A teacher-centred tool for customising learning analytics dashboard narratives. In *LAK '24*. 678–689.
- [20] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. 2023. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997* (2023).
- [21] Pauline Gibbons. 2002. *Scaffolding language, scaffolding learning*. Heinemann Portsmouth, NH.
- [22] Kirk Goldsberry. 2012. Courtvision: New visual and spatial analytics for the nba. In *2012 MIT Sloan sports analytics conference*, Vol. 9. 12–15.
- [23] Lucas Rafael Stefanel Gris, Ricardo Maracini, Arnaldo Candido Junior, Edresson Casanova, Anderson Soares, and Sandra Maria Aluisio. 2023. Evaluating OpenAI's Whisper ASR for Punctuation Prediction and Topic Modeling of life histories of the Museum of the Person. *arXiv preprint arXiv:2305.14580* (2023).
- [24] Sara Hennessy, Mutlu Cukurova, Cathy Levin, Manolis Mavrikis, and Louis Major. 2024. BJET Editorial 2024: A call for research rigour. , 5–9 pages.

- [25] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of hallucination in natural language generation. *Comput. Surveys* 55, 12 (2023), 1–38.
- [26] Yueqiao Jin, Roberto Martinez-Maldonado, Dragan Gašević, and Lixiang Yan. 2024. GLAT: The Generative AI Literacy Assessment Test. *arXiv preprint arXiv:2411.00283* (2024).
- [27] Yueqiao Jin, Lixiang Yan, Vanessa Echeverria, Dragan Gašević, and Roberto Martinez-Maldonado. 2024. Generative AI in Higher Education: A Global Perspective of Institutional Adoption Policies and Guidelines. *arXiv preprint arXiv:2405.11800* (2024).
- [28] Rogers Kaliisa, Kamila Misiejuk, Sonsoles López-Pernas, Mohammad Khalil, and Mohammed Saqr. 2024. Have Learning Analytics Dashboards Lived Up to the Hype? A Systematic Review of Impact on Students' Achievement, Motivation, Participation and Attitude. In *LAK '24 (LAK '24)*. ACM.
- [29] Rogers Kaliisa, Anders Mørch, and Anders Kluge. 2022. 'My Point of Departure for Analytics is Extreme Skepticism': Implications Derived from An Investigation of University Teachers' Learning Analytics Perspectives and Design Practices. *Technology, Knowledge and Learning* 27 (06 2022).
- [30] Hassan Khosravi, Olga Viberg, Vitomir Kovanovic, and Rebecca Ferguson. 2023. Generative AI and Learning Analytics. *JLA* 10, 3 (2023), 1–6.
- [31] Jeonghyun Kim, Il-Hyun Jo, and Yeonjeong Park. 2016. Effects of learning analytics dashboard: analyzing the relations among dashboard utilization, satisfaction, and learning achievement. *Asia Pacific Education Review* 17 (2016), 13–24.
- [32] Nam Ju Kim, Brian R Belland, and Andrew E Walker. 2018. Effectiveness of computer-based scaffolding in the context of problem-based learning for STEM education: Bayesian meta-analysis. *Educational Psychology Review* 30 (2018), 397–429.
- [33] Mohammad Amin Kuhail, Nazik Alturki, Salwa Alramlawi, and Khlood Alhejori. 2022. Interacting with educational chatbots: A systematic review. *Education and Information Technologies* 28, 1 (July 2022), 973–1018. <https://doi.org/10.1007/s10639-022-11177-3>
- [34] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 551–560.
- [35] Unggi Lee, Ariel Han, Jeongjin Lee, Eunseo Lee, Jiwon Kim, Hyeoncheol Kim, and Cheolil Lim. 2023. Prompt Aloud!: Incorporating image-generative AI into STEAM class with learning analytics using prompt data. *Education and Information Technologies* (2023), 1–31.
- [36] Haotian Li, Yun Wang, and Huamin Qu. 2024. Where are we so far? understanding data storytelling tools from the perspective of human-ai collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.
- [37] Lyn Lim, Maria Bannert, Joep van der Graaf, Shaveen Singh, Yizhou Fan, Surya Surendrannair, Mladen Rakovic, Inge Molenaar, Johanna Moore, and Dragan Gašević. 2023. Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior* 139 (2023), 107547.
- [38] Duri Long and Brian Magerko. 2020. What is AI Literacy? Competencies and Design Considerations. In *CHI '20*. ACM.
