

Article

Evaluating a Metahuman-Integrated Computer-Based Training Tool for Nursing Interventions: Usability and Expert Heuristic Analysis

Aeri Jang ¹ , Hyunju Jeong ² and Yunhee Kim ^{1,*}

¹ Department of Nursing, Mokpo National University, 1666 Yeongsan-ro, Cheonggye-myeon, Muan-gun 58554, Jeollanam-do, Republic of Korea; sseillerjang79@gmail.com

² Department of Nursing, Nambu University, 1 Nambu 1-gil, Gwangsan-gu, Gwangju 62271, Republic of Korea; jhjuddee@nambu.ac.kr

* Correspondence: kimyunhee@mokpo.ac.kr; Tel.: +82-061-450-6291

Abstract

Advances in immersive technologies, such as Metahuman Creator integrated with Unreal Engine, offer new opportunities for interactive and realistic digital learning in nursing education. While computer-based training (CBT) has demonstrated benefits for self-directed learning, limited research has examined the usability and reliability of Metahuman-based digital textbooks (DTs) in clinical nursing education. This study aims to evaluate the usability of a Metahuman-based CBT DT for nursing interventions using a multi-method approach combining user testing and expert heuristic evaluation. A total of 12 undergraduate nursing students and 4 nursing education experts used the program, which included two clinical scenarios (nursing care for ileus and upper gastrointestinal bleeding), and completed the user version of the Mobile App Rating Scale (uMARS). Experts conducted a heuristic evaluation based on eight mobile usability principles. Quantitative data were analyzed using descriptive statistics, and qualitative feedback was evaluated through inductive content analysis. The students rated the overall usability as high (mean uMARS score = 4.25/5), particularly for layout and graphics. Experts provided moderately positive ratings (mean = 3.71/5) but identified critical issues in error prevention, consistency, and user control. Qualitative feedback emphasized the need for automatic data saving, clearer navigation, improved credibility of information sources, and enhanced interactivity.

Keywords: usability evaluation; heuristic evaluation; computer-based training; digital textbook; nursing education



Academic Editor: Pedro Couto

Received: 30 September 2025

Revised: 30 October 2025

Accepted: 30 October 2025

Published: 31 October 2025

Citation: Jang, A.; Jeong, H.; Kim, Y. Evaluating a Metahuman-Integrated Computer-Based Training Tool for Nursing Interventions: Usability and Expert Heuristic Analysis. *Appl. Sci.* **2025**, *15*, 11650. <https://doi.org/10.3390/app152111650>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Metahuman Creator is a cloud-based platform tightly integrated with Unreal Engine that enables the real-time creation of high-resolution digital humans with facial, body, and clothing features [1,2]. This technology supports realistic expression, including detailed nonverbal elements, making it suitable for implementing real-world interactions in education, simulation, and medicine [2,3]. Recently released Unreal Engine 5.6 has fully integrated Metahuman Creator as an in-editor tool, allowing developers to create objects and edit animations directly without a separate web platform [2]. Furthermore, digital avatars combining Metahuman rendering and voice emotion delivery technology showed high cognitive effects in emotion expression and empathy evaluation [4]. These technological and educational effects provide a strong foundation for Metahuman-based

computer-based training (CBT) digital textbooks (DTs) to comprehensively achieve realism, immersion, and educational efficacy.

CBT promotes learners' self-regulated learning and enables repetitive learning regardless of time and location, thereby enhancing the accessibility and efficiency of digital-based education [5]. Although immersive technologies such as AR or VR provide a high sense of presence, their dependence on hardware and the high cost of building environments restrict their adoption in educational settings [6]. In contrast, CBT-based DTs can be accessed on everyday personal devices without spatial constraints, making them particularly suitable for fields such as nursing where repetitive learning is important [7]. CBT content incorporating adaptive learning technology and personalized feedback has been reported to increase learners' engagement and immersion, thereby improving the effectiveness of and satisfaction with learning, depending on digital literacy levels [8]. Therefore, CBT-based DTs serve as a practical alternative to promote qualitative improvement in nursing education by offering educational benefits of accessibility, efficiency, and personalized feedback.

The effectiveness of digital educational tools depends not only on the quality of content but also on learners' ease of access and use, making usability evaluation essential [9]. In particular, in nursing education environments, where complex information structures and procedural thinking are required, intuitive and efficient user interface (UI) design has a significant impact on learning outcomes. Heuristic evaluation, as an expert-based usability evaluation tool, is well suited for the early-stage evaluation of digital educational platforms, as it can rapidly identify multiple usability issues with a limited number of personnel [10]. Nielsen [9] suggested that just as few as five evaluators can identify about 75% of all usability issues, which has led to the widespread use of the heuristic method as a tool that satisfies both cost-effectiveness and speed. Therefore, this study also seeks to systematically derive the structural errors and improvement points of the program through a heuristic review based on expert evaluation.

Research on the effectiveness of DTs and computer-based learning has been steadily accumulating in the fields of educational technology and health education, mainly focusing on improvements in learners' satisfaction, comprehension, and self-directedness [6,11]. Prior studies such as Saab et al. and Choi [12,13] have conducted usability evaluations of VR-based nursing simulations, highlighting both learners' satisfaction and interface challenges. However, these works focused on immersive simulation modalities rather than Metahuman-integrated CBT systems.

Nevertheless, research on standardized DTs remains underexplored. While attempts have recently been made to apply realistic digital technologies such as Metahuman to education, most studies have been limited to medical simulation or metaverse contexts [3,14], and systematic research on actual educational DT development and usability evaluation remains scarce. Specifically, studies that simultaneously verify the reliability, validity, and usability of CBT platforms based on complex clinical situations in nursing are almost nonexistent. This study represents an attempt to empirically evaluate the educational effects of DTs applying Metahuman technology in nursing education, filling the gap in existing research.

