An AI-Supported Simulation-Based Serious Game for Digital Health Management Education: Design and Formative Evaluation

Annika Hering[1a], Gülsen Otman[1a], Annika Viedenz[1] and Jan-David Liebe[1,2]

[1]Center for Medical, Social and Health Informatics (MESH), Osnabrück University of Applied Sciences, Osnabrück, Germany

[2]UMIT - Private University of Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria

[a]Authors contributed equally to this paper

a.hering@hs-osnabrueck.de
g.otman@hs-osnabrueck.de
annika.viedenz@hs-osnabrueck.de
j.liebe@hs-osnabrueck.de

Abstract: Background Teaching digital transformation in healthcare management requires students to address complex organizational challenges, multiple stakeholder perspectives, and decision-making under uncertainty. Traditional case-based teaching provides limited experience of real-world decision-making complexity and feedback. Objectives This study presents the design and formative evaluation of the first version of an artificial intelligence (AI)-supported digital serious game (DSG) in Digital Health Management (DHM) education. It examined students' evaluations of the integrated management methods, generative AI (GenAI)-supported stakeholder interactions, educational design, and AI-generated model solutions. Methods Design of the DSG was informed by the co.LAB generic framework and resulted in a scenario-based learning environment integrating management methods, collaborative problem-solving, GenAI-supported stakeholder interactions, and AI-generated model solutions. The formative evaluation used an embedded mixed-methods design across bachelor's and master's degree courses at two higher education institutions. Twenty-one students participated in the DSG, of whom 15 completed the interim evaluation, while the number of method-specific evaluations varied across learning activities. Qualitative reflections were collected at group level. Results The browser-based DSG combined healthcare digitalization case scenarios, sequential management methods, GenAI-supported stakeholder interactions, collaborative problem-solving, and AI-generated model solutions. The management methods received consistently positive ratings (M = 4.19–4.55/5). Students also positively evaluated the GenAI-supported stakeholder interactions (M = 5.95/7), their usability (M = 4.18/5), the teaching-learning experience (M = 4.98/6), and the AI-generated model solutions (M = 3.98/5). Internal consistency ranged from acceptable to excellent (α = .72 to .91), while Spearman-Brown coefficients for two-item subscales ranged from .66 to .92. The qualitative findings suggest that the AI-generated model solutions supported validation, extension, and critical comparison of students' own analyses. The Net Promoter Score (NPS) was 40. Discussion The findings suggest that GenAI can support scenario-based learning in DHM when GenAI-supported stakeholder interactions and feedback are embedded in a coherent, collaborative learning design. Future studies should use a prospectively defined evaluation framework to compare different forms of AI-generated feedback and assess reflection, objective learning outcomes, competency development, and transfer in larger samples. Conclusion The study provides an initial design and empirical basis for the further development of AI-supported DSGs in DHM education. The findings indicate that the AI-supported components may support perspective taking, comparison, reflection, and method-based decision-making within a coherent learning design.

Keywords: Digital Health Management, Digital Transformation, Serious Games, Education, Artificial Intelligence

1. Background

Higher education is increasingly challenged to prepare students for professional contexts characterized by digital transformation, uncertainty and complex decision-making. In response, artificial intelligence (AI), simulation-based learning and digital serious games (DSGs) are gaining importance as didactic approaches that can support more adaptive, experiential and practice-oriented forms of learning.

AI has been applied in education in various ways, including automated assessment and feedback, personalized learning plans, adaptive instructional support and intelligent tutoring systems (Chen *et al.*, 2020). In particular, AI-driven adaptive learning systems can dynamically adapt learning content, feedback, and learning pathways to individual learners' needs and preferences, thereby supporting personalized learning experiences (Strielkowski *et al.*, 2025). Generative AI (GenAI) extends these possibilities by generating context-sensitive content and feedback in response to specific learning situations. However, its educational value depends on both technical capabilities and pedagogical integration into the learning environment. AI-generated feedback can therefore be embedded in hybrid feedback processes in which it is reviewed, supplemented, and contextualized through human judgment (Banihashem *et al.*, 2026). This is particularly important in competence-oriented higher education, where feedback should not only indicate deficiencies but also stimulate reflection, reasoning, and transfer to complex professional situations.

This need for pedagogical integration is also central to the design of serious games. Previous research has emphasized that the educational effectiveness of serious games depends on integration into a broader learning design that combines gameplay with preparation, facilitation, collaboration, and reflection (Jaccard *et al.*, 2022). DSGs and other experiential learning approaches can connect academic learning with professional practice. In operations and supply chain management education, experiential learning activities, including DSGs and university challenges, were associated with a closer alignment between students' competencies and professional role profiles. Students reported gains in soft skills such as leadership, teamwork and adaptability (Fantozzi and Schiraldi, 2026). Moreover, game-based learning environments can enable learners to experience and reflect on the consequences of their own actions. In serious games, this can support the development of professional, methodological, social, and personal competencies and contextual understanding (Schwägele, 2015).