- [39] Wenhan Lyu, Yimeng Wang, Tingting (Rachel) Chung, Yifan Sun, and Yixuan Zhang. 2024. Evaluating the Effectiveness of LLMs in Introductory Computer Science Education: A Semester-Long Field Study. In *L@S '24*. ACM, 63–74.
- [40] Pingchuan Ma, Rui Ding, Shuai Wang, Shi Han, and Dongmei Zhang. 2023. Demonstration of InsightPilot: An LLM-empowered automated data exploration system. *arXiv preprint arXiv:2304.00477* (2023).
- [41] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez Nieto, and Simon Buckingham Shum. 2020. From data to insights: A layered storytelling approach for multimodal learning analytics. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–15.
- [42] Roberto Martinez-Maldonado, Vanessa Echeverria, Gloria Fernandez-Nieto, Lixiang Yan, Linxuan Zhao, Riordan Alfredo, Xinyu Li, Samantha Dix, Hollie Jaggard, Rosie Wotherspoon, et al. 2023. Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction* 31, 1 (2023), 1–41.
- [43] Marry L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica* (2012), 276–282. <https://doi.org/10.11613/bm.2012.031>
- [44] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).
- [45] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. 2021. Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence* 2 (2021), 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- [46] Xavier Ochoa. 2022. Multimodal Learning Analytics - Rationale, Process, Examples, and Direction. In *The Handbook of Learning Analytics* (2 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, Dragan Gašević, and Agathe Merceron (Eds.). SoLAR, 54–65.
- [47] Chinedu Wilfred Okonkwo and Abejide Ade-Ibijola. 2021. Chatbots applications in education: A systematic review. *Comput. Educ.: Artif. Intell.* 2 (2021), 100033.
- [48] Keng-Boon Ooi, Garry Wei-Han Tan, Mostafa Al-Emran, Mohammed A Al-Sharafi, Alexandru Capatina, Amrita Chakraborty, Yogesh K Dwivedi, Tzu-Ling Huang, Arpan Kumar Kar, Voon-Hsien Lee, et al. 2023. The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems* (2023), 1–32.
- [49] Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*. 1–22.
- [50] Adriano Pinargote, Eddy Calderón, Kevin Cevallos, Gladys Carrillo, Katherine Chiliza, and Vanessa Echeverria. 2024. Automating Data Narratives in Learning Analytics Dashboards using GenAI. In *LAK Workshops*. 150–161.
- [51] Jeevan Pokhrel and Aruna Awasthi. 2021. *Effectiveness of Dashboard and Intervention Design*. Springer International Publishing, 93–116.
- [52] Stanislav Pozdniakov, Roberto Martinez-Maldonado, Yi-Shan Tsai, Vanessa Echeverria, Namrata Srivastava, and Dragan Gasevic. 2023. How Do Teachers Use Dashboards Enhanced with Data Storytelling Elements According to their Data Visualisation Literacy Skills?. In *LAK '23*. 89–99.
- [53] Gomathy Ramaswami, Teo Susnjak, Anuradha Mathrani, and Rahila Umer. 2022. Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. *Technology, Knowledge and Learning* 28, 3 (Aug. 2022), 959–980. <https://doi.org/10.1007/s10758-022-09613-x>
- [54] Wenqi Ren, Yang Tang, Qiyu Sun, Chaoqiang Zhao, and Qing-Long Han. 2023. Visual semantic segmentation based on few/zero-shot learning: An overview. *IEEE/CAA Journal of Automatica Sinica* (2023).
- [55] Muhittin Sahin and Dirk Ifenthaler. 2021. Visualizations and dashboards for learning analytics: A systematic literature review. *Visualizations and dashboards for learning analytics* (2021), 3–22.
- [56] David Williamson Shaffer, Wesley Collier, and Andrew R Ruis. 2016. A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *JLA* 3, 3 (2016), 9–45.
- [57] Hongbo Shao, Roberto Martinez-Maldonado, Vanessa Echeverria, Lixiang Yan, and Dragan Gasevic. 2024. Data Storytelling in Data Visualisation: Does it Enhance the Efficiency and Effectiveness of Information Retrieval and Insights Comprehension?. In *CHI'24*. 1–21.
- [58] Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval augmentation reduces hallucination in conversation. *arXiv preprint arXiv:2104.07567* (2021).