Ultimately, this study aims to evaluate the usability of a CBT program that applies Unreal Engine-based Metahuman techniques for the development of DTs for nursing education and identify problems and user needs in the process. For this purpose, nursing students will be asked to use the program, after which the overall usability will be measured using usability evaluation tools. Furthermore, a heuristic evaluation will be conducted with nursing education experts to systematically analyze improvement needs in relation to both system-level and technological aspects. This study seeks to offer an empirical foundation

for the development and dissemination of future DT content by integrating user experience and expert perspectives.

2. Materials and Methods

This study was designed as a multi-method usability evaluation, combining user testing and expert review methods, to assess the usability of a CBT-based DT applying Unreal Engine-based Metahuman techniques for nursing education and identify improvement needs. This combined approach is grounded in prior research results showing that the complementary involvement of users and experts allows for a holistic understanding of both learners' experiences and underlying design flaws in a system [15]. The usability evaluation of this study adopts a multi-method framework informed by Silva et al. [16] and Cho et al. [17], who emphasize combining user testing and heuristic inspection to comprehensively assess digital tools. In line with their approach, we conducted parallel evaluations: student usability testing to measure perceived satisfaction and engagement, and expert heuristic evaluation to uncover structural or interface-level usability issues. This dual-track design strengthens the validity of findings by integrating experiential feedback with expert-based diagnostics. Therefore, this study aims to strengthen both the design quality and learning effectiveness of educational content through a parallel analysis of actual usability data and expert diagnostic results on nursing education DTs. The overall research process is shown in Figure 1.

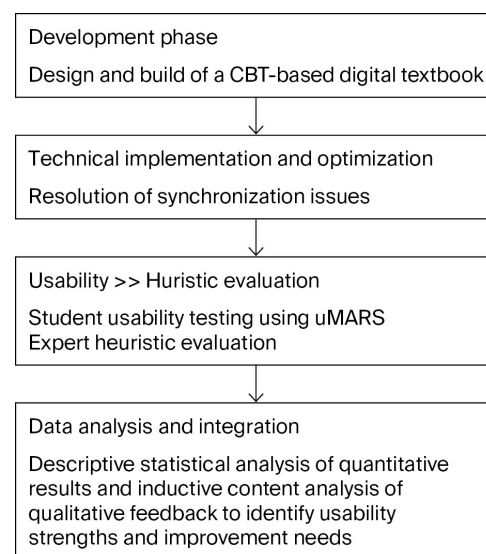


Figure 1. Flow diagram of the study process.

2.1. Development of CBT-Based DT Program Using Unreal Engine and Metahuman Technology

The developed DT was designed to enhance the effectiveness of nursing intervention education in adult nursing and implement realistic virtual patient simulation by applying Metahuman technology based on Unreal Engine 5.6. The use of the DT aims to overcome the limitations of conventional static textbooks and one-way content and improve learners' clinical reasoning abilities and practical problem-solving skills. The developed DT program consists of two scenarios, namely, nursing intervention for intestinal obstruction and nursing intervention for gastrointestinal bleeding. Each scenario is structured as a problem-centered sequence that reflects actual clinical situations, including a start interface (patient information summary, learning objective guidance prior to beginning nursing interventions), learning modules (implementation of a multiple-choice decision-making process following the sequential flow of nursing interventions), patient response animations (emotional, verbal, and behavioral responses expressed through Metahuman

characters), result feedback (clinical outcome feedback and brief explanations based on learners' decisions), and a learning summary. The developed program was built on Unreal Engine 5.0. It utilizes Metahuman Creator to present realistic patient expressions (facial expressions, gestures, eye contact, etc.), incorporates a responsive UI (touch-based internal navigation and interaction), and supports stand-alone execution on laptops and tablets (Windows OS .exe format). During development, several implementation challenges were encountered while integrating Metahuman assets and animations into the Unreal Engine environment. Synchronization issues between facial animation and voice data were resolved through calibration of the Live Link Face system and manual refinement of morph-target parameters to enhance lip-sync precision and emotional realism [1]. Rendering high-fidelity Metahuman models resulted in GPU load and frame-rate instability, which were mitigated using Level-of-Detail (LOD) optimization and baked lighting techniques [3]. These optimization processes improved real-time performance and ensured smooth interaction during user testing. The CBT-based digital textbook was developed as an interactive simulation tool that allows students to observe and participate in clinical scenarios as if they were real nurses. Learners received patient handover information and made nursing decisions throughout each scenario, mirroring the clinical handoff process. Although the program followed a guided sequence, learners experienced it as an active role-play rather than passive observation. When incorrect procedures or nursing interventions were selected, the system provided real-time error messages and corrective feedback to facilitate learning. This approach enabled both experiential engagement and reflective understanding within a structured, feedback-driven environment.

Figure 2 illustrates the sequential learning flow of the developed CBT-based digital textbook integrating Metahuman technology. The program begins with a Patient Handover screen, where learners review essential patient information (demographics, diagnosis, medications, and vital signs) and identify learning objectives. In the Learning Module, learners make clinical decisions through multiple-choice steps following the nursing process. The Patient Response stage presents the Metahuman's verbal and nonverbal reactions—such as facial expressions, eye gaze, and voice tone—reflecting the learner's selected interventions. Based on these actions, the Result Feedback screen provides real-time guidance and corrective information, enabling immediate error recognition and self-reflection. Finally, the Learning Summary presents a concise performance report and key learning points for reflection.

This interactive structure mirrors real-world clinical reasoning, combining cognitive decision-making with affective feedback through a virtual patient.

2.2. Participants

In usability evaluations, the number of participants remains debated among scholars. However, previous studies suggested that approximately 10 evaluators are sufficient to comprehensively identify usability problems [18]. Accordingly, through random sampling, 12 nursing students enrolled in an adult nursing course were recruited, comprising 4 students each from the second, third, and fourth years. For the heuristic evaluation, 4 nursing faculty members were recruited through random sampling, following the recommendation that 3–5 evaluators constitute the most efficient group size [10]; all of these participants had more than 5 years of teaching experience in nursing education and demonstrated an active interest in contemporary digital pedagogy.