Such approaches appear particularly relevant in management contexts. Managers face conflicting objectives, limited resources, and difficult-to-predict long-term consequences. Traditional teaching provides limited opportunities to experience the complexity of real-world decision-making. Simulation-based learning is particularly relevant for management education because DSGs expose learners to such situations and support communication, assertiveness, and networked strategic thinking. Such games should be grounded in educational requirements and domain-specific management challenges rather than the game format alone. Relevant design elements include a realistic organizational scenario, differentiated management roles, interdependent decision situations, and a structured didactic framework including preparation, facilitation, theoretical input and discussion or presentation phases (Häberle *et al.*, 2006). In project management education, participation in a simulation-supported course was associated with higher scores in an external certification exam, particularly in competence areas related to teamwork, change management and project design (Opas and Ćwil, 2026).

In Digital Health Management (DHM), these requirements become even more pronounced. Managers require a broad set of capabilities and competencies that enable data-driven decision-making, digital health literacy, planning and managing digital technologies, and leading organizational transformation processes (Brommeyer and Liang, 2022). At the same time, digital health innovations are embedded in complex healthcare systems and are shaped by technology, organizational readiness, integration into existing structures, management and maintenance responsibilities, financing, and the wider system context (Hellstrand Tang *et al.*, 2024). Students therefore need learning environments that allow them to apply management methods in realistic situations, rather than only learning about such methods in abstract or static formats. The integration of AI may further enhance these learning environments by supporting dynamic stakeholder interactions and providing context-sensitive feedback through AI-generated model solutions. Against this background, this study presents the design and formative evaluation of an AI-supported DSG for DHM education. The evaluation was guided by the following research questions (RQs):

> RQ1: How do students evaluate the instructional implementation, applicability, and perceived contribution of management methods?
>
> RQ2: How do students evaluate the GenAI-supported stakeholder interactions, their usability, and the educational design of the DSG?
>
> RQ3: How do students evaluate the AI-generated model solutions as a form of feedback and learning support?

2. Methods

2.1 Design and Development of the Digital Serious Game

The first version of the DSG was developed to provide a practice-oriented learning environment for DHM education. The design process was informed by the co.LAB framework (Jaccard *et al.*, 2021). Typical healthcare digitalization management challenges were translated into scenario-based cases, and management methods were sequenced from current-state analysis through target-concept development to implementation planning. The case scenarios, including the organizational context and stakeholder constellations, were generated with the support of ChatGPT (OpenAI, 2026) and subsequently reviewed and revised by the research team. The resulting learning design was implemented as a browser-based application integrating case-specific content, role-based GenAI-supported stakeholder interactions, task instructions, progress-dependent access, and AI-generated model solutions.

2.2 Formative Evaluation Design and Participants

The formative evaluation used an embedded mixed-methods design. Group-based qualitative reflections complemented quantitative evaluations by providing additional insight into how students compared and interpreted the AI-generated model solutions. Reporting of the quantitative survey component followed the Checklist for Reporting of Survey Studies (CROSS) (Sharma *et al.*, 2021).

Data collection was embedded in master's and bachelor's degree courses in the field of DHM at Osnabrück University of Applied Sciences (HSO) and Medical School Hamburg (MSH) between April and June 2026. The study population comprised students enrolled in these courses. All these students were eligible to participate, and no exclusion criteria were applied. Participants therefore constituted a convenience sample of the participating course cohorts. Quantitative data was collected at the individual level. In contrast, qualitative reflections on the AI-generated model solutions were collected at the group level because the cases had been collaboratively developed and discussed within student teams. Depending on the method, approximately 120–180 minutes were allocated to each methodological activity, followed by 20–30 minutes for subsequent discussion and reflection.

In total, 21 students participated in the DSG, including 13 at HSO and eight at MSH. As participation in the evaluation was voluntary and integrated into ongoing coursework, not all participants completed every measurement point. Responses were collected anonymously, therefore reasons for non-participation could not be systematically recorded. Of the 21 students, 15 completed the interim evaluation, corresponding to a response rate of 71.4%.

2.3 Instruments and Data Collection

To address RQ1, a self-developed method evaluation instrument comprising twelve items was used. Items were rated on an agreement scale, with higher values indicating stronger agreement.

RQ2 was examined through an interim evaluation conducted after the process analysis method. The evaluation used modified items from established instruments: nine items from the Persona Perception Scale (PPS) to assess GenAI-supported stakeholder interactions (Salminen *et al.*, 2020), seven items from the Chatbot Usability Scale (BUS-11) to evaluate usability aspects (Borsci *et al.*, 2023), and 13 items (ten multi-item scale items and three additional single items) from the questionnaire developed by the Centre for Management Simulations (CMS) to assess the educational design of the DSG (Alf and Trautwein, 2023). Two researchers independently translated the PPS and consolidated their versions. Instrument items were selected based on their relevance and adapted to the educational context. Reliability for the instruments was examined post hoc (Eisinga *et al.*, 2013).

Regarding RQ3, a self-developed nine-item scale and group-based free-text reflections assessed the AI-generated model solutions. Groups compared each model solution with their own artifact and documented a joint assessment.

The interim evaluation assessed willingness to recommend the DSG using the 11-point Net Promoter Score (NPS) item (Reichheld, 2003). Student characteristics were also collected, including prior experience with serious games, professional background, experience in developing or implementing digitalization strategies or projects in healthcare, and initial interest in the DSG.

Surveys were administered via LimeSurvey (LS) through a DSG link available only to enrolled students. The LS questionnaire was reviewed by two students to assess comprehensibility and technical functionality prior to data collection, but the instruments were not subjected to formal or psychometric pretesting. Data was collected after predefined learning activities, so assessments referred directly to experienced methods. Table 1 provides an overview of the instruments used.

