Universiteit Leiden
The Netherlands

Bachelor DSAI

Generative AI as a
Content Generator for Educational Games

Anthony Rietveld

Supervisors:
Joost Broekens & Mike Preuss

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

15/12/2025

Abstract

Secondary school students often struggle to remain focused on their studies, as external distractions and traditional method of instruction reduce their motivation to learn. Serious games development aims to combat this issue by creating interactive experiences that combine engaging gameplay with educational objectives. Generative AI can also be integrated into serious games in order to generate custom content that contributes to a more personalized experience. This project investigates how AI could be applied to various genres of serious games in order to enhance them. A focus group consisting of seven Dutch secondary school and higher education students voiced their preferences for the eventual prototype game, which guided the final design. Their interaction with the game was used for an empirical analysis of the game, alongside a statistical analysis based on Davis's Technology Acceptance Model (TAM). The analysis shows that the Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) were statistically significantly above a neutral midpoint, suggesting that students evaluated the game more positively than neutral on average. However, the empirical analysis of the user experience indicates that several AI-related issues still need to be resolved before the game can function as an effective educational tool.

Contents

1 Introduction

1.1 The Problem Context

Student Motivation Towards Learning

Studying is often perceived as a tedious and demanding task by many secondary school students. Some studies suggest that this is due to the use of smartphones, which distract students from paying attention during lectures and study sessions [Bul23] [BD18]. Others researchers argue that the traditional instruction methods are not well suited to the students. They argue that a lack of involvement from the students part in the learning process will lead them to become demotivated to learn, as students struggle to find the practical application for study material [Kho17] [Ami25]. It is thus imperative to find solutions for the problem of poor student motivation.

1.2 Research Motivation

Educational Games

The design of Educational Games provides an opportunity to motivate students through the use of an engaging learning environment. Jackson et al. [JOMJ18] highlight how such games can be more effective than traditional instructional methods. The results of their review suggest that narrative driven educational games promote attitude change, engagement, motivation and skill acquisition. Educational games also appear to be slightly more effective in promoting enjoyment and knowledge acquisition compared to traditional methods. However, a careful balance between the educational value and entertainment value of an educational game should be struck, as an overemphasis on either aspect can diminish the effectiveness of the other [EAD+11]. When a balance between the two factors is achieved, educational games facilitate the subconscious absorption of knowledge, such as through emotional engagement and strategic decision making [PCB92]. This results in a deeper level of learning compared to the passive learning of traditional media [GAD02].

Role of Generative AI

Generative AI can enhance the experience of educational games and support student motivation ny providing personalized content suited to the needs of the user. Mittal et al. [MSCS24] in *A Comprehensive Review on Generative AI for Education* showed how traditional schooling systems often struggle to give students what the authors describe as "a curious heart". They argue that the traditional methods not only fail to inspire genuine curiosity but are also unable to provide the attention needed for each individual student. The study suggests that generative AI could address both of these issues by invoking curiosity into students when it is actively used. Furthermore, the dynamic and generative nature of AI-produced content creates an engaging and personalized learning environment that motives students to learn more deeply.

1.3 Research question

Main Research Question

> *Do secondary school students find an educational game with AI content generation useful and easy to use?*

In order to answer the main research question, this thesis examines three subquestions. A prototype text adventure game with AI content generation was developed in order to facilitate a practical analysis of the topic. Seven teenagers agreed to function as the focus group for the experiment. They were instructed to interact with the prototype game and share their thoughts on it through the use of a questionnaire. Their feedback gave insight into the engagement and educational values of the game.

Research Subquestions

- **(SQ1)** In what ways can generative AI be used to enhance different types of games?
- **(SQ2)** Which type of game or design elements does the focus group want?
- **(SQ3)** How could a game be tested for its usefulness and ease of use?

1.4 Research Objectives & Thesis overview

This research project aims to evaluate how an educational game with an AI content generator could be constructed and tested for its usefulness and ease of use. Chapter 2 will cover various works related to the field of serious educational games development and those covering the use of AI in educational context. Chapter 3 covers four theoretical frameworks for serious games with AI content generation. It also covers the design of the final prototype text adventure game, based on the preferences of the focus group. Chapter 4 details the results of the statistical analysis, alongside the users' empirical feedback and their opinions on the game. Chapter 5 concludes the thesis whilst also discussing the limitations and suggestions for future work.