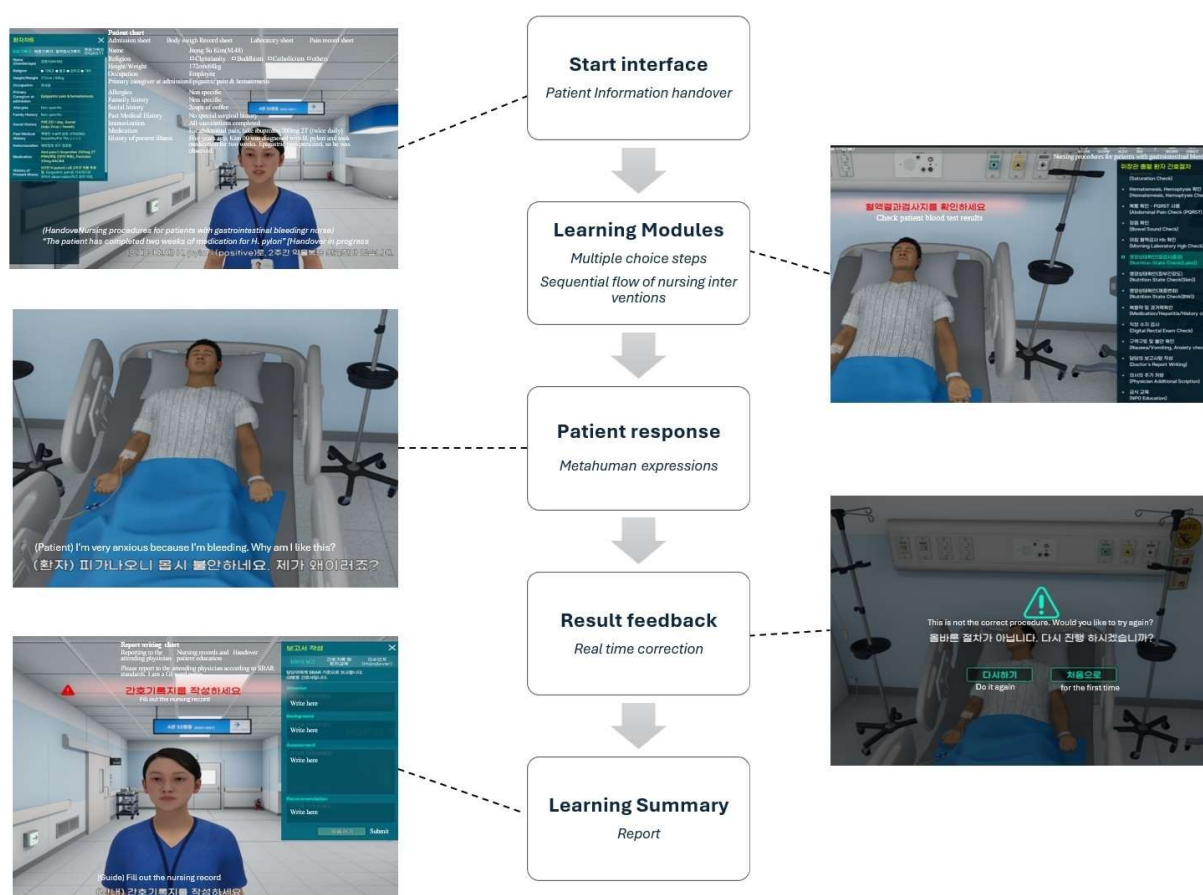


Figure 2. Structure and flow of the Metahuman-based CBT digital textbook program.

A total of 12 students participated in this study, with a mean age of 21.07 (± 1.03) years. The participants consisted of 4 s-year, 4 third-year, and 4 fourth-year students, 7 of whom had prior experience using DTs. For the heuristic evaluation, 4 faculty experts participated, including 3 professors of adult nursing and 1 professor of critical care nursing. The experts' mean clinical experience was 4.91 (± 2.10) years, and their mean educational experience was 12.00 (± 8.03) years (Table 1).

Table 1. General characteristics of participants ($N = 16$).

Characteristics	Categories	n (%) or $M \pm SD$
Students ($n = 12$)		
Age (year)		21.07 $\pm$ 1.03
Using the experience of digital textbooks	Yes	7 (58.3)
	No	5 (41.7)
Learning experience in nursing care for upper gastrointestinal bleeding	Yes	8 (66.7)
	No	4 (33.3)
Learning experience in nursing care for patients with ileus	Yes	5 (41.7)
	No	7 (58.3)
Experts ($n = 4$)		
Age (year)		48.25 $\pm$ 6.29
Area of expertise	Adult nursing	3 (75.0)
	Critical care nursing	1 (25.0)

Table 1. *Cont.*

Characteristics	Categories	<i>n</i> (%) or <i>M</i> ± <i>SD</i>
Research field	Web-based education	1
	Mobile-based education	1
	Virtual reality-based education	1
	Smartphone application-based education	1
Clinical experience (year)		4.91 ± 2.10
Educational experience (year)		12.00 ± 8.03

2.3. Research Tools

In this study, the User Version of the Mobile App Rating Scale (uMARS) was used to evaluate the usability of the digital application [19]. Although the evaluated program is a CBT tool rather than a mobile app, we selected the uMARS because its multidimensional structure—focusing on engagement, functionality, aesthetics, information quality, and subjective experience—is broadly applicable to digital educational tools beyond mobile contexts.