Table 1: Overview of the instruments

No.	Instrument	Timepoint of Data Collection	Construct Definition	Items (n)	Response Scale
1	Method evaluation	After each management method	Perceived instructional support, clarity and structure of the method implementation, understanding of its purpose, independent applicability, and contribution to analyzing and planning the digitalization project	12	5-point scale (Strongly disagree – Strongly agree)

No.	Instrument	Timepoint of Data Collection	Construct Definition	Items (n)	Response Scale
2	Modified PPS	Interim evaluation	Perceived realism, consistency and usefulness of the GenAI-supported stakeholder interactions	9	7-point scale (Completely disagree – Completely agree)
3	Modified BUS-11	Interim evaluation	Usability of the GenAI-supported stakeholder interactions	7	5-point scale (Strongly disagree – Strongly agree)
4	Modified CMS Teaching-Learning Experience	Interim evaluation	Perceived structure, comprehensibility, realism, feedback opportunities, learning support and motivation	10	6-point scale (Strongly disagree – Strongly agree)
4a	CMS Pace	Interim evaluation	Perceived pace of the course	1	3-point scale (a bit too fast – just right – a bit too slow)
4b	CMS Time	Interim evaluation	Perceived time available per method	1	3-point scale (a bit too short – just right – a bit too long)
4c	CMS Scope	Interim evaluation	Perceived scope of joint discussion phases	1	3-point scale (a bit too short – just right – a bit too long)
5	AI-generated model solutions	After each management method	Validation of students' own work, extension and refinement of students' analyses, and critical reflection and contextual adaptation	3	Free text
		Interim evaluation	Perceived quality, completeness, comprehensibility and reflective value of the AI-generated model solutions	9	5-point scale (Strongly disagree – Strongly agree)
6	NPS	Interim evaluation	Willingness to recommend the DSG	1	11-point scale (0–10)

Note: PPS = Persona Perception Scale; BUS-11 = Chatbot Usability Scale; CMS = Centre for Management Simulations; DSG = Digital serious game.

2.4 Data Analysis

Quantitative data was analyzed using R (v. 4.4.1). Descriptive analyses included mean scores (M) and standard deviations (SD) with scale scores calculated as the mean of the respective items. Responses were included in the dataset if all items of the respective measure had been completed. Due to variation in participation across evaluation points, the number of valid responses was reported separately for each analysis. As students could evaluate multiple management methods, the method-specific data included repeated observations. Analyses were therefore descriptive and method-specific. Cronbach's α or Spearman-Brown was calculated for the PPS, BUS-11, CMS Teaching–Learning Experience, and for the AI-generated model solutions instrument. Reliability was not assessed within the management method evaluation ratings because the items represented distinct evaluative criteria rather than interchangeable indicators of a single latent construct. Nevertheless, an equally weighted mean score was calculated as a descriptive summary of students' overall evaluation across these criteria. The NPS was calculated as the proportion of promoters (9–10) minus detractors (0–6) and ratings of 7–8 were classified as passive.

Free-text responses regarding the AI-generated model solutions were analyzed using qualitative content analysis following Kuckartz and Rädiker (2024). Initial main categories were derived deductively from the reflection tasks, while additional subcategories were developed inductively from the data. A meaningful quote within a group-based reflection response constituted the coding unit. The material was coded by two researchers, and coding was iteratively refined during the analysis. Ambiguous quotes were discussed until consensus was reached.

3. Results

3.1 Learning Design and Case Studies of the Digital Serious Game

The DSG combined scenario-based challenges, role-based stakeholder interactions, rule-based sequential task progression, collaborative decision-making, method-based problem-solving, and structured feedback and reflection. The first version implemented simulation-based learning through qualitative, scenario-based decision-making and GenAI-supported stakeholder interactions rather than numerical simulation of system

consequences. The case studies addressed the introduction of (1) a patient portal, (2) a digital medication management system, (3) a telemonitoring program for chronic heart failure, (4) a mobile documentation and route-planning system in long-term care, (5) an AI-supported risk-identification system in child and youth welfare, and (6) a digital platform for cross-sectoral case coordination. Although fictionalized for educational purposes, they were based on typical digitalization projects in healthcare organizations.

At the beginning of the intervention (1), the lecturer introduced the learning objectives, case studies, and task sequence. Each group worked on one case throughout the intervention. The learning path followed three guiding questions: “Where do we stand?”, “Where do we want to go?”, and “How do we get there?”. These questions structured the gradual development of a method-based implementation strategy for the digital transformation project in healthcare. After familiarizing themselves with the selected case study and the assignment logic, students worked through the management methods. Figure 1 provides an overview of the learning design and instructional sequence of the DSG.

