Evolving a Digital Game for Architectural Patterns Education

Maria Clara Ribeiro de Menezes
Fluminense Federal University (UFF)
Niterói, Brazil
mcrmenezes@id.uff.br

Lara da Silva Dias
Federal University of Juiz de Fora (UFJF)
Juiz de Fora, Brazil
lara.dias@estudante.ufjf.br

Pedro Henrique Valle
University of São Paulo (IME-USP)
São Paulo, Brazil
pedrohenriquevalle@usp.br

Alessandreia Marta de Oliveira
Federal University of Juiz de Fora (UFJF)
Juiz de Fora, Brazil
alessandreia.oliveira@ufjf.br

Abstract

Software Architecture Education (SAE) is challenging due to the abstract nature of architectural reasoning and the difficulty of applying theoretical knowledge in real-world scenarios. To address these limitations, this work presents PatternQuest Game , a Digital Educational Game (DEG) designed to support the teaching and learning of Architectural Patterns through narrative-driven challenges that promote diagnostic reasoning and decision-making under constraints. The game was implemented as a scalable Web-based game and incorporates gameplay mechanics aligned with a symptom-oriented educational design, in which players diagnose architectural issues and select appropriate patterns under constraints. To evaluate its user experience as an initial validation step, we conducted a heuristic evaluation based on established usability principles, identifying strengths and areas for improvement in the interface and interaction design. The results highlight strengths in usability and interaction design, suggesting the approach's potential to support engagement and contextual reasoning in SAE.

Keywords

Educational Game, Heuristic Evaluation, Architectural Patterns

1 Introduction

Software Architecture (SA) plays an important role in the development of complex systems [3]. Foundational knowledge in these areas is consistently ranked among the most frequently applied skills in daily software development tasks by industry practitioners [1]. In particular, architectural patterns serve as proven, reusable solutions to recurring problems, establishing a shared lexicon that facilitates team communication and system maintainability [12, 13]. Recognizing this relevance, societies such as ACM/IEEE [17] and the Brazilian Computer Society (SBC) [27] address software design and architectural patterns in computer science-related curricula as a fundamental component of Software Engineering Education.

Despite this, teaching and learning SA remain challenges [11]. From the students' perspective, grasping architectural concepts is difficult due to the subject's abstract nature and the lack of real-world context in classroom settings. Traditional lecture-based approaches emphasize structural diagrams and pattern definitions. However, they struggle to foster the diagnostic reasoning required to identify design degradation and apply the pattern under constraints [13]. Consequently, a gap persists between academic education and the practical demands of the software industry.

In professional practice, identifying design degradation and architectural problems is not a trivial task. When architectural documentation is unavailable, developers must rely on source-code symptoms, such as code smells and multi-smell patterns, to spot degradation and determine which refactoring operations or design patterns to apply [22]. This analysis is fundamental to Software Engineering (SE), yet students are rarely exposed to environments where they can safely practice this "symptom-to-pattern" cognitive process. While some active learning approaches, such as cooperative methods and educational games, have been explored in SE education, there is a scarcity of games that simulate the pressure and trade-offs of architectural decision-making.

Therefore, this work presents the evolution and evaluation of PatternQuest Game , a Digital Educational Game (DEG) designed to provide a practical environment for exercising architectural pattern selection. This version of the game was developed based on findings from a preliminary evaluation with students, focusing on addressing identified gaps in interaction and engagement. In this paper, we describe the design decisions behind the game's evolution and report an expert-based heuristic evaluation. This study provides evidence regarding the perceived educational quality, design suitability, and usability of the proposed approach. To guide this investigation, the Research Questions (RQs) were defined: **RQ1:** What usability, gameplay, and interaction strengths and limitations can be identified in the new version of PatternQuest Game through heuristic evaluation? and **RQ2:** To what extent does the evolved game design satisfy pedagogical and educational heuristics for supporting the teaching and learning of Architectural Patterns?

The **main contributions of this work** are: (i) the evolution of an existing DEG for Architectural Patterns education, incorporating improvements derived from student feedback obtained in a preliminary evaluation; (ii) provide a detailed discussion of the design and gameplay refinements introduced in the new version of the game; (iii) report a heuristic evaluation conducted with professors, providing evidence regarding the usability, gameplay quality, and pedagogical adequacy of the proposed approach for supporting Software Architecture Education (SAE).

The remainder of this work is organized as follows. Section 2 discusses related work. Section 3 presents the design and evolution of PatternQuest Game . Section 4 details the heuristic evaluation procedure, while Section 5 discusses and analyzes the obtained results. Section 6 presents the threats to validity. Finally, Section 7 concludes the work and outlines directions for future work.

2 Related Work

SAE is a challenging domain due to the abstract nature of architectural concepts and the need to reason about design decisions, trade-offs, and quality attributes [3]. Traditional lecture-based approaches often emphasize theoretical content and diagram interpretation, but students frequently struggle to connect these concepts with software design practice [11]. In addition, limited experience with large-scale systems makes it difficult for learners to understand how architectural decisions affect system evolution and maintainability.

Previous studies have explored strategies to address these challenges, including adaptive learning environments, digital and non-digital games, and gamification. Hidalgo et al. [14] investigated the learning styles of software architects using Kolb's Learning Style Inventory and found a predominance of analytical and problem-solving-oriented profiles, suggesting that SAE benefits from activities emphasizing practical reasoning and contextual decision-making over purely theoretical exposition.

One line of research explores adaptive educational tools. DECORA [5] recommends architectural alternatives based on requirements and constraints, enabling students to compare trade-offs while learning architectural patterns and design decisions. Its evaluation used motivational instruments such as the Attention, Relevance, Confidence, and Satisfaction (ARCS) model and the Instructional Materials Motivation Survey (IMMS) [16] to analyze students' perceptions regarding motivation and understanding.

Another group of studies investigates game-based approaches for teaching architectural reasoning and trade-offs. DecidArch [7] uses collaborative card-based activities in which players negotiate architectural decisions while considering stakeholder concerns and system events. Similarly, ATAM-RPG [19] adopts role-playing mechanics to simulate Architecture Trade-off Analysis Method (ATAM) sessions, allowing students to discuss quality attributes and prioritize architectural scenarios collaboratively.

More recently, Architectural Stories [9] proposed a narrative-driven non-digital card game to support the learning of Architectural Patterns. Inspired by investigative deduction games, players analyze contextualized software problems by asking questions and inferring the architectural pattern that best addresses the observed symptoms. Unlike previous card-based approaches focused on trade-offs or design decisions, the game emphasizes diagnostic reasoning through narrative interpretation. The proposal was evaluated with 34 undergraduate students using the MEEGA+ model, reporting positive results regarding usability, engagement, motivation, and perceived learning.

The transition from physical to DEGs has also been explored in SAE. D-LEARN [4], adapted from the LEARN board game [24], presents SA concepts through quiz-oriented interactions in a digital environment. The authors evaluated the game using the Model for the Evaluation of Educational Games (MEEGA+) [23] to investigate students' perceptions regarding player experience and usability.