This tool consists of a total of 20 items on a 5-point Likert scale (1 point = strongly disagree; 5 points = strongly agree) and evaluates 5 subareas: engagement, functionality, information, aesthetics, and subjective quality. At the time of development, internal reliability was Cronbach’s $\alpha = 0.90$ for the overall scale and 0.70 or higher for each subscale, and test–retest reliability was reported as ICC = 0.66–0.70.

To further analyze the application’s technical errors and design flaws, a heuristic evaluation tool [10] tailored for mobile environments was additionally employed. This tool evaluates usability defects based on 8 principles: visibility of system status, match between the system and the real society, consistency and standard, user control and freedom, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, and error prevention. Prior studies have reported that it detects more defects than Nielsen’s heuristics while achieving higher inter-rater agreement.

2.4. Data Collection Methods and Ethical Considerations

This study was conducted after receiving approval for ethical procedures from the Institutional Review Board (IRB) of Mokpo National University (#MNUIRB-250706-SB-016-01, approved on 19 June 2025). In consideration of the participants’ ethical rights, the purpose and procedures of the study were explained in advance, and written informed consent was obtained from those who voluntarily agreed before the survey was conducted. The survey was conducted from 7 to 14 July 2025. Participants used a laptop with the CBT-based DT installed and subsequently completed a questionnaire.

2.5. Data Analysis Methods

Usability evaluation was conducted separately for expert and subject groups, and the authors analyzed the collected data using descriptive statistical analysis to calculate the respective means and standard deviations. SPSS version 25.0 [20] was used for data analysis. In this study, 4 experts conducted the evaluation based on Nielsen’s 8 mobile-specific heuristic items. In the evaluation, issues were prioritized and incorporated into application improvements if they were identified by at least 2 experts under the same item or if a single expert assigned a severity rating of 4 or higher [21]. In addition, qualitative evaluation was conducted using responses to open-ended questions included at the end of the questionnaire (“What do you consider to be the strengths and areas for improvement of this application?”). The collected response data were then analyzed using inductive content analysis, and the procedure was as follows: First, two researchers repeatedly

read all response texts to derive initial meaning units (coding units). Second, codes with similar meanings were categorized to form subthemes and themes, and in this process, the reliability of the interpretation was secured through consensus. Third and last, the frequency of each theme and representative quotes were organized and used as supplementary data for the qualitative evaluation [22]. This qualitative analysis contributed to a more in-depth identification of the program's strengths and improvement needs, along with quantitative results.

3. Results

This section presents the outcomes of both expert heuristic evaluations and user usability assessments conducted for the CBT-based digital textbook. The results are organized to provide a comprehensive understanding of usability performance, highlighting both quantitative scores and qualitative insights from participants.

3.1. Heuristic Evaluation

The four selected experts were introduced to the CBT-based nursing DT evaluation and provided with a manual outlining the installation and use of the application developed with Unreal Engine Metahuman technology. The results of the heuristic evaluation of the program are shown in Table 2.

Table 2. Result of the heuristic evaluation for the computer-based training (CBT)-based digital textbook (DT) application ($N = 4$).

Heuristics	Severity Score				Examples of Qualitative Evaluation Narratives
	E1	E2	E3	E4	
Visibility of System Status	0	0	0	0	Location-based learning guide function E4 ^a : "The app did not clearly indicate the current status or progress shown to users."
Match Between the System and the Real Society	0	2	1	0	None
Consistency and Standards	0	4	2	0	Insufficient screen transition or button connectivity: E1: "The unclear Back or Home buttons make navigation within the app difficult." E1: "During learning, there is a lack of a function to control content back and forth or go back."
User Control and Freedom	0	3	0	1	E4: "Providing error hints and corrective feedback would improve learning."
Recognition Rather than Recall	1	0	2	0	E1: "Sound settings were missing, and navigation was difficult."
Flexibility and Efficiency of Use	0	0	3	0	E1: "Lack of poor navigation hindered."
Aesthetic and Minimalist Design	0	0	0	0	None
Error Prevention	2	3	0	3	Lack of automatic Save button and interface status display: E3: "Because automatic saving is not available, input data may be lost easily." E4: "It would be good to provide hints and enable correction when errors occur." E2: "It would be good to provide ways to correct errors."

^a E = Expert.

Five problem areas were identified in the heuristic evaluation, with the most serious problem confirmed in the "Error Prevention" item, which was pointed out by three experts, with the highest severity recorded at 3 points. In the qualitative opinions, one specific criticism was that "because automatic saving is unavailable, input data may be lost easily." The second most problematic item was "Consistency and Standards," raised by two or more experts, with one expert assigning a severity of 4 points. Qualitative responses included the opinion that "the unclear Back or Home buttons make navigation within the app difficult,"

confirming the need for improvement in connectivity between functions and screen layout design. The third most problematic were “User Control and Freedom,” “Recognition rather than recall,” and “Flexibility and Efficiency of Use,” each of which had two experts raise problems. However, all experts scored “Visibility of System Status” and “Aesthetic and Minimalist Design” with 0 points, confirming that no problems occurred in these areas. However, some qualitative responses also noted that “the app did not clearly indicate the current status or progress shown to users.”

3.2. Usability Evaluation

The students ($n = 12$) recorded average total scores of 4.0 or higher in most items. In particular, the highest satisfaction was recorded for the visual components “Layout” ($M = 4.75$, $SD = 0.45$) and “Graphics” ($M = 4.75$, $SD = 0.45$). On the other hand, relatively low scores were recorded for “Quantity of Information” ($M = 3.83$, $SD = 1.47$) and “Credibility of Source” ($M = 3.92$, $SD = 1.31$), and some students commented that “the content was beneficial, but more detailed explanations would be good.” The experts ($n = 4$) gave the highest scores for “Target Group” ($M = 4.50$, $SD = 0.58$) and “Recommendation” ($M = 4.25$, $SD = 0.96$), but showed generally lower evaluation tendencies compared to students. They particularly gave low scores in functional and information reliability aspects such as “Customization” ($M = 3.00$, $SD = 1.41$), “Interactivity” ($M = 3.00$, $SD = 1.63$), and “Credibility of Source” ($M = 3.25$, $SD = 2.22$), noting that “due to insufficient interaction design, it was difficult to feel a sense of actual engagement,” and “information with weak credibility needs improvement.”