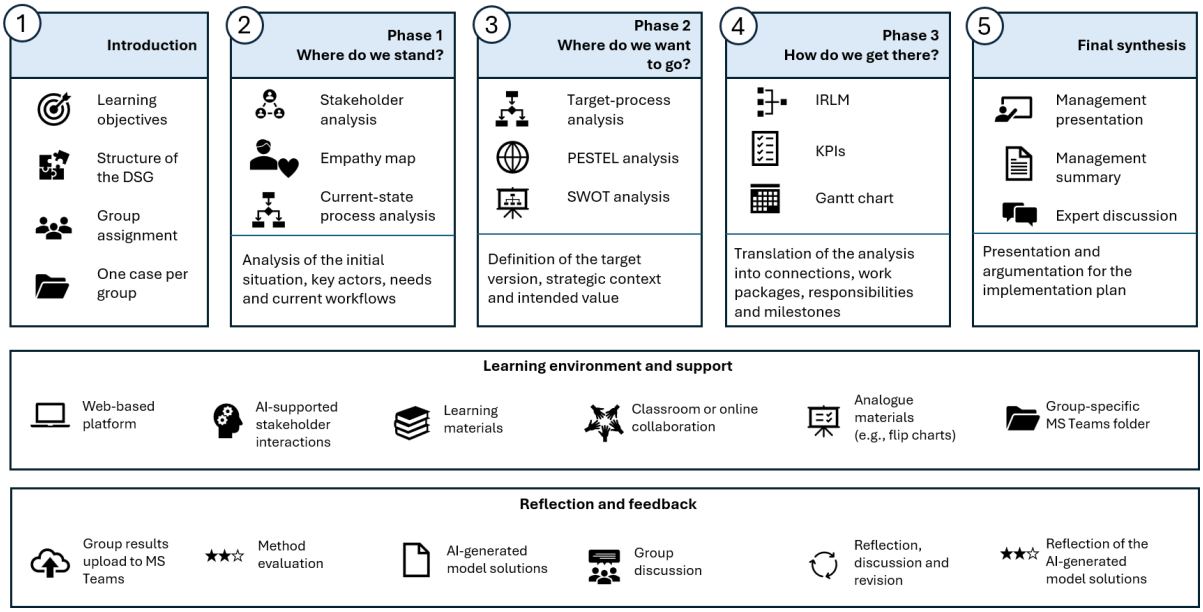


Figure 1: Learning design and instructional sequence of the GenAI-supported digital serious game (DSG)

In the first phase (2), students analyzed the initial situation, underlying problem, and contextual conditions of their case. They applied stakeholder analysis and empathy mapping and conducted a current-state process analysis to examine existing analogue or inefficient workflows.

In the second phase (3), students developed a target concept for the digitalization project. This included modeling a target-state process, analyzing external influencing factors through Political, Economic, Social, Technological, Environmental, and Legal (PESTEL) analysis, and synthesizing internal strengths and weaknesses as well as external opportunities and threats through Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. This phase supported students in defining the intended value of the project for the healthcare organization and its stakeholders.

In the third phase (4), students translated their previous analyses into an implementation logic and a project management plan. They developed an Implementation Research Logic Model (IRLM), including key performance indicators (KPIs), to specify relevant determinants, implementation strategies, and intended outcomes. Based on these results, students created a Gantt chart in which they derived work packages, activities, responsibilities, and milestones for their case.

For the final synthesis (5), students integrated the results of the applied methods and their Gantt chart into a management presentation and prepared a summary for senior management. During their presentation they responded to questions from the lecturer about their implementation plan and justified the approach they had taken. The DSG formed the digital component of a blended learning arrangement. Students used it to access case information, GenAI-supported stakeholder interactions, learning materials, task-specific instructions, and interactive checklists. The management methods were mainly applied collaboratively in the physical classroom using analogue materials such as flip charts and sticky notes. The resulting artifacts were documented and stored

in group-specific Microsoft Teams (MS Teams) folders. After applying each management method, students discussed their results in class and evaluated the respective methods.



Figure 2: "Empathy map" as exemplary project artifact of the digital serious game (DSG)

3.2 Technical Implementation

The DSG was implemented as a browser-based web application using client-server architecture. This supported flexible integration into higher education teaching and location-independent browser access. A central component of the platform was the GenAI-supported stakeholder simulation. Students could interview virtual stakeholders through a conversational interface using natural-language questions. Stakeholder responses were generated via the Anthropic Messages API using Claude 3.5 Sonnet (Anthropic, 2026). Stakeholder-specific profiles of responsibilities, interests, objectives, and concerns informed the prompting process. Students could save stakeholder statements as favorites, add notes, and revisit conversations for subsequent project work.

Progress was tracked within the DSG, and subsequent sections were unlocked only after all tasks within the current method had been completed. This mechanism guided students systematically through the sequence of management methods. Figure 3 illustrates the interface for the use case of an AI-supported risk-identification system in child and youth welfare.