Beyond educational games, some studies investigate the mechanisms of gamification integrated into SA courses. The gamification approach proposed by Araújo et al. [2] used badges, missions, and point systems to encourage participation in undergraduate SAE activities. Similarly, the study by Carneiro et al. [10] analyzed how different motivational profiles, identified through the HEXAD scale [18], influence students' engagement in gamified SA courses.

Table 1 summarizes the main characteristics of the studies discussed in this section. Overall, previous work has explored collaboration, trade-off analysis, adaptive recommendations, quizzes, and gamification mechanisms as strategies to support SAE. However, most approaches rely on structured question-and-answer activities, predefined decision flows, or instructor-guided interactions.

In contrast, only a few studies investigate narrative-driven scenarios to encourage students to interpret contextual information, identify architectural symptoms, and associate software problems with architectural patterns during gameplay. Among them, Architectural Stories explores this instructional strategy in a non-digital format, whereas digital educational games for SAE still predominantly rely on quizzes or structured interactions. Moreover, many studies evaluate motivation, usability, or player experience through questionnaires and surveys, with less emphasis on inspecting interaction and pedagogical issues through heuristic analysis.

In this context, PatternQuest Game differs from previous approaches in two main aspects. First, the DEG organizes gameplay around narrative-driven investigative missions in which players analyze contextualized software problems and select architectural patterns according to inferred system constraints and quality concerns. Second, the evaluation combines evidence from a previous student-based experience study using MEEGA+ with a heuristic evaluation conducted with specialists using a Heuristic Evaluation for Digital Educational Games (AHJED) [6], allowing the identification of both perceived player experience issues and expert-observed interaction and pedagogical limitations.

As shown in Table 1, most studies rely on surveys and perception-based instruments to evaluate motivational or usability aspects of SAE approaches. In contrast, this work combines a previous student experience evaluation using MEEGA+ with a heuristic inspection based on AHJED, allowing the analysis of interaction, gameplay, and pedagogical aspects from both student and specialist perspectives.

3 Description of PatternQuest Game

PatternQuest Game [1] is an educational game designed to support the teaching of architectural patterns through diagnostic problem-solving. It combines pedagogical principles, narrative elements, gamification mechanisms, and a modular SA to create an interactive environment where students analyze software quality issues and select appropriate architectural patterns.

The educational design is grounded in Problem-Based Learning (PBL) [15, 25] and Cognitive Load Theory [26]. To bridge the gap between abstract architectural concepts and practice, the game adopts a narrative-driven metaphor that contextualizes technical challenges within an investigative journey.

PatternQuest Game is set in ArchPattern City, where students assume the role of an "Architectural Investigator." In this fictional setting, software failures are represented as incidents to be investigated, and architectural deficiencies are treated as clues that must be analyzed. Rather than answering isolated technical questions, students are asked to interpret software symptoms, identify underlying structural problems, and select architectural patterns that

[1] Available at: https://patternquestgame.web.app/

Table 1: Comparative analysis of related work in SAE.

Work	Topic	Edu.	Type	Strategy	Evaluation	Digital
DECORA [5]	AP	UG	Adaptive Tool	Adaptive Recommendations	Survey	•
DecidArch [7]	TO	UG	Non-digital Game	Card-based Role-play	Case Study	○
ATAM-RPG [19]	AT	UG	Non-digital Game	RPG-based Collaboration	Case Study	○
Architectural Stories [9]	AP	UG	Non-digital Game	Narrative-driven Investigation	MEEGA+	○
D-LEARN [4]	SA	UG	DEG	Quiz-based Interaction	MEEGA+	•
Araújo et al. [2]	SA	UG	Gamified Course	Badges and Missions	Survey	•
Carneiro et al. [10]	SA	UG	Empirical Study	Structured Q&A	Survey	•
PatternQuest Game	**AP**	**UG**	**DEG**	**Narrative-driven Investigation**	**HE + Student Experience**	**•**

Legend: LS = Learning Styles; HE = Heuristic Evaluation; AP = Architectural Patterns; TO = Trade-offs; AT = Architecture Tradeoff Analysis Method (ATAM); SA = Software Architecture Fundamentals.
UG = Undergraduate Students; GE = General Audience/Practitioners.
DEG = Digital Educational Game. • = Digital approach; ○ = Non-digital approach.

address the scenario. This metaphor was designed to make abstract concepts more concrete and to situate the problem-solving process in a professional practice-inspired setting.

To reinforce immersion, the game includes an avatar customization system that lets players assemble their investigator by combining elements like hairstyles, outfits, and accessories. This feature was incorporated as an element, although it is secondary to the pedagogical structure centered on diagnostic reasoning.

Figure 1 summarizes the transition from the baseline version (v1) to the refined version (v2), highlighting the main refinements implemented after analyzing evidence collected during a previous evaluation with students. These refinements addressed user experience, pedagogical support, gameplay flow, and system security.

The implemented changes addressed limitations related to onboarding, navigation, feedback presentation, and progression mechanics. In addition to interface refinements, the updated version introduced server-side validation and Firebase security rules to improve system consistency and reliability.

The content in PatternQuest Game is organized into five districts that represent successive stages of conceptual complexity. The current version of the game includes 39 enigmas distributed across these districts, each addressing recurring software design problems and the architectural patterns commonly used to solve them.

- **Garage District:** Introduces basic principles of modularization and separation of concerns through patterns such as *Layers* and *Model-View-Controller (MVC)*.
- **Commercial Center and Industrial District:** Presents scenarios involving system integration and distributed communication, including patterns such as *Event-Driven Architecture*, *Service-Oriented Architecture (SOA)*, and *Microservices*.
- **Digital District and Central Tower:** Addresses more complex concerns, such as scalability and architectural flexibility, with patterns including *Hexagonal Architecture* and *Proxy*.

This progression is presented through the district map interface shown in Figure 2, where players navigate the city and select challenges based on their current level of advancement.

This organization reflects a scaffolding strategy in which students initially work with localized and concrete design problems and progressively encounter scenarios that involve broader architectural concerns. The objective is to sequence the content from lower to higher levels of abstraction, allowing concepts introduced in earlier districts to be revisited in more complex contexts.

The central gameplay mechanic of PatternQuest Game is a diagnostic reasoning cycle in which students analyze software problems and select architectural patterns that address the identified issues. This structure was designed to emphasize the association between recurring design problems and corresponding architectural solutions. The cycle is organized into three sequential stages that guide the player's reasoning from problem interpretation to the selection of an appropriate architectural solution:

1. **Symptom Analysis:** The player reads a scenario describing a software quality problem, such as performance degradation, excessive coupling, or limited scalability.
2. **Constraint Identification:** Based on the scenario, the player infers the underlying architectural issue or design constraint.
3. **Pattern Selection:** The player chooses the architectural pattern that appropriately addresses the identified problem.

The investigation interface illustrated in Figure 3 presents each scenario and the available pattern alternatives. During gameplay, the Heads-Up Display (HUD) provides real-time information on score, remaining lives, and progress.

Each mission is governed by a limited number of lives and a time constraint. Incorrect answers reduce the number of remaining lives, while the timer establishes a bounded decision-making context. These mechanics were incorporated to discourage random guessing and to encourage more deliberate analysis of the alternatives.