2 Background and Related Work

This chapter outlines various works related to the field of educational game design, as well as those covering the use of AI in an educational context.

2.1 Making a game of the Old Testament

Engström et al. [EAD+11] outline the development of an example educational game, as well as emphasizing the importance of aligning the design with the goals of authenticity, education and entertainment. The research article discusses the process of creating an action role-playing game, centered around various events from the Old Testament. A small team of 12 final-year students from computer game development programmes at the University of Skövde developed the game whilst discussing the design with the Church of Sweden, who came up with the initiative in response to the dropping confirmation rates of the Swedish youth. The article noted how the collaboration of the two parties led to the creation of three distinct goals, which were used to guide the final design of the game:

Goals of the Old Testament game:

- Authenticity: As the Old Testament game takes inspiration from the Bible, it has to be faithful in its representation of the various stories. To alter the source, which holds an important status in both past and present societies, would risk upsetting a lot of people and would go against the wishes of the Church of Sweden's initiative.
- Education: The game was funded and designed for its use during confirmation, a period in which Christian teenagers (around 14-15 years old) learn about Christianity, the Bible and the values of their religion. It would therefore stand to reason that a game set in Bible will have the player reflect on the game's content and thus learn more, compared to non-specialized commercial games.
- Entertainment: Whilst the game was made for the purposes of education, it would still need qualities similar to commercial computer games in order to attract adolescents such that they attend confirmation.

These goals can be translated to this study design, serving as the general objectives that need to be kept in mind for an educational game. In this case, a game would have to be truthful to source material so as to prevent a student from learning incorrect information. A educational game must also be entertaining enough to keep the users attention, or it would be unable to effectively support meaningful learning.
The article also suggested focusing on a well-defined game genre, which will serve as a reference point to compare the quality of entertainment values against.

2.2 Jesse Schell "A book of lenses"

The Art of Game Design [Sch08] presents the reader with various general guidelines on how one should approach the development of a game. Below is a list of chapters that are most relevant for the design of this study:

- Chapter 6: The game begins with an idea. Schell argues that most ideas are weak and so it is important for good designers to evaluate their ideas. By defining the constraints and goals for your game, you are able to more quickly arrive at a clear game design idea. You may also look at other designs and iterate on existing ideas.
- Chapter 8-9: A game is designed with the player in mind. You should identify your target player, what they would like to see from a game and what skills they posses. A mismatch between a player's mental model and understanding of the game can lead to a frustrating experience.
- Chapter 25: Good games are created through playtesting. Only once people sit down and play your game, will you gain a true understanding of the experience which you have created. Besides observing how players interact with the game, you may also use a survey or conduct a post-game interview.

2.3 Serious Games Development

Educational games fall under the category of serious games, which Zyda [Zyd05] describes as games with an ambition to give the user more than a purely entertaining experience. Various articles also show that motivation is higher for students who play and learn through games rather than through "standard" media ([GAD02], [PCB92]). Petranek et al. argue for the use of computer based learning, stating participants unconsciously process all types of information: facts, emotions, strategies, outcomes, relationships, feelings, and much more. Wouters et al. [WVNVOVDS13] meta-analysis specifies where these upsides of educational games lie:

> "Additional moderator analyses on the learning effects revealed that learners in serious games learned more, relative to those taught with conventional instruction methods, when the game was supplemented with other instruction methods, when multiple training sessions were involved, and when players worked in groups."

2.4 AI in education

AI has been extensively used in education to create personalized learning experience that adapt to the users needs. A systematic review by Duarte et al. [DPBG23] argues how AI is able to deliver personalized feedback in real time as one of most promising benefits of this technology. In addition, AI platforms could dynamically adjust their educational materials based on individual student performance, cognitive abilities and engagement levels [PSL23].
However, some researchers caution the excessive use of AI, stating that students may become reliant on it. One study by Lehman et al. [LCS24] on LLM's showed that while the AI can improve learning when used as a tutor, students who rely on it to do exercises for them showed worse learning outcomes. Students may also overestimate how much they have learned when relying to heavily on the LLM's.