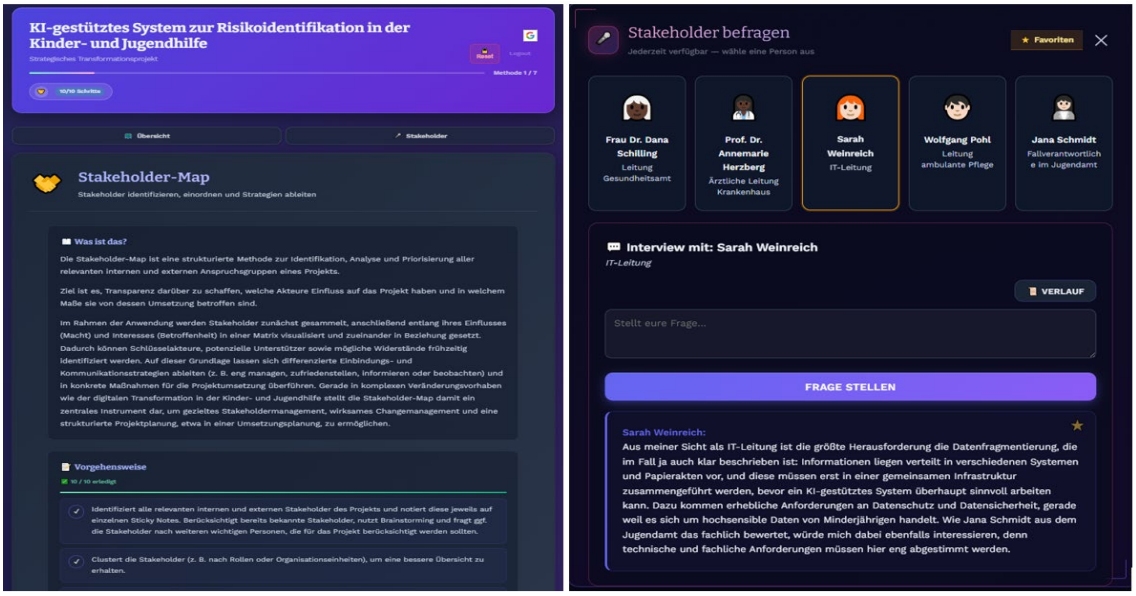


Figure 3: Exemplary "stakeholder analysis" interface with access to GenAI-supported stakeholders in the digital serious game (DSG)

For reflection and self-assessment, AI-generated model solutions were produced for each management method within the platform using Claude (Anthropic, 2026). The model solutions drew on each case and stakeholder interviews and followed the method's predefined steps and guiding questions. Although method-structured, the results were primarily textual rather than visual method-specific artifacts. They were not presented as single

correct answers but as possible approaches against which students could critically compare their own results and reflect on alternative implementation strategies. After completing a method and uploading their results to MS Teams, the student groups received the corresponding model solution from the lecturer.

3.3 Evaluation of Management Methods

Students evaluated all management methods positively, with mean ratings above 4.0 on the five-point agreement scale. SWOT analysis received the highest mean rating (M = 4.55), followed by the empathy map (M = 4.34). PESTEL analysis received the lowest mean rating, although its evaluation was still positive overall (M = 4.19). The results are presented in Table 2. Detailed item-level results are available on the project website (Hochschule Osnabrück, 2026).

Table 2: Descriptive analysis of management methods; n indicates responses summarized across HSO and MSH

Management Method	Responses (n)	M (SD)
Stakeholder analysis	20	4.26 (0.84)
Empathy map	18	4.34 (0.86)
Process analysis (current-state and target-state analysis)	33	4.24 (0.79)
PESTEL analysis	16	4.19 (0.79)
SWOT analysis	14	4.55 (0.74)
IRLM and KPIs	3	4.22 (0.64)

Note: M = Mean score; SD = Standard deviation. Rating on a five-point agreement scale. For process analysis, evaluations of current-state and target-state analyses were pooled.

3.4 Interim Evaluation

3.4.1 Student Characteristics

As shown in Table 3, the interim evaluation included seven HSO and eight MSH students. Most participants had no prior experience with serious games (11/15). Professional backgrounds were heterogeneous across both institutions, with medicine (3/15), healthcare administration, and social services (2/15 each) representing the most common fields. Experience with healthcare digitalization projects was more common among MSH students (7/8) than HSO students (1/7). The initial interest in the DSG was generally positive (12/15).

Table 3: Student characteristics of the interim evaluation; n = 15

Student Characteristics	HSO (n)	MSH (n)	Total (N)
Prior Participation in Serious Games			
Not at all	7	4	11
1-2 times	-	2	2
3-4 times	-	2	2
More than 4 times	-	-	-
Professional Background			
Medicine (e.g. doctors)	3	-	3
Care (e.g., healthcare and nursing, care for the elderly)	-	1	1
Therapeutic professions	-	1	1
Health Management / Health Economics	-	1	1
Administration in the healthcare sector	-	2	2
Computer Science / IT	-	1	1
Social services (e.g., social work, case management)	2	-	2
Business Administration / Management	-	1	1

Student Characteristics	HSO (n)	MSH (n)	Total (N)
Other	2	1	3
Experience with Digitalization Projects in Healthcare			
Yes	1	7	8
No	6	1	7
Initial Interest in the DSG			
Very low	-	-	-
Low	-	-	-
Rather low	3	-	3
Rather high	3	2	5
High	1	4	5
Very high	-	2	2

Note: IT = Information technology; DSG = Digital serious game.

3.4.2 Instrument Results

Students evaluated the GenAI-supported stakeholder interactions, their usability, the CMS teaching-learning experience, and the AI-generated model solutions positively. The findings are presented in Table 4. Detailed item-level results are available on the project website (Hochschule Osnabrück, 2026).