Player advancement is represented by a career system with titles such as "Intern Investigator" and "Senior Architect." The game also includes achievements and a leaderboard based on indicators such as the number of completed enigmas, response accuracy, and solution time. These elements were incorporated as motivational mechanisms to encourage continued participation and are not intended as direct measures of learning.

To support decision making during gameplay, PatternQuest Game provides an integrated Architecture Library (Figure 5) containing textual descriptions and diagrams of the available patterns.

	V1	V2
UXUI Design	Mixed languages (PT/EN) and limited audio control	Full i18n support and individualized audio controls
Pedagogical scaffolding	Text-only patterns, optional theoretical material, and short time windows	Architectural diagrams, mandatory study onboarding, and 60s timers
Core mecanichs	Simple quiz-based progression with potential for question repetition	Expanded question bank with non-repetition logic
System Engineering	Frontend-only validation (*localStorage*) and client-side authentication risk	Server validation via Firebase Security Rules and backend authentication

Figure 1: Comparative mapping of V1 to V2 improvements.

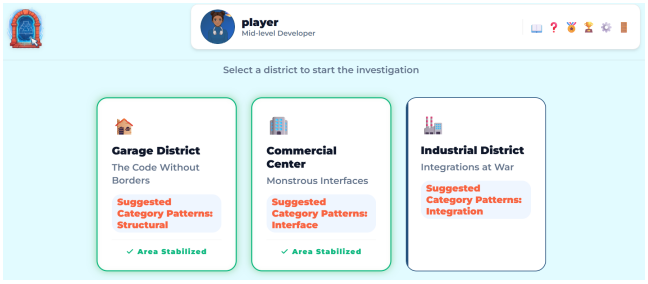

Figure 2: Map interface, where players select challenges.

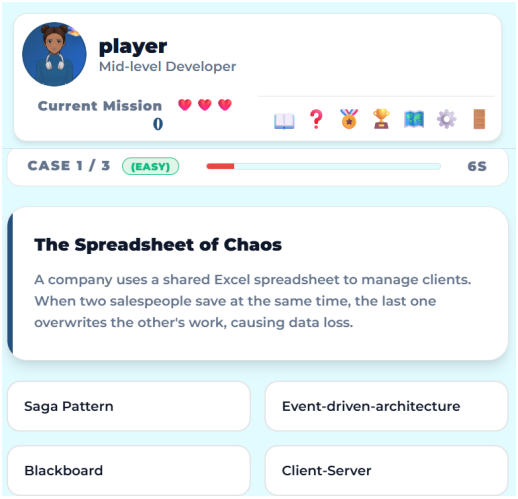

Figure 3: Investigation interface, presenting architectural scenarios, and HUD displaying score, lives, and progression.

Players can consult this material while solving the enigmas, allowing the game to function both as a diagnostic exercise and as a reference resource. In addition, before starting the missions, players have access to a Game Manual (Figure 4) describing the interface, rules, and scoring system. Together, these support materials were designed to reduce the effort required to understand the game mechanics so that students can focus on the architectural challenges.

To facilitate access in educational settings without requiring software installation, PatternQuest Game was implemented as a web-based application using React, TypeScript, and Firebase. The system combines a component-based front-end with cloud services for data storage, authentication, and deployment.

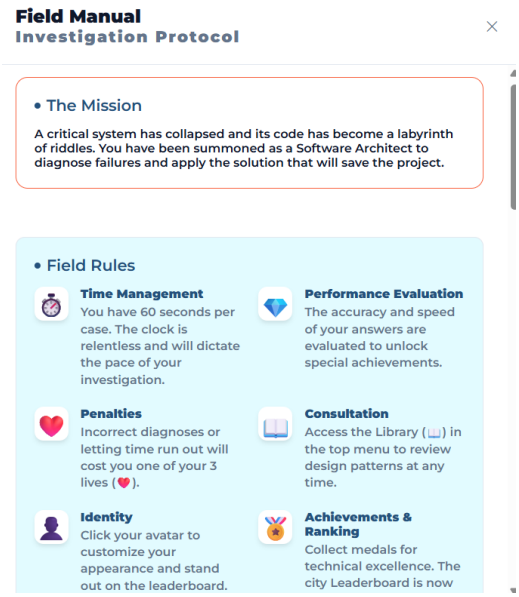

Figure 4: Game Manual explaining rules and progression.

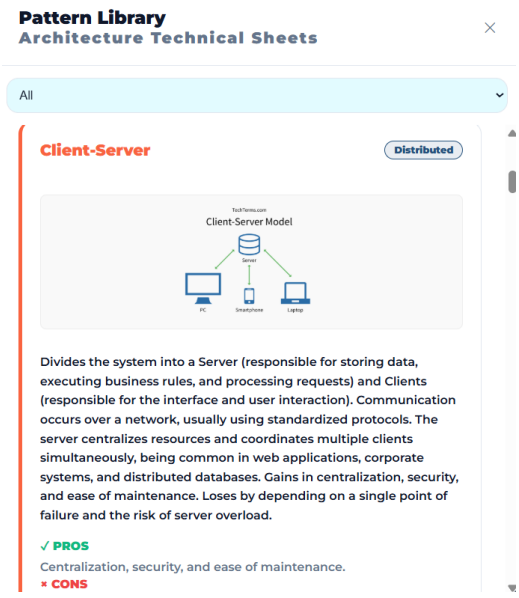

Figure 5: The in-mission Architecture Library.

The SA follows a layered architecture composed of Presentation, Logic, and Data layers, as shown in Figure 6. This separation isolates the user interface, gameplay rules, and persistence services, improving maintainability and allowing challenges, architectural patterns, and districts to be incorporated into the codebase with changes.

The Presentation layer implements the interfaces through reusable React components, including the district map, investigation screen, Game Manual, and Architecture Library. The Logic layer manages the diagnostic cycle, scoring rules, and player progression. The Data layer uses Cloud Firestore to store the enigmas, architectural patterns, and user progress, while Firebase Authentication and Hosting provide user management and application delivery.

The SA was designed to separate user interface components, gameplay logic, and data management services. This organization improves maintainability and facilitates the incorporation of new challenges, architectural patterns, and instructional resources within the game environment.

4 Evaluation

The evaluation of PatternQuest Game was conducted through two complementary studies. First, a previous study reported in [8] evaluated an earlier version of the game with students using the MEEGA+. The results of that study are used in this article as complementary evidence regarding students' perceptions of player experience and learning support. Subsequently, the game version was inspected by specialists through a heuristic evaluation based on AHJED. Together, these studies provide complementary perspectives from both the target audience and domain experts.

4.1 Previous Evaluation Using MEEGA+

In an earlier version of PatternQuest Game , we evaluated the game with 30 undergraduate and graduate students from computing-related programs, as reported in our previous work [8]. The complete responses from that study are available at [2].

The objective of that study was to examine students' perceptions of usability, player experience, and perceived learning after 30 minutes of gameplay.