3 Methods & Experimental Design

This chapter details the various genre frameworks for educational games, the integration of generative AI into these frameworks, and the final text adventure prototype. The first part of this chapter covers the various frameworks which were considered for the final prototype, as well as how generative AI could be used the enhance the experience. The second part will go over the final prototype, its internal architecture, and how the experiment was conducted.

3.1 The Choice of a Genre

A genre is imperative to the design of any game, as the expected mechanics associated with the genre guide the general flow of the gameplay loop. By focusing on a specific genre, or even on a particular game, developers have clear point of reference to build upon [EAD+11]. Serious games differ from commercial games, by having additional values beyond pure entertainment. It is therefore not required for a serious game to be innovative in its design. Kniestedt et al [KGML+21] stated that:

> "The results show that adding additional game mechanics to a core gameplay loop did not lead to participants playing more or longer, nor did it improve their game experience."

Educational games will therefore follow most of the core features associated with a genre, as the fundamental gameplay elements are already well established. William et al. [GM17] argue for the use of genres as boundaries, in order to support effective research within the field of game design research. These foundations help researchers place their work in a meaningful context and communicate their findings more effectively. This makes it easier to compare studies following similar genres. Lastly, the choice of a genre also accounted for the development time and complexity of a game. In practice, this meant selecting a genre in which a typical game could reasonably be produced by a single developer, with room to incorporate feedback quickly to support rapid prototyping.

3.2 Genres and AI application

The following section covers various genres that were considered for the final educational game prototype. It will also discuss how Generative AI could be used to enhance the player experience for each of the different genres.

3.2.1 Quiz Game

Quizzes are a simple but effective type of serious game, because the underlying mechanics closely resembles the tests and assessments used in schools (for specific subjects such as for example history, social studies and geography). The core gameplay will have its player answer one-off questions based solely on their prior knowledge, which in the case of school assessments must be gained through studying. *Trivial Pursuit* (and the dutch version *Triviant*) are examples of entertaining quiz games. Players are asked to answer questions from various different categories in order to win. Similar to this, in the Dutch tv-quiz show *De Slimste Mens*, contestants are tested on their general

knowledge as they gain points by answering questions correctly. The person who gains the highest amount of points wins the game.

Whilst there does not exist one generally agreed upon design framework for the quiz game genre, Abejide Ade-Ibijola et al [AIA18] provides a simple design for an educational quiz system. Their 'Abeced' game uses the following flow diagram:

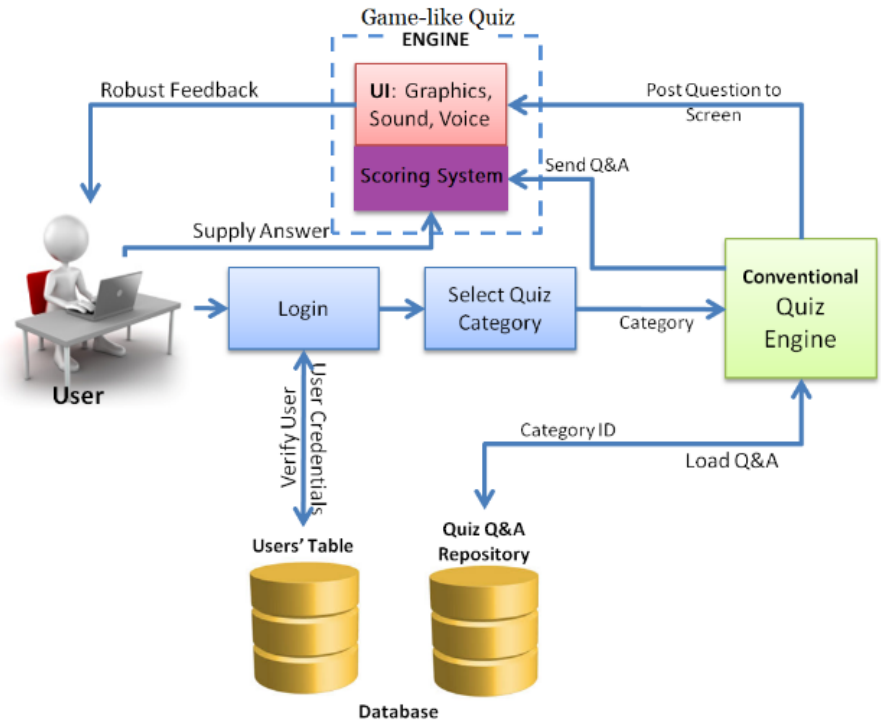

Figure 1: The 'Abeced' flow diagram. Users are able to login and choose the quiz category. From there, a question is fetched from the Quiz repository which the user must answer. The user is given feedback through the UI, in the form of visuals and sounds.