Table 4: Descriptive analysis and reliability of the quantitative interim evaluation; individual rating per student summarized across HSO and MSH (n = 15 for each instrument)

Instrument	Dimension	M (SD)	Reliability
PPS	-	5.95 (1.01)	.89
BUS-11	-	4.18 (0.70)	.72
CMS teaching-learning experience	-	4.98 (1.03)	-
	Comprehensibility	5.17 (0.75)	.66*
	Realism	5.27 (0.94)	.87*
	Teaching design	5.11 (0.80)	.80
	Learning outcomes	4.53 (1.29)	.91
CMS Pace	-	2.00 (0.38)	-
CMS Time	-	1.67 (0.49)	-
CMS Scope	-	1.80 (0.41)	-
AI-generated model solutions	-	3.98 (0.66)	-
	Completeness	4.10 (0.55)	.82*
	Traceability	3.90 (0.55)	.80*
	Justification	3.84 (0.67)	.74
	Overall evaluation	4.13 (0.82)	.92*

Note: M = Mean score; SD = Standard deviation. Reliability for the CMS teaching-learning experience and AI-generated model solutions is reported at the subscale level; reliability was not calculated for CMS Pace, CMS Time, or CMS Scope because each construct was assessed using a single item. Unmarked coefficients represent Cronbach's α, coefficients marked with an asterisk represent Spearman-Brown coefficients.

The dimension "Realism" was the highest-rated CMS subscale (M = 5.27), while "Completeness" (M = 4.10) and "Overall evaluation" (M = 4.13) received the highest ratings among the AI-generated model solutions. Course pace was rated as broadly appropriate (M = 2.00), whereas the time available per method tended to be perceived as somewhat short (M = 1.67). The scope of the joint discussion phases was rated as largely appropriate, with a slight tendency toward being too short (M = 1.80).

The NPS stood at 40. Overall, 53.3% (n = 8) of the students were classified as “promoters”, 33.3% (n = 5) as “passives”, and 13.3% (n = 2) as “detractors”.

Following George and Mallery (2003), internal consistency for the instruments ranged from acceptable to excellent across the respective scales and subscales (α = .72–.91). For the two-item subscales of the CMS teaching-learning experience, Spearman-Brown coefficients ranged from .66 to .87, with Comprehensibility showing the lowest coefficient. For the two-item subscales of the AI-generated model solutions instrument, Spearman-Brown coefficients ranged from .80 to .92.

3.5 Reflection of AI-Generated Model Solutions

As shown in Table 5, the qualitative analysis was based on the three overarching reflection tasks of the AI-generated model solutions: validation of students’ own work, extension and refinement of students’ analyses, and critical reflection and contextual adaptation. The qualitative analysis comprised 343 coded meaning units. Confirmation was the most frequently coded function, followed by extension and concretization. Critical comparison, alternative perspectives, and the identification of omissions and blind spots occurred less frequently. All qualitative results are available on the project website (Hochschule Osnabrück, 2026).

Table 5: Reflection of AI-generated model solutions; group reflections are summarized over the management methods across HSO and MSH

Feedback Area	Feedback Function	Coded Quotes (n)	Illustrative Quotes
Validation of students’ own work	Confirmation	119	*“All details have been confirmed.”* (HSO, IRLM) *“Our list of stakeholders matches the model solution almost 100 %. The clusters we have selected (management vs operational users vs end users) also closely resemble those in the model solution.”* (MSH, Stakeholder analysis)
Extension and refinement of students’ analyses	Extension and concretization of analysis	108	*“The characters’ thoughts were more detailed.”* (HSO, Empathy map) *“The model solution is much more detailed.”* (MSH, PESTEL analysis)
	Identification of omissions and blind spots	27	*“We have not provided a detailed answer to Step 4.”* (HSO, Empathy map) *“We have not identified any key findings.”* (MSH, Process analysis)
Critical reflection and contextual adaptation	Alternative perspective	59	*"Further key stakeholders identified: patients, IT/software providers and payers (holistic approach)"* (MSH, Stakeholder analysis) *"We have chosen a different stakeholder […] For us, this was the more appropriate choice in order to gain two different perspectives."* (HSO, Empathy map)
	Critical comparison	30	*“In general, we find our solution more suitable because it is somewhat more detailed; […].”* (HSO, PESTEL analysis) *“Our visual presentation makes our solution much clearer.”* (MSH, SWOT analysis)

Note: Coded quotes (n) refer to the number of coded meaning units; a single free-text response could contribute to more than one coded quote.

4. Discussion

This study presents the design and formative evaluation of an AI-supported DSG for DHM education under authentic teaching conditions. By combining GenAI-supported stakeholder interactions and AI-generated model solutions with established management methods, the study examines a configuration for which empirical evidence remains limited.

With respect to RQ1, students positively evaluated the instructional implementation, applicability, and perceived contribution of the management methods (M = 4.19–4.55/5). The item-level evaluations further indicate that instructions and methodological steps were generally perceived as clear and comprehensible and that students largely felt able to apply the methods independently. However, ratings were somewhat lower for translating individual method results into subsequent project-planning activities, particularly the derivation of actions for the Gantt chart, and for the perceived applicability of some methods in professional practice. This suggests that while the methods were generally perceived as well implemented within the DSG, their connection to subsequent implementation planning could be made more explicit.

Regarding RQ2, students positively evaluated the GenAI-supported stakeholder interactions, usability, and educational design. Realism was the highest-rated CMS dimension, suggesting that the learning environment was perceived as practice-oriented and plausible, whereas perceived learning outcomes were comparatively lower. Course pace was considered appropriate, but time per method and, to a lesser extent, discussion phases tended to be perceived as too short. Additional time for method application and reflection may therefore strengthen the learning design. The findings indicate that GenAI-supported stakeholder interactions can be integrated into a coherent DHM learning design and were positively perceived by students. The positive NPS further indicates overall acceptance of the first DSG version.