Table 2 summarizes the descriptive statistics obtained in the student-based evaluation. The highest mean scores were observed in Accessibility ($\mu = 4.88$), Aesthetics ($\mu = 4.77$), Operability ($\mu = 4.63$), Relevance ($\mu = 4.50$), and Perceived Learning ($\mu = 4.42$). These results indicate that participants reported favorable perceptions regarding ease of use, visual presentation, and the educational relevance of the game. As the instrument captures self-reported perceptions, these findings should be interpreted as preliminary evidence of acceptance rather than as direct measures of learning outcomes.

Although the MEEGA+ provided an overview of students' perceptions, it did not identify specific interface and pedagogical design issues. For this reason, we conducted a complementary expert-based study. While the previous study offered a user-centered perspective, the heuristic evaluation was designed to identify concrete usability and instructional issues and to guide refinements in the current version of the game.

[2] https://bit.ly/meegaplus

Table 2: MEEGA+ Descriptive Statistics from the Previous Student-Based Evaluation

Dimension	Items	Mean (μ)	Std. Dev. (σ)
Aesthetics	1–2	4.77	0.46
Learnability	3–5	4.54	0.74
Operability	6–7	4.63	0.64
Accessibility	8–10	4.88	0.42
User Error Protection	11–12	4.17	1.11
Confidence	13–14	4.45	0.72
Challenge	15–17	3.94	1.15
Satisfaction	18–21	4.42	0.84
Fun	25–26	4.25	0.93
Focused Attention	27–29	3.63	1.26
Relevance	30–33	4.50	0.85
Perceived Learning	34–35	4.42	0.91

4.2 Planning of the New Evaluation

The heuristic evaluation was planned to identify interface, gameplay, and pedagogical issues in the current version of PatternQuest Game . For this purpose, we adopted AHJED (Heuristic Evaluation for Digital Educational Games) proposed by Azevêdo et al. [6], as it combines usability and educational criteria within a single inspection framework.

The evaluation considered seven AHJED dimensions: Interface and Usability (IN), Gameplay and Mechanics (JG), Multimedia Elements (MM), Story and Narrative (HJ), Educational Design (ED), Content (CN), and Educational Agent (AE). The questionnaire administered in this study contained 42 heuristics: nine IN, eight JG, five MM, three HJ, eight ED, five CN, and four AE heuristics. The CN5 heuristic, concerning the implicit presentation of educational content through game elements, was not included, and the Social Interaction (IS) dimension was excluded because PatternQuest Game is currently a single-player game and lacks collaborative or multiplayer features.

To operationalize the inspection, we created a questionnaire in Google Forms. For each of the 42 heuristics, specialists assigned a severity score from 0 to 4, where 0 indicates that no problem was identified and 4 represents a critical issue requiring immediate correction. Optional open-ended questions were included to collect descriptions of the detected problems and suggestions for improvement.

We selected the participants based on experience in SA, educational technologies, digital games, and usability evaluation. We defined this profile to ensure that both the technical and pedagogical aspects of the game were examined.

Participation was voluntary, and all evaluators were informed of the study's objectives and the procedures involved before completing the assessment. No personally identifiable information was collected, and all responses were analyzed anonymously. The collected data were used exclusively for academic research purposes and are reported only in aggregated form, ensuring the confidentiality and privacy of the participants.

4.3 Execution of New Evaluation

We conducted the heuristic evaluation with a panel of $N = 5$ specialists, intentionally selected to provide complementary expertise

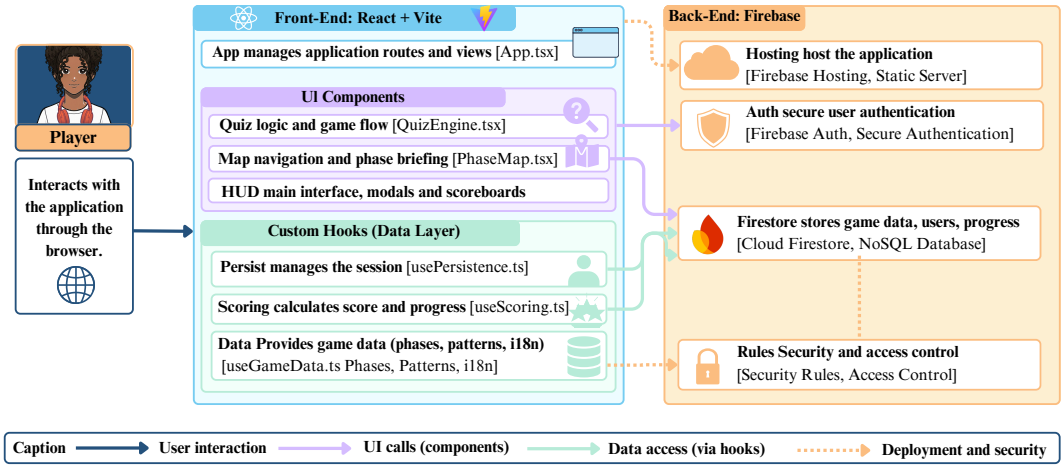

Figure 6: Architecture of PatternQuest Game .

to the study's objectives. This sample size is consistent with recommendations for heuristic evaluations, in which small groups of three to five evaluators are considered sufficient to identify most usability problems while maintaining an efficient inspection process [21]. Four evaluators had advanced knowledge in SA, ensuring that the educational content, architectural patterns, and diagnostic reasoning tasks could be critically assessed from a domain perspective. In addition, one participant reported extensive experience with digital games, contributing a player-centered perspective regarding gameplay, engagement, and immersion. Among the five participants, four had prior experience conducting heuristic evaluations, thereby increasing the reliability and consistency of the identified issues.

We conducted the evaluation remotely via an online protocol. Each participant received access to the playable version of the PatternQuest Game and a link to the Google Forms questionnaire containing the AHJED heuristics. No formal training session or prior explanation of the instrument was provided. We made this decision intentionally, as one of the study's goals was to verify whether the heuristics and questionnaire structure were sufficiently self-explanatory to support independent expert assessment. This approach also simulated a realistic first contact with the game, similar to the experience of future users encountering the system for the first time.

Participants were instructed to interact freely with the game and explore all available features, including the introductory tutorial, case narratives, suspect analysis, clue interpretation, pattern selection, and feedback screens. No time limit was imposed, allowing each evaluator to examine the game according to their own pace and depth of analysis. This strategy aligns with heuristic evaluation practices, in which experts are encouraged to thoroughly inspect the interface and gameplay to identify potential violations of the adopted principles.

After completing the gameplay session, the evaluators completed the AHJED-based questionnaire. Besides assigning severity ratings to each heuristic, participants were encouraged to provide written comments describing the issues encountered, explaining the rationale behind their assessments, and suggesting concrete improvements. These qualitative observations were particularly valuable for interpreting the ratings and understanding the underlying causes of the identified problems.

The use of Google Forms standardized data collection and ensured that all responses were recorded in a structured format. Once all questionnaires had been submitted, the responses were exported for analysis. Quantitative data were analyzed using descriptive statistics for each heuristic and AHJED dimension, while qualitative comments were grouped into recurring themes. Together, these data provided a comprehensive view of PatternQuest Game strengths and weaknesses, supporting refinements to its technical implementation and pedagogical design.