- [59] Shamane Siriwardhana, Rivindu Weerasekera, Elliott Wen, Tharindu Kalu-arachchi, Rajib Rana, and Suranga Nanayakkara. 2023. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering. *Transactions of the Association for Computational Linguistics* 11 (2023), 1–17.
- [60] Teo Susnjak, Gomathy Suganya Ramaswami, and Anuradha Mathrani. 2022. Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education* 19, 1 (Feb. 2022). <https://doi.org/10.1186/s41239-021-00313-7>
- [61] Robert M Thorndike, George K Cunningham, Robert Ladd Thorndike, and Elizabeth P Hagen. 1991. *Measurement and evaluation in psychology and education*. Macmillan Publishing Co, Inc.
- [62] Katrien Verbert, Xavier Ochoa, Robin De Croon, Raphael A Dourado, and Tinne De Laet. 2020. Learning analytics dashboards: the past, the present and the future. In *Proceedings of the 10th Learning Analytics and Knowledge Conference*. 35–40.
- [63] Cai-Ting Wen, Chen-Chung Liu, Ching-Yuan Li, Ming-Hua Chang, Shih-Hsun Fan Chiang, Hung-Ming Lin, Fu-Kwun Hwang, and Gautam Biswas. 2024. The learning analytics of computational scientific modeling with self-explanation for subgoals and demonstration scaffolding. *Computers & Education* 215 (2024), 105043.
- [64] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnasar, Jesse Spencer-Smith, and Douglas C Schmidt. 2023. A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382* (2023).
- [65] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. 2023. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155* (2023).
- [66] GE Xun and Susan M Land. 2004. A conceptual framework for scaffolding III-structured problem-solving processes using question prompts and peer interactions. *Educational technology research and development* 52, 2 (2004), 5–22.
- [67] Lixiang Yan, Vanessa Echeverria, Yueqiao Jin, Gloria Fernandez-Nieto, Linxuan Zhao, Xinyu Li, Riordan Alfredo, Zachari Swiecki, Dragan Gašević, and Roberto Martinez-Maldonado. 2024. Evidence-based multimodal learning analytics for feedback and reflection in collaborative learning. *British Journal of Educational Technology* 55, 5 (2024), 1900–1925.
- [68] Lixiang Yan, Samuel Greiff, Ziwen Teuber, and Dragan Gašević. 2024. Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour* 8, 10 (2024), 1839–1850.

- [69] Lixiang Yan, Roberto Martinez-Maldonado, and Dragan Gasevic. 2024. Generative Artificial Intelligence in Learning Analytics: Contextualising Opportunities and Challenges through the Learning Analytics Cycle. In *LAK '24*. 101–111.
- [70] Lixiang Yan, Roberto Martinez-Maldonado, Linxuan Zhao, Samantha Dix, Hollie Jaggard, Rosie Wotherspoon, Xinyu Li, and Dragan Gašević. 2023. The role of indoor positioning analytics in assessment of simulation-based learning. *BJET* 54, 1 (2023), 267–292.
- [71] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gašević. 2024. Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology* 55, 1 (2024), 90–112.
- [72] Lixiang Yan, Linxuan Zhao, Vanessa Echeverria, Yueqiao Jin, Riordan Alfredo, Xinyu Li, Dragan Gašević, and Roberto Martinez-Maldonado. 2024. VizChat: Enhancing Learning Analytics Dashboards with Contextualised Explanations Using Multimodal Generative AI Chatbots. In *International Conference on Artificial Intelligence in Education*. Springer, 180–193.
- [73] Ahmed Mohamed Fahmy Yousef and Ahmed Ramadan Khatiry. 2021. Cognitive versus behavioral learning analytics dashboards for supporting learner's awareness, reflection, and learning process. *Interactive Learning Environments* 31, 9 (Dec. 2021), 5460–5476.
- [74] Linxuan Zhao, Zachari Swiecki, Dragan Gašević, Lixiang Yan, Samantha Dix, Hollie Jaggard, Rosie Wotherspoon, Abra Osborne, Xinyu Li, Riordan Alfredo, and Roberto Martinez-Maldonado. 2023. METS: Multimodal Learning Analytics of Embodied Teamwork Learning. In *LAK '23*. 186–196.
- [75] Xin Zhao, Andrew Cox, and Liang Cai. 2024. ChatGPT and the digitisation of writing. *Humanit. Soc. Sci. Commun.* 11, 1 (April 2024).