In relation to RQ3, students evaluated the AI-generated model solutions positively as a form of feedback and learning support (M = 3.98/5). Within this evaluation, completeness and the overall assessment received somewhat higher ratings, whereas traceability and justification were rated slightly lower. Students used the model solutions to validate and extend their own work, while they also identified alternative perspectives, omissions, and opportunities for critical comparison. Rather than being treated as authoritative answers, the model solutions therefore served as reference points for comparison and reflection. This feedback function aligns with the role of reflection and debriefing in serious games (Crookall, 2010) and suggests that AI-generated model solutions may support reflective processes. At the same time, confirmation and extension occurred more frequently than critical comparison, indicating that more critical forms of engagement were less prominent in the reflections.

Students' critical engagement with the model solutions further suggests that the pedagogical design of AI-generated feedback, including how directive, staged, and open it is, may be as important as its availability. This interpretation is consistent with recent research showing that the design of AI-generated feedback can influence students' revision behaviour (Alsaiari *et al.*, 2026). Future iterations could move beyond primarily textual model solutions by allowing students to upload their artifacts and receive interactive, artifact-specific feedback. Instead of presenting a complete reference solution, the AI could identify gaps, ask critical follow-up questions, and prompt students to justify or rethink their results.

Several limitations should be considered. The study was conducted in two university courses with a small convenience sample, and voluntary participation resulted in incomplete measurement points. The findings are therefore exploratory and not representative of DHM students more broadly. Participants differed in professional background, prior digitalization experience, and exposure to serious games, particularly between institutions, which may have influenced evaluations. Response numbers also varied across management methods, with very few evaluations for IRLM and KPIs. The adapted and newly developed instruments were not formally validated or psychometrically pretested, and the prior review by two students addressed only comprehensibility and technical functionality. Reliability estimates should therefore be interpreted cautiously. The study assessed perceptions rather than objective learning outcomes or competency development. Because the AI-supported components were embedded in the overall learning design, their specific effects cannot be isolated. Their quality also depended on the stakeholder profiles, prompts, and case information used. Finally, the model solutions were not reviewed before release; students' reflections therefore indicate engagement with the generated content but do not constitute formal validation of its factual accuracy.

Future research could examine whether and how AI-supported components in serious games affect learning outcomes and competency development using larger samples, and whether more explicit prompts for questioning and contextualizing AI-generated feedback strengthen critical reflection. DSGs should incorporate documented practice-based cases to increase contextual depth and organizational realism. A prospective evaluation framework should define validated instruments, outcome measures, measurement points, and procedures for assessing the contribution of individual DSG components.

5. Conclusion

The design of the AI-supported DSG demonstrates how scenario-based healthcare digitalization cases, sequential management methods, GenAI-supported stakeholder interactions, collaborative problem-solving, and AI-generated model solutions can be integrated into a coherent learning environment for DHM education. The educational value of the AI components appears to lie not in providing authoritative answers but in supporting comparison, reflection, and method-based decision-making. The study provides an empirically informed basis for the further development of AI-supported DSGs in DHM education.

Acknowledgements

The authors would like to thank the students who participated in the study and contributed their perspectives during the DSG.

Funding

This study is part of the research project "SMiLeH – Strategic Management in Learning Health Systems" (Project No. 3004-0241), funded by the Stiftung Innovation in der Hochschullehre (StIL). The funder had no role in the study design, data collection, analysis, interpretation of data, or in the writing of the manuscript.

Ethics Declaration

An ethics waiver was issued by the Ethics Committee of Osnabrück University of Applied Sciences (Approval No. W_HSOS/2026/1/3), 21 April 2026, Prof. Dr. Frank Ollermann. All students were informed about the purpose and voluntary participation in the surveys and provided informed consent prior to participation.

AI Declaration

The authors used DeepL (DeepL, 2026) to support language refinement and ChatGPT (OpenAI, 2026) for structural revision during the development of the manuscript. The AI-generated suggestions were used to improve clarity, coherence, and wording. All suggestions were critically reviewed, edited, and approved by the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Data Availability Statement

The complete survey items, quantitative results, and qualitative findings are available on the project website (Hochschule Osnabrück, 2026) or from the authors upon request.

Author Contributions

AH: Conceptualization, Methodology, Writing – Original Draft. GO: Conceptualization, Validation, Writing – Original Draft. AV: Formal analysis, Data Curation, Writing – Review & Editing. JDL: Conceptualization, Supervision.

References

Alf, T. and Trautwein, F. (2023), "Planspielbasierte Lehrveranstaltungen: Ein Vergleich zwischen Präsenz- und Onlinelehre", *Zeitschrift Für Hochschulentwicklung*, Vol. 18 No. Sonderheft Planspiele, pp. 345–363, doi: 10.21240/zfhe/SH-PS/19.

Alsaiari, O., Baghaei, N., Lodge, J.M., Noroozi, O., Gašević, D., Boden, M. and Khosravi, H. (2026), "Directive, metacognitive, or a blend of both? A comparison of AI-generated feedback types on student engagement, confidence, and outcomes", *Computers and Education: Artificial Intelligence*, Vol. 10, p. 100553, doi: 10.1016/j.caeai.2026.100553.

Anthropic. (2026), "Claude 3.5 Sonnet. version: claude-3-5-sonnet-20241022".