5 Results and Discussion

This section presents the results of the heuristic evaluation conducted with specialists[3] and discusses the implications of the identified findings for the evolution of PatternQuest Game . The analysis combines quantitative severity ratings and qualitative observations obtained through the AHJED evaluation. To answer the research questions, the results are organized into two analytical perspectives: (RQ1) interface and interaction design, and (RQ2) pedagogical design and contextual reasoning.

Figure 7 presents the distribution of severity ratings across the seven AHJED dimensions; each rating corresponds to 20% of the horizontal bar, thus a 40% segment indicates that two of the five evaluators assigned the corresponding severity level. Higher-severity ratings were concentrated primarily in the ED (Educational Design) dimension, which received eight level-3 and three level-4; HJ (Story and Narrative) received two level-4; IN (Interface and Usability) received four level-3; JG (Gameplay and Mechanics) received three level-3; AE (Educational Agent) received one level-3. No level-3 or level-4 ratings were assigned to the Multimedia Elements or Content dimensions.

[3]The complete responses collected through the evaluation questionnaire are publicly available at: https://bit.ly/4f20u42

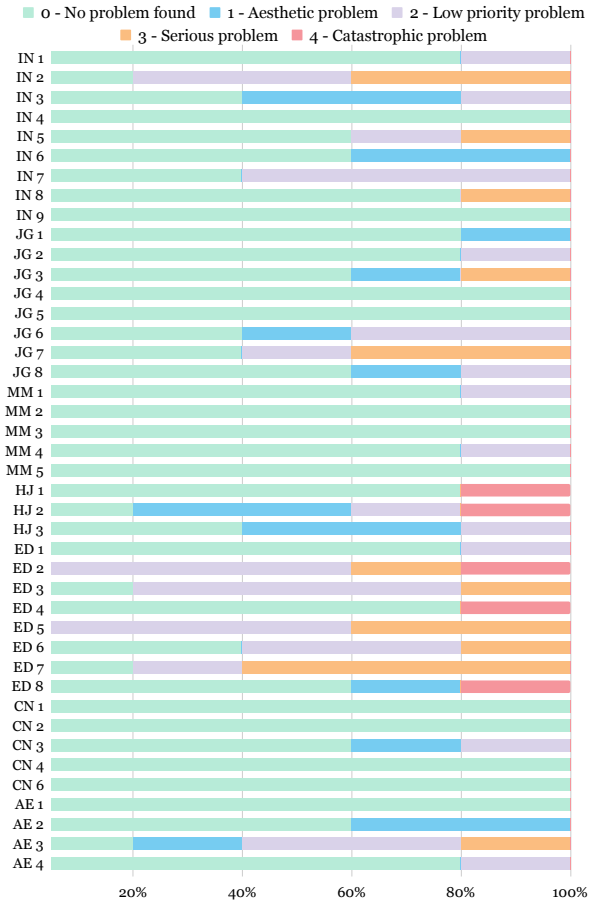

Figure 7: Distribution of heuristic severity ratings across AHJED dimensions.

The findings from the heuristic evaluation are consistent with those from the previous student-based evaluation conducted using MEEGA+ [8]. In that study, participants reported positive perceptions regarding Accessibility ($\mu = 4.88$), Aesthetics ($\mu = 4.77$), Operability ($\mu = 4.63$), Relevance ($\mu = 4.50$), and Perceived Learning ($\mu = 4.42$). While the MEEGA+ results reflect students' self-reported perceptions, the predominance of low-severity heuristic ratings suggests that specialists identified relatively stable interaction and pedagogical structures, providing complementary evidence on the game's usability and educational organization.

Before discussing the identified problems, specialists also reported positive perceptions regarding the overall interaction structure and the contextualized nature of the missions. Some evaluators highlighted achievements, progress systems, and visual organization as motivating aspects of the experience. Others emphasized that narrative-based missions differed from traditional theoretical question-and-answer activities by encouraging the application of architectural concepts in context. These perceptions provide preliminary indications that the game structure may support engagement and contextual reasoning during gameplay.

The findings selected for detailed discussion were not determined solely by the maximum severity assigned to each heuristic. We considered: (i) the recurrence of level-3 or level-4 ratings across evaluators; (ii) the presence of a level-4 rating accompanied by explanatory qualitative evidence; and (iii) open-ended comments describing a direct effect on task completion, gameplay interpretation, narrative engagement, or pedagogical reasoning.

Heuristics receiving level-3 ratings from at least two evaluators were treated as recurring higher-severity findings. Isolated level-3 ratings were reported as localized concerns. Qualitative observations that were not reflected in the corresponding numerical ratings were explicitly identified as emergent qualitative findings rather than quantitative heuristic violations.

5.1 Interface and Interaction Design (RQ1)

The analysis related to RQ1 focused on the IN (Interface and Usability), JG (Gameplay and Mechanics), and MM (Multimedia Elements) dimensions. Overall, specialists were able to complete the missions and interact with the main gameplay features without external assistance, suggesting that the game already provides a stable baseline usability structure for first-time users.

Experts positively evaluated the interface's visual organization and the consistency of interaction elements throughout the missions. One participant stated that *"the visual elements are colorful, and the achievement system motivates progression."* Similarly, evaluators reported that the arrangement of screens and interaction components facilitated navigation and reduced the effort required to understand the gameplay flow. These observations align with the previously reported positive MEEGA+ perceptions for Accessibility, Learnability, and Operability.

Another recurrent aspect in the specialists' comments concerned the integration of instructional materials into the gameplay flow. Evaluators considered the possibility of consulting support materials during missions as a relevant characteristic of the interaction design. One participant highlighted that *"having the possibility to access help materials at any moment"* reduced frustration during the challenges. This finding suggests that integrating instructional support directly into the interface may reduce context switching and help maintain task continuity during gameplay.

The evaluation also indicated that most gameplay mechanics were perceived as intuitive and compatible with common interaction patterns found in digital games. Specialists were generally able to understand the mission progression, clue interpretation, and pattern selection process without formal onboarding or prior training. Considering that the evaluation intentionally omitted tutorial explanations, these observations suggest that the interface and mission structure already communicate core interaction goals with reasonable clarity.

Despite these positive perceptions, the heuristic evaluation identified specific interaction frictions that may affect navigation clarity and gameplay fluency. Table 3 summarizes the most critical heuristic violations identified during the inspection, together with the proposed refinements.

One of the main interaction issues concerned the navigation affordance. Although the game already included a return-to-map feature, some evaluators failed to identify the *Home* button during

Table 3: Selected Higher-Severity and Qualitatively Relevant Findings

Heuristic	Evaluation Focus	Max Severity	Expert Feedback & Proposed Refinement
CN1	Content clarity and lack of ambiguity.	4	**Feedback:** "I didn't understand if I was supposed to answer which pattern the prompt was describing or propose one that would solve it." **Refinement:** Reword mission prompts to explicitly separate the diagnostic (problem) from the therapeutic (solution) phase.
HJ2 / MM4	Narrative immersion and time constraints.	4	**Feedback:** "Due to the short time, I was more worried about responding than engaging with the narrative." **Refinement:** Implement a 'Story Mode' where the timer is paused during narrative reading and is active only during pattern selection.
IN2 / JG3	Player control and navigation freedom.	3	**Feedback:** "There is no explicit 'HOME' button..." **Refinement:** The return button exists, but redesign it with higher contrast, an explicit icon, and standardized placement to improve visual affordance.
ED2 / ED7	Learning autonomy and customization.	3	**Feedback:** "The sequence of missions is predefined... missions seem static without customization." **Refinement:** Allow instructors to create custom mission sequences and unlock multi-mission selection for students.

active missions and incorrectly assumed that the *Refuse* option should be used to exit the mission. This suggests that the visual representation and placement of navigation controls were insufficiently explicit for first-time interaction.