The 'Abeced' uses a predefined Quiz Q&A repository for both category and question selection. Scoring is done through the use of Levenshtein distance algorithms which measures string similarity by calculating the minimum number of single-character edit operations required to transform one string into another [Aun19], however it does not account for the semantic similarities between strings.

Generative AI can be adapted to this framework, serving as the quiz's category selection and question generator. An LLM for instance, can dynamically generate questions from user provided material, which makes the content adaptive to the user preference unlike a static repository [DWB+24] [IGK+25]. In addition, the AI can also be used to grade the answers of students as Luigi Di Caro et al [DCVM+23] found that GPT LLM model can capture significant semantic information, with performance highly dependent on contextual information. This removes the original flaw of the 'Abeced', where semantically similar strings would be incorrectly marked as wrong. The following

shows a simple modified flow diagram, where generative AI was integrated into the interactive process:

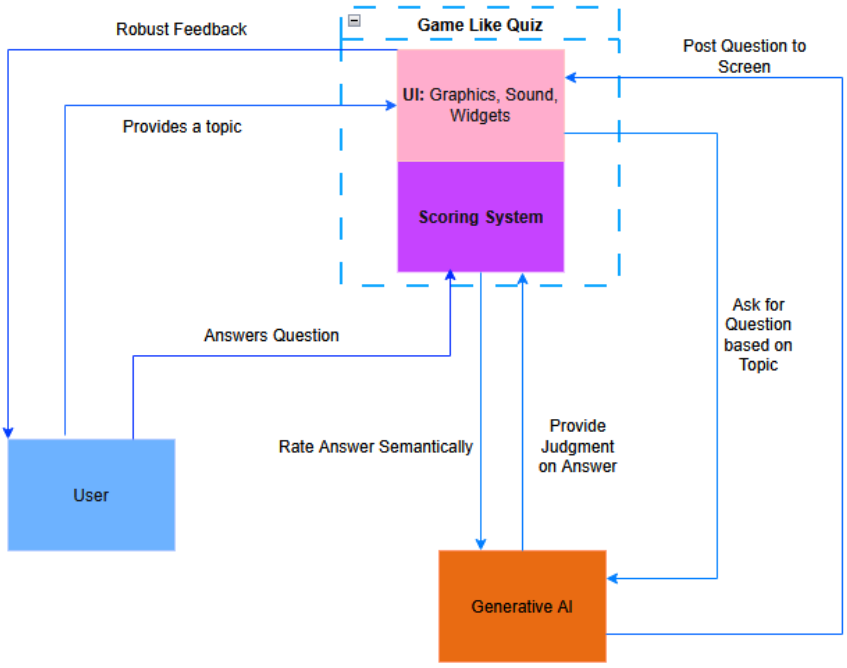

Figure 2: The modified Quiz flow diagram. Whenever the user provides a topic or answers a question, a corresponding request is send to the generative AI system. This will provide generated questions and feedback to answers.

3.2.2 Text Adventure

Games of the text adventure genre give players meaningful choices that influence the story's progression, which creates a sense of personal involvement that boosts engagement. A literature review by Tyler B Wright et al. [WW24] also uncovered how this interactive fiction (IF) has the additional benefit of improving literacy by encouraging active participation, as students have to consider more details of the story. This approach is especially useful for studying history, as it allows players to step into important historical periods and explore them as if they were experiencing them first hand.