For the major area-wise evaluation results, under “Engagement,” students recorded an average of 4 points or higher in most items, positively evaluating interest and interactivity with learning content. On the other hand, experts gave low scores ($M = 3.00$) for “Interactivity,” pointing out the absence of interactive design elements, and presented opinions that “apart from basic interactions, the design does not sufficiently promote learner engagement.” In the area of “Functionality,” students were generally positive, but experts gave low scores in “Performance” ($M = 3.25$, $SD = 0.96$) and “Navigation” ($M = 3.75$, $SD = 0.96$), presenting criticism that “functional consistency is poor, and navigation is limited.” For “Aesthetics,” both students and experts gave high scores overall, with “Visual Appeal” averaging 4 points or higher for both groups. For “Information,” students gave high scores for “Quality of Information” ($M = 4.67$) but evaluated the “Quantity” and “Credibility” items relatively low. Experts pointed out that the “source citation is insufficient, and detailed explanations are inadequate.” In the “Subjective Quality” area, students showed high satisfaction in “Recommendation” ($M = 4.33$) and “Overall Satisfaction” ($M = 4.42$), while experts gave relatively low evaluations in “Usage Intent” ($M = 2.75$) and “Willingness to Pay” ($M = 3.00$), stating that “it is still insufficient for the commercialization stage” (see Appendix A).

4. Discussion

This study evaluated the usability of a Metahuman-based CBT DT for nursing education using a multi-method approach to derive the validity and areas for improvement in user-centered design. Heuristic evaluation results confirmed the necessity for improvement in “Error Prevention,” “Consistency and Standards,” “User Control Freedom,” and “Flexibility and Efficiency of Use” items as follows.

In the “Error Prevention” item, three experts evaluated its severity as 2–3 points, emphasizing the potential loss of user-entered data following the lack of an automatic saving feature and the necessity of functions to correct errors or provide hints for errors during the learning process. This evaluation suggests that the developed DT can cause not

only execution errors (slips) due to user carelessness but also “planning errors arising from inconsistency between system design and user expectations.” Prior research has reported that errors during the learning process increase learners’ frustration and decrease their levels of immersion and engagement [23]. Therefore, the CBT-based DT developed in this study should be improved by incorporating automatic saving and alert functions that warn users when they attempt inappropriate nursing interventions, thereby preventing errors proactively. Although web-based interactive content has been reported to enhance learning efficiency compared with traditional e-textbooks, most automated feedback systems still remain at a limited answer-based comparison level. They do not yet provide structures capable of detecting and warning learners of complex decision-making errors in real time [24].

In contrast, the Metahuman-based CBT digital textbook developed in this study provides real-time nonverbal feedback—such as facial expressions, gaze, and voice responses of the virtual patient—based on the learner’s choices and actions. This design demonstrates a clear distinction from conventional static or simple reactive textbooks by offering an interactive and preventive learning structure that allows learners to recognize and adjust potential errors before they occur.

In the “Consistency and Standards” item, one expert evaluated the severity as 4 points. Participants in the qualitative evaluation noted that the unclear Back or Home buttons on the screen hindered screen navigation. This incurred additional time for learners to learn interface navigation methods, making it difficult to concentrate on the learning content and increasing the cognitive load [24]. Therefore, to enhance the visual clarification of navigation buttons, in addition to the current standard icon-type Home button in the upper right, a Back button should be added and changed to a color that contrasts with the screen, consistently placed at the top of the screen. A previous study [22] reported that when e-textbooks lack consistency in screen navigation rules or feedback structures, learners’ predictability and sense of control decrease, leading to increased cognitive load. Similarly, the Metahuman-based CBT digital textbook developed in this study also exhibits some issues with consistency. Therefore, future improvements should focus on designing a consistent response structure in which nonverbal feedback operates according to standardized rules, enabling learners to better predict system reactions and maintain a stable understanding of the learning context.

In the “User Control Freedom” item, one expert evaluated severity as 3 points. This item evaluates whether the safeguards are available for learners to recover from mistakes [24], emphasizing the freedom to undo. However, the current design forces complete termination when the Exit button is toggled, preventing learners from recovering from minor errors without losing overall progress. Therefore, redo/undo functions that allow partial reversal during task performance are needed to prevent termination of the entire program.

Similarly, a previous study [25] evaluating user responses to a virtual nursing simulation for nursing students reported that, once learners made an error during the learning process, there was no option to return to the previous stage, requiring them to restart the entire sequence. This indicates that existing DTs and VR-based learning systems, which largely depend on linear progression, have structural limitations that hinder learners from recovering from errors or adjusting their learning paths.

In contrast, the Metahuman-based CBT DT developed in this study includes an interactive learning structure that enables learners to recognize and adjust their judgments through nonverbal feedback—such as the virtual patient’s facial expressions, gaze, and voice—when inappropriate nursing interventions are selected. Such a design provides an educational medium with a technological foundation that could evolve into an interactive

error-preventive and recovery structure through the future incorporation of redo/undo functions and error recovery modules.

In the “Flexibility and efficiency of use” item, one expert scored the severity 3 points. Prior studies [10,24] emphasized that this item should incorporate user customization settings tailored to various user needs, as well as accessibility improvements and interface adaptations to enhance learning efficiency and accommodate different devices or situations. Virtual simulation offers advantages in enhancing learners’ immersion and sense of realism. However, its implementation requires platforms such as head-mounted displays, 360-degree videos, computers, and a stable internet connection, which can hinder accessibility and practical use for learners [26]. In contrast, the Metahuman-based CBT digital textbook developed in this study was designed to operate independently on a standard laptop without the need for additional equipment. Therefore, it has the advantages of minimizing platform dependency and enabling flexible use across various learning environments.