Similarly, some interface feedback mechanisms generated interpretation issues. One evaluator reported a lack of clear auditory feedback during error situations (MM 1), while another misunderstood the countdown timer because its visual representation resembled a conventional progress bar. As one participant noted, *"the progress bar, contrary to the traditional model, was too fast."* This finding suggests that the timer metaphor may not adequately communicate its intended function, particularly under time pressure, leading users to interpret it as a standard progression indicator rather than a countdown mechanism.

Another issue identified during the evaluation involved login and authentication affordances. One specialist reported difficulty recognizing the access point after leaving the game, observing that *"the door does not look like a door."* Although this issue did not prevent interaction completion, it reinforces the importance of using more recognizable visual metaphors and standardized interaction patterns for essential navigation elements.

Overall, the heuristic findings suggest that the current version of PatternQuest Game already provides a relatively stable interaction structure, while also revealing opportunities for refinement related to visibility, feedback communication, and affordance clarity. Importantly, the identified problems did not prevent specialists from completing the proposed activities, indicating that the observed limitations were localized rather than systemic.

The convergence between the heuristic findings and the previous MEEGA+ evaluation strengthens this interpretation. While students previously reported positive perceptions of aesthetics, accessibility, and operability, specialists identified more specific interaction frictions, mainly in visual communication and navigation affordances. Together, these results suggest that the interface supports interaction continuity well, though refinements to feedback presentation and visual signaling could improve usability for novice players.

5.2 Contextual Reasoning and Pedagogical Design (RQ2)

The analysis related to RQ2 focused on the ED (Educational Design), CN (Content), HJ (Story and Narrative), and AE (Educational Agent) dimensions. Overall, specialists positively evaluated the use of contextualized missions and scenario-based reasoning as mechanisms to present SA concepts.

One specialist emphasized the narrative structure of the game, stating: *"The narrative is different, fleeing the conventional theoretical questions and answers. It is great to think about the applicability of the patterns."*. Another reported that *"in the Garage District, I understood the importance of patterns to organize systems that grew in a disorderly manner."*. Together, these comments indicate that the specialists perceived the missions as an alternative to conventional theoretical questions and as an opportunity to reflect on how architectural patterns can be applied to practical software design problems.

Specialists also perceived the use of immediate feedback and integrated instructional materials during gameplay as positive. Multiple participants reported that the option to consult support materials while solving challenges facilitated concept review and reduced interruptions during the activity. One evaluator stated that *"what motivated me the most was consulting the materials, as it facilitated reviewing concepts without leaving the page."* In addition, experts considered the explanatory feedback after correct and incorrect answers useful for reinforcing the architectural concepts explored during the missions.

The game's progression structure was another aspect that participants discussed positively. Evaluators reported that organizing missions around architectural concepts contributed to a gradual increase in difficulty throughout the gameplay experience. One specialist observed that *"the progression of difficulty within the same block and concept was adequate,"* while another stated that *"the questions increase in complexity, which makes the game more interesting."* These perceptions indicate that the mission sequencing and district

organization may support incremental exposure to more complex reasoning tasks.

The evaluation also suggested that some game mechanics may foster learner autonomy during the interaction. Specialists appreciated the ability to replay missions, review instructional content, and retry questions. One participant highlighted that *"the possibility of replaying missions with new questions helps reinforce the content."*. These observations suggest that the gameplay structure may accommodate students with different levels of prior knowledge by allowing repeated interaction with the educational content.

Despite these positive perceptions, the heuristic evaluation identified important pedagogical frictions, mainly related to content interpretation and cognitive load. The most critical issue emerged in heuristic CN 1, associated with content clarity and ambiguity. Some evaluators reported uncertainty regarding whether they should identify the architectural problem presented in the mission or select the pattern that would solve it. One participant stated that *"I didn't understand if I was supposed to answer which pattern the prompt was describing or propose one that would solve it."*

This ambiguity suggests that the diagnostic reasoning loop implemented in the missions was not always communicated clearly to first-time users. From a pedagogical perspective, this issue is relevant because misunderstanding the task objective may compromise the intended reasoning process and increase unnecessary cognitive effort during problem solving. As a refinement strategy, future versions should more explicitly separate the identification of architectural symptoms from the proposal of solution patterns.

Another important issue involved the relationship between time pressure and narrative immersion. Although the countdown timer was intended to simulate urgency and maintain gameplay rhythm, some specialists perceived the mechanic as detrimental to reflective reading and narrative engagement. One evaluator observed that *"due to the short time, I was more worried about responding than engaging with the narrative."* This finding suggests that the imposed time constraints may have shifted player attention toward rapid task completion rather than conceptual reflection.

The proposed refinement for this issue involves separating narrative reading from timed decision-making moments. In practice, future versions may pause the timer during story exposition and activate it only during pattern selection activities. This adjustment may reduce cognitive overload while preserving the challenge of decision-making.

The evaluation also identified opportunities to improve customization and pedagogical flexibility. Some specialists considered the mission sequence overly predefined and suggested mechanisms that allow instructors to customize mission progression based on learning objectives. Although this limitation did not prevent interaction completion, it suggests that the current pedagogical structure may benefit from more flexible sequencing strategies that can support different instructional contexts.

Overall, the heuristic findings provide preliminary evidence that the contextualized, narrative-driven structure of PatternQuest Game may support engagement and diagnostic reasoning regarding SA concepts. However, the evaluation also revealed important limitations associated with task clarity, time pressure, and pedagogical customization, indicating that further refinements are necessary before broader classroom adoption.

5.3 Discussion

The combined findings from the heuristic evaluation and the previous student-based evaluation provide complementary evidence about the quality of PatternQuest Game as a DEG for SAE. While MEEGA+ captured students' perceptions of usability, engagement, and perceived learning, AHJED identified specific interaction and pedagogical issues not captured by perception-based questionnaires. Together, these perspectives provide a broader understanding of the game's strengths and areas for improvement.

Regarding RQ1, the predominance of low-severity violations indicates that PatternQuest Game provides an interaction structure with limited usability barriers. The identified issues were related to navigation affordances, interface feedback, and visual communication rather than fundamental interaction problems. These findings suggest that the game supports effective interaction while revealing refinements to improve the experience for first-time players.

Regarding RQ2, specialists recognized the value of contextualized scenarios, narrative-driven investigations, and instructional resources for supporting architectural reasoning. Unlike traditional quiz-based approaches, PatternQuest Game encourages learners to analyze software symptoms, infer architectural constraints, and justify pattern selection in realistic scenarios, helping connect theoretical concepts with practical decision-making.