Banihashem, S.K., Noroozi, O., Khosravi, H., Schunn, C.D. and Drachsler, H. (2026), "Pedagogical framework for hybrid intelligent feedback", *Innovations in Education and Teaching International*, Vol. 63 No. 2, pp. 554–570, doi: 10.1080/14703297.2025.2499174.

Borsci, S., Schmettow, M., Malizia, A., Chamberlain, A. and Van Der Velde, F. (2023), "A confirmatory factorial analysis of the Chatbot Usability Scale: a multilanguage validation", *Personal and Ubiquitous Computing*, Vol. 27 No. 2, pp. 317–330, doi: 10.1007/s00779-022-01690-0.

Brommeyer, M. and Liang, Z. (2022), "A Systematic Approach in Developing Management Workforce Readiness for Digital Health Transformation in Healthcare", *International Journal of Environmental Research and Public Health*, Vol. 19 No. 21, p. 13843, doi: 10.3390/ijerph192113843.

Chen, L., Chen, P. and Lin, Z. (2020), "Artificial Intelligence in Education: A Review", *IEEE Access*, Vol. 8, pp. 75264–75278, doi: 10.1109/ACCESS.2020.2988510.

Crookall, D. (2010), "Serious Games, Debriefing, and Simulation/Gaming as a Discipline", *Simulation & Gaming*, Vol. 41 No. 6, pp. 898–920, doi: 10.1177/1046878110390784.

DeepL. (2026), "DeepL", 15 July, available at: https://www.deepl.com/de/write.

Eisinga, R., Grotenhuis, M.T. and Pelzer, B. (2013), "The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown?", *International Journal of Public Health*, Vol. 58 No. 4, pp. 637–642, doi: 10.1007/s00038-012-0416-3.

Fantozzi, I.C. and Schiraldi, M.M. (2026), "Bridging education and industry: university challenges and serious games for professional skill development", *The TQM Journal*, Vol. 38 No. 3, pp. 595–610, doi: 10.1108/TQM-12-2024-0511.

George, D. and Mallery, P. (2003), *SPSS® for Windows® Step by Step: A Simple Guide and Reference*, 4th ed., Vols. 1-11.0 Update, Allyn & Bacon, Boston.

Häberle, O., Baume, M. and Krcmar, H. (2006), "Realitätsnahes Entscheidungstraining für zukünftige IT-Manager: Ein computerunterstütztes Planspiel: Aufbau und Lessons Learned".

Hellstrand Tang, U., Smith, F., Karilampi, U.L. and Gremyr, A. (2024), "Exploring the Role of Complexity in Health Care Technology Bottom-Up Innovations: Multiple-Case Study Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability Complexity Assessment Tool", *JMIR Human Factors*, Vol. 11, p. e50889, doi: 10.2196/50889.

Hochschule Osnabrück. (2026), "SMiLeH - Strategic Management in Learning Health Systems", 4 September, available at: https://www.hs-osnabrueck.de/smileh/ergebnisse-und-publikationen/.

Jaccard, D., Bonnier, K.E. and Hellström, M. (2022), "How might serious games trigger a transformation in project management education ? Lessons learned from 10 Years of experimentations", *Project Leadership and Society*, Vol. 3, p. 100047, doi: 10.1016/j.plas.2022.100047.

Jaccard, D., Suppan, L., Sanchez, E., Huguenin, A. and Laurent, M. (2021), "The co.LAB Generic Framework for Collaborative Design of Serious Games: Development Study", *JMIR Serious Games*, Vol. 9 No. 3, p. e28674, doi: 10.2196/28674.

Kuckartz, U. and Rädiker, S. (2024), *Qualitative Inhaltsanalyse. Methoden, Praxis, Umsetzung mit Software und künstlicher Intelligenz*, 6. Auflage., Juventa Verlag, Weinheim.

Opas, M. and Ćwil, M. (2026), "Simulation-based Learning and Project Management Certification Outcomes: Evidence from IPMA Level D Training", *Electronic Journal of E-Learning*, Vol. 24 No. 3, pp. 79–91, doi: 10.34190/ejel.24.3.4740.

OpenAI. (2026), "ChatGPT", GPT 5.5, available at: https://chatgpt.com/.

Reichheld, F.F. (2003), "The one number you need to grow", *Harvard Business Review*, Vol. 81 No. 12, pp. 46–54, 124.

Salminen, J., Santos, J.M., Kwak, H., An, J., Jung, S. and Jansen, B.J. (2020), "Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas", *International Journal of Human-Computer Studies*, Vol. 141, p. 102437, doi: 10.1016/j.ijhcs.2020.102437.

Schwägele, S. (2015), *Planspiel – Lernen – Lerntransfer Eine Subjektorientierte Analyse von Einflussfaktoren*, Monografie, Otto-Friedrich-Universität Bamberg, Bamberg.

Sharma, A., Minh Duc, N.T., Luu Lam Thang, T., Nam, N.H., Ng, S.J., Abbas, K.S., Huy, N.T., *et al.* (2021), "A Consensus-Based Checklist for Reporting of Survey Studies (CROSS)", *Journal of General Internal Medicine*, Vol. 36 No. 10, pp. 3179–3187, doi: 10.1007/s11606-021-06737-1.

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G. and Vasileva, T. (2025), "AI -driven adaptive learning for sustainable educational transformation", *Sustainable Development*, Vol. 33 No. 2, pp. 1921–1947, doi: 10.1002/sd.3221.