However, the CBT-based DT in this study was designed exclusively for laptop environments, which may have limitations regarding usage flexibility and efficiency. Therefore, to improve usability in the future, it is necessary to ensure compatibility with various devices, such as smartphones or tablets, as suggested in a previous study [27].

Usability was scored 4.25 points by students and 3.71 points by experts, both showing good levels (uMARS is 3.0 points or higher) according to the criteria presented by Stoyanov [19]. The discussion of the usability evaluation results is as follows.

The students gave 4.75 and 4.42 points for “Layout, Graphics” and “Overall satisfaction, Recommendation,” respectively, suggesting that visual interest and user-friendly components worked as positive learning experiences for learners familiar with digital environments. Indeed, qualitative responses also confirmed opinions that the “screen composition and graphics are simple and intuitive, so they do not interfere with learning flow,” supporting this implication.

However, experts gave relatively lower scores for items such as “Quantity of Information” and “Credibility of Source,” with 3.83 and 3.92 points, respectively, and there were also qualitative responses stating that “information sources are unclear, making them difficult to trust.” These results suggest that content design must ensure the credibility of learning materials and clearly specify their sources, demonstrating that DTs should address not only interface-level usability but also learning credibility.

Expert usability evaluations were generally lower than those of students, with particularly critical perspectives presented in terms of functionality and information aspects. Experts gave scores of 3.00, 3.00, and 3.25 points for “Customization,” “Interactivity,” and “Credibility of Source,” respectively, pointing out that the program remains in a standardized content structure. Along with this, heuristic evaluation repeatedly identified problems in the “Error Prevention,” “Consistency and Standards,” and “User Control and Freedom” items, meaning that the current system is inadequate in individual path exploration or situational adaptive response design for learners. Interestingly, qualitative opinions revealed that features identified by learners as strengths were perceived as weaknesses in heuristic evaluation. For example, positive responses, such as “simple and easy to view,” were interpreted by experts as “insufficient functions and limited user options.” This interpretation shows that the dual characteristics of user-centered design (simplicity and limitation) can be perceived differently according to evaluation perspectives.

The technical implementation strategies applied in this study directly influenced the system’s usability outcomes. Resolving synchronization issues between Metahuman facial animation and voice data through Live Link Face calibration improved the naturalness of patient expressions, which was reflected in high student ratings for the graphics and layout. Similarly, optimizing performance via Level-of-Detail control and baked lighting enhanced

stability and responsiveness during real-time interactions. These refinements not only strengthened the reliability of user testing but also demonstrated the importance of balancing visual fidelity with computational efficiency in educational simulation design [1,3]. Furthermore, documenting this process increases methodological transparency and provides practical guidance for developers integrating high-fidelity virtual humans into nursing education. Future studies may build upon these findings by applying adaptive rendering techniques and automated synchronization tools to expand cross-device usability and learner immersion.

Based on the above discussion, the implications are as follows. First, considering the educational aspects, visual design and intuitive composition play important roles in the level of immersion of and initial usage satisfaction with DTs. However, having a simple composition alone is insufficient, and it is necessary to ensure learning reliability and depth by verifying the information quality, gathering sources, and providing context. Second, from a developmental perspective, system flexibility (e.g., learners' choice of usage devices), user setting functions, and error prevention design are proposed as core elements for future interface improvement. Third, from a methodological perspective, this study moved beyond dependence on a single evaluation indicator by integrating user experience and expert analysis, thereby evaluating both the authenticity of the content and its technical robustness. This multi-method approach can be considered a standard method for future digital educational content evaluation.

This study has several limitations. The modest sample size constrains generalizability, and evaluations were conducted solely on laptop platforms, limiting insight into cross-device usability. Only two nursing scenarios were assessed, which may not fully represent the diversity of clinical interventions. Additionally, the short-term evaluation design precludes conclusions about knowledge retention or sustained behavioral change.

To overcome these limitations, future work should enroll a larger and more diverse participant cohort and incorporate additional clinical contexts. Furthermore, longitudinal studies are essential to verify long-term learning retention and clinical performance. Content credibility must also be strengthened by embedding verifiable references or links to original sources. An iterative design process—guided by a qualitative coding framework—should be adopted to continuously refine the system based on user feedback. Moreover, enriching interactivity through gamified modules, real-time feedback loops, and branching simulations is recommended to boost learner engagement and motivation. Cross-platform compatibility (e.g., tablets, smartphones) is also critical to enhance accessibility and adoption across varied learning environments.

5. Conclusions

This study evaluated the usability of a Metahuman-based CBT digital textbook (DT) developed with Unreal Engine to support self-directed learning in nursing education. Usability was evaluated with both students and experts, along with a heuristic assessment to identify areas for improvement in user-centered design. Heuristic results revealed the need for improvement in error prevention, consistency and standards, user control and freedom, and flexibility and efficiency of use, indicating potential limitations such as insufficient recovery paths, unclear interface layout, and device-specific restrictions. As the system is in the prototype stage, high-fidelity graphical implementation using Unreal Engine and Metahuman entails potential technical challenges in performance optimization and real-time interaction stability across devices.

Although overall usability scores were satisfactory, suggestions included enhancing the credibility of learning materials through source citation, expanding cross-device compatibility, and adding functions for error correction and recovery. Thus, future development

will focus on integrating auto-save, redo/undo, and multi-device support to improve flexibility and efficiency.

In summary, the effective use of digital textbooks requires a user-centered design that ensures learning reliability, user autonomy, and real-time feedback. Although it was conducted during the development phase, this study provides meaningful insights into the technical and pedagogical directions needed to enhance the interactivity, accessibility, and reliability of Metahuman-based digital textbooks in nursing education.