These findings align with the previous MEEGA+ evaluation [8], where students reported positive perceptions of Relevance ($\mu = 4.50$), Perceived Learning ($\mu = 4.42$), Confidence ($\mu = 4.45$), Satisfaction ($\mu = 4.42$), and Fun ($\mu = 4.25$). Although these results reflect perceived educational value rather than direct learning outcomes, the heuristic evaluation helps explain how design elements such as contextualized narratives, support materials, progressive missions, and feedback contributed to these perceptions. The convergence between evaluations strengthens the interpretation that these features positively influenced both student perceptions and expert assessments.

The heuristic evaluation also highlighted that educational effectiveness depends not only on content quality but also on how learning tasks are communicated. Ambiguous instructions and excessive time pressure may increase cognitive load and limit reflection, even when the content is considered relevant. Thus, interface and interaction design should be considered alongside pedagogical content when designing DEGs for SAE.

Overall, the combined evidence suggests that PatternQuest Game contributes to SAE by integrating narrative-based diagnostic reasoning with complementary evaluation approaches. Although this study does not provide direct evidence of learning effectiveness, it offers initial evidence of the game's usability, pedagogical organization, and potential to support contextual reasoning about architectural patterns. Future classroom studies with larger student groups and pre- and post-tests are needed to determine whether these characteristics lead to measurable improvements in architectural knowledge and reasoning skills.

6 Threats to Validity

This study has limitations that must be considered when interpreting the reported findings. Following recommendations commonly

adopted in empirical SE studies [20], we discuss the main threats to validity and the strategies adopted to reduce their impact.

Construct Validity: This threat concerns whether the evaluation measured usability, gameplay quality, and pedagogical adequacy. To mitigate it, we adopted the AHJED framework [6], specifically designed for DEGs. The questionnaire combined quantitative severity ratings with qualitative comments, enabling the triangulation of numerical scores and textual feedback. Furthermore, evaluators were selected for complementary expertise in SA, usability, educational technologies, and digital games. Nevertheless, heuristic evaluations inherently rely on subjective perceptions, which may introduce evaluator bias.

Internal Validity: Internal threats relate to factors that may have influenced the results. The evaluation was conducted remotely and asynchronously, without researcher supervision. Although this reduced researcher interference and approximated a first-contact scenario, participants may have interacted with the DEG under different environmental conditions and exploration strategies. To minimize inconsistencies, all evaluators received the same instructions, questionnaire, and game version. Additionally, the absence of onboarding may have amplified first-time interaction difficulties, while the countdown timer may have increased stress and affected reflective reading during missions, potentially influencing severity ratings of narrative engagement and gameplay flow.

External Validity: This threat concerns the generalization of the findings. The main limitation is the sample size ($N = 5$) and the evaluator profile. Although five specialists are consistent with recommendations for heuristic evaluations [21], they do not represent the target audience of undergraduate students, who may interpret interface elements, gameplay, and architectural concepts differently. To mitigate this limitation, the panel included complementary expertise in SA, educational technologies, usability, and digital games, providing a broader inspection perspective. Nevertheless, the findings should be interpreted as preliminary evidence from expert inspection rather than classroom use. Moreover, the current version of PatternQuest Game includes a limited set of scenarios and architectural patterns, and the evaluation considered only short-term interactions, leaving long-term adoption, learning retention, and content expansion for future investigation.

Conclusion Validity: To reduce interpretation bias during data analysis, heuristic violations and proposed refinements were cross-referenced with evaluators' qualitative comments and severity ratings, ensuring that findings remained grounded in the inspection evidence. The study does not claim actual learning gains or educational effectiveness, as heuristic evaluations primarily identify usability, gameplay, and pedagogical design issues. Future classroom studies will be necessary to investigate perceived learning, engagement, and learning outcomes associated with the DEG.

7 Conclusion and Future Work

This paper presented PatternQuest Game , a DEG designed to support SAE through narrative-driven missions centered on architectural pattern selection and diagnostic reasoning. The game was designed to contextualize SA concepts within investigative scenarios in which players analyze software problems and select architectural patterns based on the constraints and quality concerns.

This work described the evolution of the DEG after a previous student-based evaluation and reported a complementary heuristic evaluation with specialists using AHJED. The evaluation combined severity ratings and qualitative feedback to investigate interaction, gameplay, and pedagogical aspects.

Regarding RQ1, the heuristic evaluation indicated that specialists completed missions and interacted with gameplay features without assistance. Most violations were low severity, suggesting a stable interaction structure, while navigation, feedback, and timer visualization issues revealed usability improvement opportunities.

Regarding RQ2, specialists positively perceived contextualized missions, support materials, and progressive challenges as mechanisms for presenting architectural concepts. The evaluation highlighted narrative scenarios and in-game materials as alternatives to conventional activities, while also revealing limitations related to mission ambiguity and time pressure effects on reflective reading and engagement.

The findings should be interpreted considering the scope and limitations of the evaluation methods. The results provide preliminary evidence of perceived usability, gameplay experience, and pedagogical organization based on heuristic inspection and participant perceptions, but do not demonstrate learning gains or retention. Moreover, the limited number of specialists and short-term evaluation sessions restrict the generalization of findings to broader educational contexts and long-term classroom adoption.

As future work, we plan to implement the refinements identified during the heuristic evaluation, including improvements to mission clarity, navigation affordances, gameplay indicators, and timing mechanics. We also intend to investigate alternative gameplay configurations with interaction flows.

To address ambiguity identified during the heuristic evaluation, future versions will better separate problem diagnosis from architectural pattern selection through clearer guidance and interface cues. We also plan to support instructor-customized missions aligned with learning objectives and course contexts.

Future versions of PatternQuest Game will also explore collaborative learning features associated with the Social Interaction (IS) dimension of the AHJED framework, including cooperative investigation activities and peer discussion mechanisms to support reflection on architectural decisions and trade-offs.

Finally, future studies will involve classroom evaluations with undergraduate SE and Computer Science students to investigate perceived learning, engagement, usability, and learning outcomes. These studies will employ pre- and post-tests, comparisons with traditional instruction, and longer-term classroom use to assess learning gains, architectural reasoning, and the educational potential and limitations of narrative-driven DEGs in SAE.

Artifact Availability

The artifacts generated and analyzed during this study are publicly available. The game can be accessed at https://patternquestgame.web.app/ . The data from the previous student experience evaluation using MEEGA+ is available at https://bit.ly/meegaplus , and the complete responses collected through the expert heuristic evaluation are available at https://bit.ly/4f20u42 .

References

[1] Deniz Akdur. 2022. Analysis of Software Engineering Skills Gap in the Industry. *ACM Trans. Comput. Educ.* 23, 1, Article 16 (Dec. 2022), 28 pages. doi:10.1145/3567837

[2] Adryan Araújo, Adrian Coelho, Maria Rodrigues, Windson Viana, and Anna Marques. 2024. Todo Esforço Será Recompensado: Gamificação no Ensino de Arquitetura de Software com o Uso de Badges. In *Anais do XXXII Workshop sobre Educação em Computação* (Brasília/DF). SBC, Porto Alegre, RS, Brasil, 341–352. doi:10.5753/wei.2024.3064

[3] Len Bass, Paul Clements, and Rick Kazman. 2021. *Software architecture in practice.* Addison-Wesley Professional, New Jersey.