An example of an IF for entertainment would the game *AI Dungeon*, where players may choose their own course of action through a story. The game asks for a setting from the user, which it will use as a background for the story [Fer25]. Players will be asked what actions they would like to take in each different scenario, which will influence the direction of the story. This game provides the basis for the general flow of an IF game, which could be slightly modified to function as serious game. This is shown in the following flow diagram:

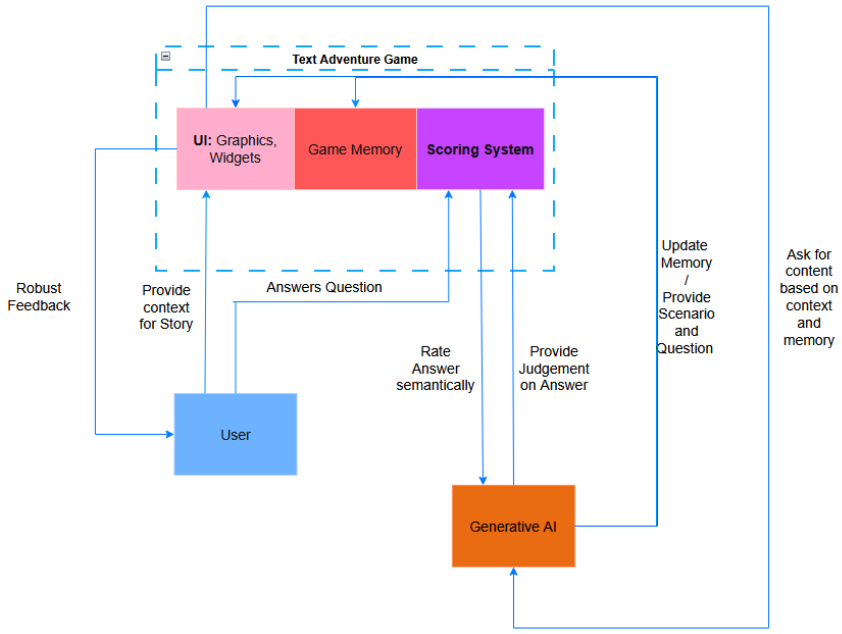

Figure 3: A text adventure flow diagram. User provide a context for the overarching story, and answer questions. The game then uses an AI system to generate the story. The game memory will be updated with previously used scenarios in order to avoid repeating them.

The user interacts with the UI elements of the text adventure game, which connects to an AI system backend that generates stories and questions. A memory is used to keep track of used scenarios and questions, so it does not repeat them again. The system also reuses the scoring system of the quiz game framework alongside the question generation mechanisms. The stories and actions may also not stray to far from true events, as in this would lead to learning false information in the case of for example history.

3.2.3 Board Game

Board Games are commonly defined as games in which players move one or more pieces on a marked board according to established rules. Beyond this broad definition however, board games and their mechanics may vary vastly from one another. Whilst certain board games promote engagement through complex strategic depth [CK15] or team work [Kat25], they were not considered for this research project as the additional mechanics might conflict with the primary goal of learning a subject. The final flow diagram is therefore based on a combination of *Game of the Goose* and *Tivial Pursuit*. Players must throw dice in order to advance their pieces on the board. The positions which players land on determine the difficulty of a question with the exception of special positions that effect the pieces instead (similar to to those in *Game of the Goose*. The player will be asked a question after moving their piece. Should the player answer this question correctly, they may advance additional positions on the board. Question of higher difficulties would reward

more positions than those of lower difficulties. However, the player must move the piece back one position if they answers a question incorrectly. Randomization keeps the game engaging through the dice rolls and different questions, whilst a players knowledge on a chosen topic will be rewarded as additional positions in the event of a correctly answer question. The flow diagram of this design is shown below:

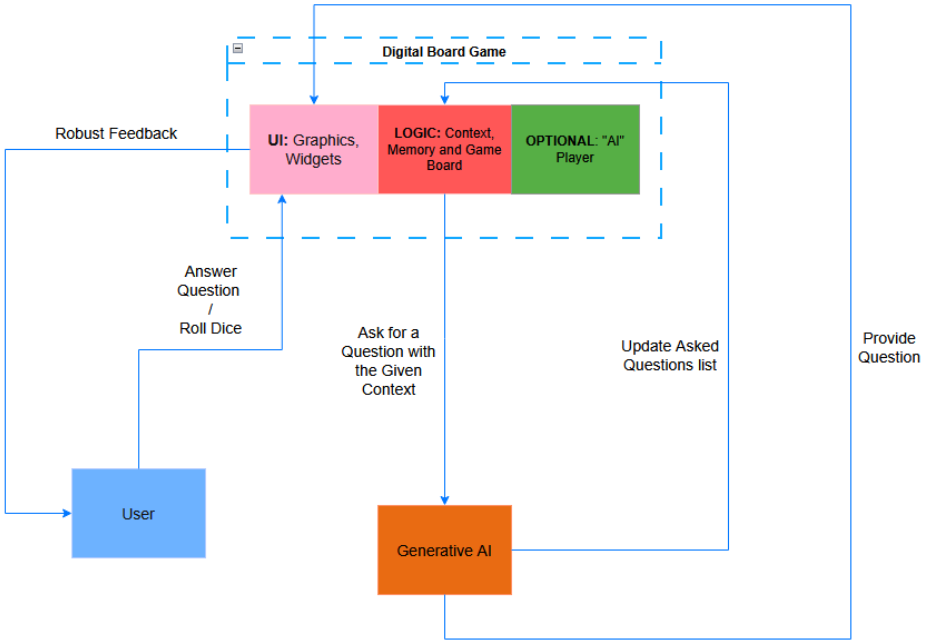

Figure 4: A board game flow diagram. Users provide context for the questions. They interact with a digital dice in order to move their board pieces. Generative AI is used to generate questions and update the context (what questions were already asked). An optional AI agent may play alongside the player.

Generative AI will be used in this interaction to generate questions and update the list of already asked questions. In addition, an optional AI agent may be introduced into the game which creates a friendly competitive environment for the player. This may be in the form of a simple rule base agent however, as the theoretical board game is sufficiently simple enough.

3.2.4 Card Game

Card games form a very broad genre, as the only requirement is that they involve playing cards the player can interact with. Card games, much like board games, promote engagement through strategic depth, although this is dependent on the context of the game [KGML+21]. An element of competition may also boost the engagement of the player with the game [Rom12]. Kristan et al. [KBCVdC20] suggest that an AI opponent could enhance the experience if it makes use of Dynamic Difficulty Adjustment (DDA). This enables the game to be challenging enough to engage experienced player, but not so difficult that it becomes frustrating. Otherwise, the game risks the

player disengaging entirely, which fundamentally contradicts the principles of serious games research.

The main inspiration for the card game design framework was the game *Greatest* by Beppe et al (2018) [BdAA+18]. The game was originally designed for teaching software testing to students, as they would have to match the appropriate test card to the corresponding scenario card. This could be adapted to standard secondary school subjects, where scenarios and questions may be based on the chosen topic. Testing cards could instead be transformed into hint cards, giving the player small pieces of information to better answer the question. The goal is to reach a certain amount of points gained through answering questions correctly. A LLM may be utilized as generative AI in order to create both questions and hints, depending on the cards that the player has. On top of that, a DDA agent will serve as a challenging but fair opponent. This results in an engaging yet still primarily education focused game with the following flow diagram:

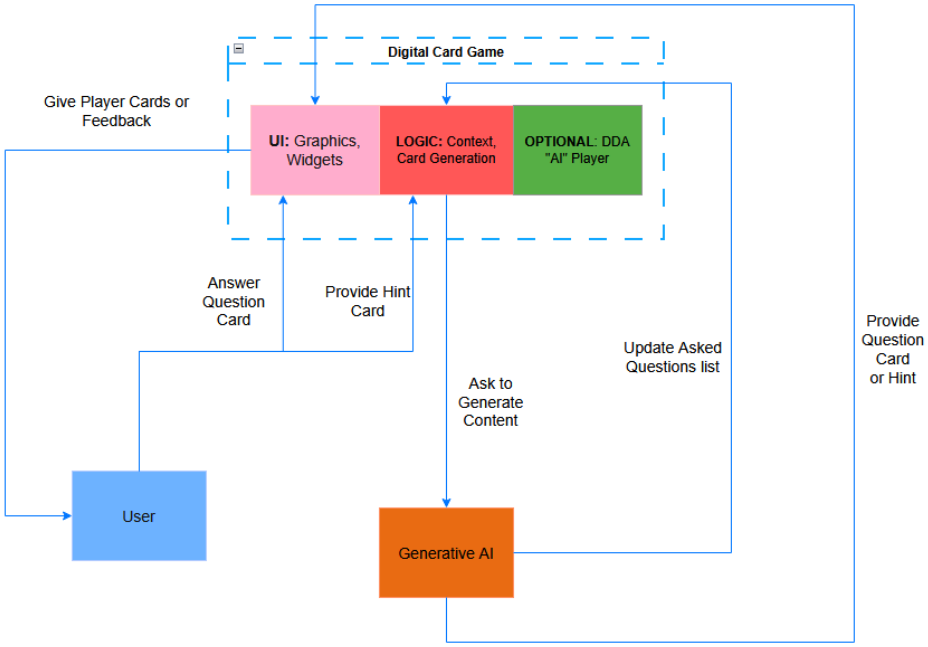

Figure 5: A card game flow diagram. The user provides context for the question cards and interacts with the game by drawing cards from the UI elements, and handing in hint cards if they posses them. Generative AI is used to generate questions and hints for the game, with questions being stored such that they are not asked again. A Dynamic Difficulty Adjustment (DDA) agent creates a friendly comitative environment for the user.

Table of the considered genres

Genre	Description	AI Application
Quiz	The player is asked questions about a chosen topic, which they must answer correctly in order to receive points.	• Question generation • Custom feedback
Text Adventure	The player takes part in a generated story and must choose the correct actions to continue.	• Story generation • Scene image generation • Feedback on story performance
Board Game	The player moves around a digital board game space. Randomized elements such as dice rolls and question types keep the game engaging.	• Question generation • AI agent
Card Game	The player draws cards from a deck and must answer them correctly. Hint cards may assist the player in answering the questions. The goal is to reach a target score before the opponent does.	• Question card generation • Hint generation • Custom feedback • Dynamic difficulty adjustment (DDA)

Table 1: Summary of different game genres, their descriptions, and AI applications.

3.3 About the Experiment

3.3.1 Focus Group and pre-experiment survey

The focus group consisted of seven volunteers around the ages of 15-17. Most of the volunteers were Dutch highschool studens; two male Dutch highschool students (VWO 6), four female highschool students (HAVO 4 & 5, VWO 4) and one female graduate school student (first-year). These students were asked to answer a preliminary survey to determine the genre of the prototype game, corresponding to one of the genres of the previous section. From this pre-experiment survey, half of the volunteers preferred a text adventure game over the other genres. One student who preferred this genre noted how they like custom character creators in games and so they would like to see this feature be added into the prototype. Whilst the feature was considered early on in development, it was ultimately omitted as the custom character could violate the rule of 'Authenticity'. A custom created character may feel out of place in a historical setting, as it does not originate from the represented time period. Another participant noted how they had an aversion to competitive style games, such as the board and card games. This is inline with the text adventure framework, where the user has to only interact with the story and not any opposing players.

3.3.2 Experimental Material: The final prototype

The final prototype text adventure game consists of website like interface. Users first begin by either uploading image files or typing a text prompt describing their desired adventure topic. This is done through the interaction with the story control panel located on the left of the interface. Users may deleted any unwanted images, or click the Start button to begin playing the game with the given context. Once the game starts, the interface displays both a story and question on the left of the interface, as well as a scene image depicting the story on the right. Users must then answer the question, which is done through widgets located at the bottom of the interface. The answering method alternates between multiple-choice and open-ended answers based on the previous response method. The game will then evaluate the users response against the correct answers using semantic matching, and provides immediate feedback through a styled overlay panel and score counter. Players may conclude their game session by pressing the *Beëindig spel* (End Game) button, which will trigger the generation of a personalized summary based on the users performance. The prototype text adventure app is shown in Figure 6.