Author Contributions: Formal analysis, A.J. and Y.K.; data curation, Y.K. and H.J.; writing—original draft preparation, A.J., Y.K., and H.J.; writing—review and editing, A.J., Y.K., and H.J.; funding acquisition, A.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Research Foundation of Korea, grant number 2021R1I1A3046198. However, the Foundation did not and will not have a role in the study design, data collection, analysis, publication decision, or manuscript preparation.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Mokpo National University (protocol code MNUIRB-250409-SB-005-01, approved on 19 June 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.

Acknowledgments: All individuals included in this section have consented to the acknowledgement. The authors would like to express their sincere thanks to the experts and nursing students who participated. Furthermore, the authors express their deepest gratitude to Samwoo Immersion for their hard work during the entire research and development process.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CBT	computer-based training
DT	digital textbook
IRB	Institutional Review Board
UI	user interface
uMARS	user version of the Mobile App Rating Scale

Appendix A

Table A1. Mean and standard deviation of each item.

Category	Item	Students (<i>n</i> = 12) M ± SD		Experts (<i>n</i> = 4) M ± SD			Examples of Qualitative Evaluation Narratives User	Expert
Engagement	Entertainment	4.25 ± 0.45		3.75 ± 0.50		Initially difficult but immersive once familiar, though lacking engaging design	S1 ^a “It felt unfamiliar at first, but I was quite immersed once I got used to it” S2 “It felt like actually performing nursing, so it was fun.”	E1 ^b “Initially difficult, but OK once familiar” E2 “Apart from basic interactions, the design does not sufficiently promote learner engagement”
	Interest	4.50 ± 0.52		4.50 ± 0.58				
	Customization	3.67 ± 0.98	4.15	3.00 ± 1.41	3.75			
	Interactivity	4.08 ± 0.79	± 0.51	3.00 ± 1.63	± 0.44			
	Target Group	4.25 ± 0.45		4.50 ± 0.58				
Functionality	Performance	4.33 ± 0.78		3.25 ± 0.96		Functions are intuitive and easy to use but require more refinement	S8 “The buttons are intuitive, so I could quickly learn how to use them.” S6 “It would be more convenient if there was a function to rotate the direction of the nurse’s viewpoint with the mouse.”	E4 “Almost no personalization features” E3 “Performance is OK, but there’s a slight delay,” “Functional consistency is poor, and navigation is limited”
	Ease of Use	4.00 ± 0.74		3.75 ± 0.96				
	Navigation	4.25 ± 0.75	4.27	3.75 ± 0.96	3.69			
		± 0.43		± 0.63				
	Gestural Design	4.50 ± 0.52		4.00 ± 0.82				
Aesthetics	Layout	4.75 ± 0.45		3.50 ± 1.00		Design is pretty and visually satisfying but simplification is needed	S10 “The design is clean, and the colors are good, so I could concentrate well.” S11 “The design was nice, and it helped me a lot because I could learn the sequence of nursing procedures.”	E2 “The design is satisfactory, but more simplification needed for professionals”
	Graphics	4.75 ± 0.45	4.64	4.25 ± 0.96	3.92			
	Visual Appeal	4.42 ± 0.51	± 0.30	4.00 ± 0.82	± 0.88			
Information	Quality of Information	4.67 ± 0.49		4.00 ± 0.82		Information structure is easy to understand but more detailed explanations needed	S11 “The content was beneficial, but more detailed explanations would be good.” S8 “It would be good if patient information could be viewed simultaneously when writing SBAR.”	E1 “Source citation is insufficient, and detailed explanations are inadequate” E2 “Difficult to feel engagement because of insufficient interaction design,” “Information with weak credibility needs improvement” E4 “Information structure is appropriate, but source citation is insufficient”
	Quantity of Information	3.83 ± 1.47		3.75 ± 0.96				
	Visual Information	4.17 ± 1.40	4.15	3.75 ± 1.89	3.69			
		± 0.98		± 1.20				
	Credibility of Source	3.92 ± 1.31		3.25 ± 2.22				
Subjective Quality	Recommendation	4.33 ± 0.49		4.25 ± 0.96		Positive response that it is worth trying but still insufficient for commercialization	S2 “Overall, it was satisfactory and seems recommendable to others.” S9 “Overall, the design was user-friendly and posed no difficulties for use.”	E4 “It is worth trying but needs improvement” E2 “It is still insufficient for the commercialization stage”
	Usage Intent	3.75 ± 0.62	4.02	2.75 ± 0.50	3.50			
	Willingness to Pay	3.58 ± 0.79	± 0.46	3.00 ± 0.00	± 0.54			
	Overall Satisfaction	4.42 ± 0.51		4.00 ± 0.82				
Total			4.25 ± 0.24		3.71 ± 0.15			

^a S = Student, ^b E = Expert.