[4] Luis Caminha and Anna Marques. 2024. D-LEARN: A digital game for Software Architecture education. In *Anais do XX Simpósio Brasileiro de Sistemas de Informação* (Juiz de Fora/MG). SBC, Porto Alegre, RS, Brasil. https://sol.sbc.org.br/index.php/sbsi/article/view/30883

[5] Eduarda Carvalho, Jacson Barbosa, Renato Bulcão Neto, Valdemar Graciano Neto, and Pedro Valle. 2025. DECORA: Um Sistema de Apoio ao Ensino de Decisões de Projetos Arquiteturais. *Anais do V Simpósio Brasileiro de Educação em Computação* 5 (2025), 719–729. doi:10.5753/educomp.2025.5429

[6] Max André de Azevêdo Silva, Danielle Rousy Dias da Silva, and Clauirton de Albuquerque Siebra. 2018. AHJED - Avaliação Heurística para Jogos Educacionais Digitais. In *Nuevas Ideas en Informática Educativa*, Vol. 14. TISE, Santiago de Chile, 126–136.

[7] Remco C De Boer, Patricia Lago, Roberto Verdecchia, and Philippe Kruchten. 2019. Decidarch v2: An improved game to teach architecture design decision making. In *2019 IEEE International Conference on Software Architecture Companion (ICSA-C).* IEEE, Hamburg, Germany, 153–157.

[8] Maria Clara Ribeiro de Menezes, Lara da Silva Dias, Pedro Henrique Dias Valle, and Alessandreia Marta de Oliveira. 2026. PatternQuestGame: A Digital Game to Support the Architectural Patterns Education. In *Proceedings of the 2026 IEEE Frontiers in Education Conference (FIE).* IEEE, Paphos, Cyprus.

[9] Maria Clara Ribeiro de Menezes, Pedro Henrique Dias Valle, and Alessandreia Marta de Oliveira. 2026. A Card Game for Architectural Patterns Education. In *Simpósio Brasileiro de Educação em Computação (EDUCOMP).* SBC, 715–728.

[10] Carlos Freitas Filho, Anna Marques, and Lince Oliveira. 2025. Desvendando as Mentes dos Jogadores: Utilizando a Escala HEXAD para Identificar Preferências de Gamificação no Ensino de Arquitetura de Software. In *Anais do XXXIX Simpósio Brasileiro de Engenharia de Software* (Recife/PE). SBC, Porto Alegre, RS, Brasil, 599–609. doi:10.5753/sbes.2025.11271

[11] Matthias Galster and Samuil Angelov. 2016. What makes teaching software architecture difficult?. In *Proceedings of the 38th International Conference on Software Engineering Companion* (Austin, Texas) *(ICSE '16).* Association for Computing Machinery, New York, NY, USA, 356–359. doi:10.1145/2889160.2889187

[12] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. 1994. *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley Professional, Reading, MA, USA.

[13] Lina Garcés and Brauner R. N. Oliveira. 2025. Teaching software design patterns: An experience using the Jigsaw classroom. In *Proceedings of the XXXIX Brazilian Symposium on Software Engineering (SBES 2025).* SBC, São Carlos, SP, Brazil, 1–10. doi:10.5281/zenodo.16161087

[14] Mauricio Hidalgo, Hernán Astudillo, and Laura M Castro. 2024. A Study of Software Architects' Cognitive Approaches: Kolb's Learning Styles Inventory in Action. In *International Conference on Applied Informatics.* Springer, Morocco, 231–246.

[15] Cindy E. Hmelo-Silver. 2004. Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review* 16, 3 (2004), 235–266. doi:10.1023/B:EDPR.0000034022.16470.f3

[16] John M Keller. 2009. *Motivational design for learning and performance: The ARCS model approach.* Springer Science & Business Media, New York.

[17] Amruth N. Kumar and Rajendra K. Raj. 2023. *Computer Science Curricula 2023 (CS2023): Community Engagement by the ACM/IEEE-CS/AAAI Joint Task Force.* Association for Computing Machinery, New York, NY, USA. 1212–1213 pages. doi:10.1145/3545947.3569591

[18] Andrzej Marczewski. 2021. Hexad: Um Framework de Tipos de Jogadores para o Design de Gamificação. Gamified UK. Tradução de G. F. Tondello. Trabalho original publicado em 2018.

[19] Claudia Hidalgo Montenegro, Hernán Astudillo, and María Clara Gómez Álvarez. 2017. ATAM-RPG: A role-playing game to teach architecture trade-off analysis method (ATAM). In *2017 XLIII Latin American Computer Conference (CLEI).* IEEE, Cordoba, Argentina, 1–9. doi:10.1109/CLEI.2017.8226416

[20] Erica Mourão, João Felipe Pimentel, Leonardo Murta, Marcos Kalinowski, Emilia Mendes, and Claes Wohlin. 2020. On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and software technology* 123 (2020), 106294.

[21] Jakob Nielsen and Rolf Molich. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 249–256.

[22] Anderson Oliveira, Willian Oizumi, Leonardo Sousa, Wesley K. G. Assunção, Alessandro Garcia, Carlos Lucena, and Diego Cedrim. 2022. Smell Patterns as Indicators of Design Degradation: Do Developers Agree?. In *Proceedings of the XXXVI Brazilian Symposium on Software Engineering (SBES 2022).* ACM, New York, NY, USA, 311–320. doi:10.1145/3555228.3555243

[23] Giani Petri, Christiane Gresse von Wangenheim, and Adriano Ferreti Borgatto. 2017. *MEEGA+, Systematic Model to Evaluate Educational Games.* Springer International Publishing, Cham, 1–7. doi:10.1007/978-3-319-08234-9_214-1

[24] Tamires A. S. Sousa and Anna B. S. Marques. 2020. LEARN Board Game: A game for teaching Software Architecture created through Design Science Research. In *Proceedings of the XXXIV Brazilian Symposium on Software Engineering* (Natal, Brazil) *(SBES '20).* Association for Computing Machinery, New York, NY, USA, 834–843. doi:10.1145/3422392.3422494

[25] Maurício Souza, Renata Moreira, and Eduardo Figueiredo. 2019. Students Perception on the use of Project-Based Learning in Software Engineering Education. In *Proceedings of the XXXIII Brazilian Symposium on Software Engineering* (Salvador, Brazil) *(SBES '19).* Association for Computing Machinery, New York, NY, USA, 537–546. doi:10.1145/3350768.3352457

[26] John Sweller. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12, 2 (1988), 257–285. doi:10.1016/0364-0213(88)90023-7

[27] Aderbal F. Zorzo, Daltro Nunes, Edelberto Matos, Igor Steinmacher, Jair Leite, Renata M. Araujo, Ricardo Correia, and Simone Martins. 2017. *Referenciais de Formação para os Cursos de Graduação em Computação.* Technical Report. Sociedade Brasileira de Computação (SBC), Porto Alegre, RS, Brasil. 153 pages. doi:10.5753/sbc.ref.2017.134