References

- Sharma, D.; Sharma, J. The potential of virtual cloud character creation technology-meta human creator: A review. *AIP Conf. Proc.* **2023**, 2782, 020153. [\[CrossRef\]](#)
- Fang, Z.; Cai, L.; Wang, G. MetaHuman creator the starting point of the metaverse. In Proceedings of the 2021 International Symposium on Computer Technology and Information Science (ISCTIS), Guilin, China, 4–6 June 2021; pp. 154–157. [\[CrossRef\]](#)
- Chojnowski, O.; Eberhard, A.; Schiffmann, M.; Müller, A.; Richert, A. Human-like nonverbal behavior with metahumans in real-world interaction studies: An architecture using generative methods and motion capture. In Proceedings of the 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), Melbourne, Australia, 3–6 March 2025. [\[CrossRef\]](#)
- Salehi, P.; Sheshkal, S.A.; Thambawita, V.; Halvorsen, P. From flat to feeling: A feasibility and impact study on dynamic facial emotions in AI-generated avatars. *arXiv* **2025**. [\[CrossRef\]](#)
- Schunk, D.H.; Zimmerman, B.J. *Self-Regulation of Learning and Performance: Issues and Educational Applications*, 3rd ed.; Routledge: New York, NY, USA, 2023; pp. 255–280. ISBN 9780805813357.
- Zhang, D.; Zhou, L.; Briggs, R.O.; Nunamaker, J.F., Jr. Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Inf. Manag.* **2006**, 43, 15–27. [\[CrossRef\]](#)
- Taylor, D.L.; Yeung, M.; Bashet, A.Z. Personalized and adaptive learning. In *Innovative Learning Environments in STEM Higher Education: Opportunities, Challenges, and Looking Forward*; Ryoo, K., Winkelmann, K., Eds.; Springer: Cham, Switzerland, 2021; pp. 17–34. ISBN 9783030589479.
- Yaseen, H.; Mohammad, A.S.; Ashal, N.; Abusaimeh, H.; Ali, A.; Sharabati, A.-A.A. The impact of adaptive learning technologies, personalized feedback, and interactive AI tools on student engagement: The moderating role of digital literacy. *Sustainability* **2025**, 17, 1133. [\[CrossRef\]](#)
- Nielsen, J. *Usability Engineering*; Morgan Kaufman: San Francisco, CA, USA, 1994; ISBN 0125184069.
- Bertini, E.; Gabrielli, S.; Kimani, S.; Catarci, T.; Santucci, G. Appropriating and assessing heuristics for mobile computing. In Proceedings of the AVI '06: Proceedings of the Working Conference on Advanced Visual Interfaces, Venezia, Italy, 23–26 May 2006. [\[CrossRef\]](#)
- Choi, H.; Tak, S.H.; Lee, D. Nursing students' learning flow, self efficacy and satisfaction in virtual clinical simulation and clinical case seminar. *BMC Nurs.* **2023**, 22, 454. [\[CrossRef\]](#) [\[PubMed\]](#)
- Saab, M.M.; McCarthy, M.; O'Mahony, B.; Cooke, E.; Hegarty, J.; Murphy, D.; Walshe, N.; Noonan, B. Virtual reality simulation in nursing and midwifery education: A usability study. *CIN Comput. Inform. Nurs.* **2023**, 41, 815–824. [\[CrossRef\]](#) [\[PubMed\]](#)
- Choi, K.S. Virtual reality simulation for learning wound dressing: Acceptance and usability. *Clin. Simul. Nurs.* **2022**, 68, 49–57. [\[CrossRef\]](#)
- Pour, M.E.; Gießer, C.; Schmitt, J.; Brück, R. Evaluation of metahumans as digital tutors in augmented reality in medical training. *Curr. Dir. Biomed. Eng.* **2023**, 9, 41–44. [\[CrossRef\]](#)
- Ivory, M.Y.; Hearst, M.A. The state of the art in automating usability evaluation of user interfaces. *ACM Comput. Surv.* **2001**, 33, 470–516. [\[CrossRef\]](#)
- Silva, A.G.; Caravau, H.; Martins, A.; Almeida, A.M.P.; Silva, T.; Ribeiro, Ó.; Santinha, G.; Rocha, N.P. Procedures of user-centered usability assessment for digital solutions: Scoping review of reviews reporting on digital solutions relevant for older adults. *JMIR Hum. Factors* **2021**, 8, e22774. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cho, H.; Yen, P.Y.; Dowding, D.; Merrill, J.A.; Schnall, R. A multi-level usability evaluation of mobile health applications: A case study. *J. Biomed. Inform.* **2018**, 86, 79–89. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hwang, W.; Salvendy, G. Number of people required for usability evaluation: The 10 ± 2 rule. *Commun. ACM* **2010**, 53, 130–133. [\[CrossRef\]](#)
- Stoyanov, S.R.; Hides, L.; Kavanagh, D.J.; Zelenko, O.; Tjondronegoro, D.; Mani, M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR mHealth uHealth* **2015**, 3, e27. [\[CrossRef\]](#) [\[PubMed\]](#)
- IBM Corp. *IBM SPSS Statistics for Windows*, Version 25.0; IBM Corp: Armonk, NY, USA, 2017.
- Jeffries, R.; Miller, J.R.; Wharton, C.; Uyeda, K.M. User interface evaluation in the real world: A comparison of four techniques. In Proceedings of the CHI '91: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New Orleans, LA, USA, 27 April–2 May 1991; pp. 119–124. [\[CrossRef\]](#)
- Elo, S.; Kyngäs, H. The qualitative content analysis process. *J. Adv. Nurs.* **2008**, 62, 107–115. [\[CrossRef\]](#) [\[PubMed\]](#)
- Novak, E.; McDaniel, K.; Li, J. Factors that impact student frustration in digital learning environments. *Comput. Educ. Open* **2023**, 5, 100153. [\[CrossRef\]](#)
- Nielsen, J. 10 Usability Heuristics for User Interface Design. Available online: <https://www.nngroup.com/articles/ten-usability-heuristics/> (accessed on 7 August 2025).
- Liu, Z.; Zhang, Q.; Liu, W. Perceptions and needs for a community nursing virtual simulation system: A qualitative study. *Heliyon* **2024**, 10, e28473. [\[CrossRef\]](#) [\[PubMed\]](#)

26. Kiegaldie, D.; Shaw, L. Virtual reality simulation for nursing education: Effectiveness and feasibility. *BMC Nurs.* **2023**, *22*, 488. [[CrossRef](#)] [[PubMed](#)]
27. Lee, J.; Soyulu, M.Y.; Ou, C. Exploring insights from online students: Enhancing the design and development of intelligent textbooks for the future of online education. *Int. J. Innov. Online Edu.* **2023**, *7*, 29–54. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.