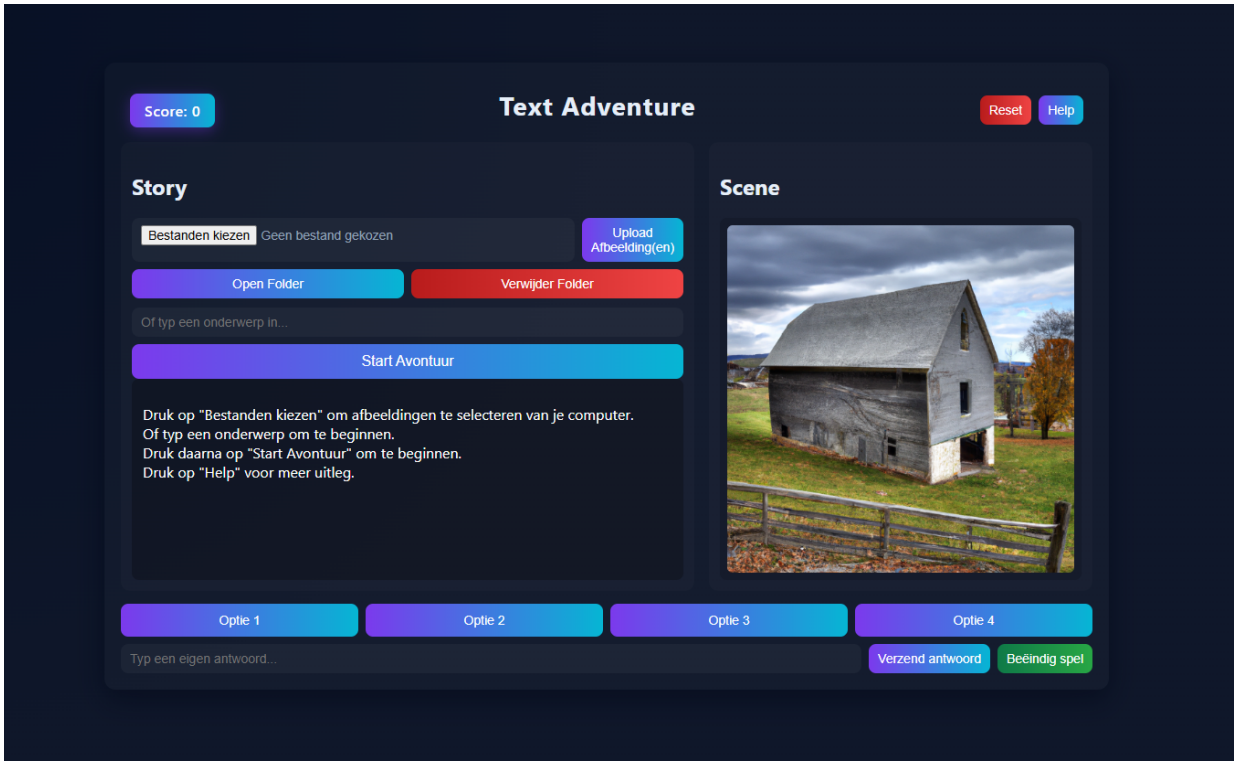

Figure 6: The text adventure Flask application. The left-hand side controls the story, whilst the image of the scene is shown on the right. Answering questions must be done through the widgets on the bottom.

The application was built using a Python-Flask backend in combination with a responsive HTML, CSS and Javascript frontend. The backend manages the user sessions with the Flask-Session, which stores game state variables (e.g., user score, question answers and used scenarios) server-side in the file system. The application integrates with OpenAI's API in order to generate custom stories and scene descriptions using the GPT-5 Nano model. It also made use of the DALL-E 2 model for procedural image generation. Uploaded images were stored server side in session-specific folders. The local images are Base64-encoded into the expected format, before being passed to the GPT-5 Nano vision endpoint as image content. The AI generated stories and scene images are then served to the frontend for display.

3.3.3 Experimental Setup

Each participant of the focus group was asked to attend a short experiment session where they would play around with the game and talk about their experience during play testing. Participants were told that a session was estimated to take 15 minutes, although this could be extended should they enjoy the game experience. The amount of preliminary information about the program, beyond its overall purpose, was kept to a minimum. This was done to facilitate a less biased final opinion from the participants, by enabling observation of a genuine blind experience with the game. Students were also asked to prepare and submit two distinct pictures to the AI: one image taken from a page of one of their textbooks in history, geography, social studies or biology, and any random image of their choosing. This was done to observe the experience with the AI content for its intended educational purpose, as well as its use in a setting purely focused around entertainment. Upon a participants decision to stop playing, they were asked some questions in person to detail their experience with the game and its AI components. They were additionally asked to answer an online questionnaire, with items relating to the Perceived Usefulness (PU) and Perceived Ease of Use (PEOU).

3.3.4 Technology Acceptance Model (TAM)

The PU and PEOU are concepts introduced in Fred Davis's Technology Acceptance Model [D+89]. These are considered latent variables, psychological attributes that are not directly observable and must therefore be inferred from observable instances. Davis suggested using a multi-item questionnaire for gathering participants opinions on the usefulness of a system, as well as its ease of use. The final perceived usefulness and ease of use questionnaires consisted of six different questions with a scale from 1-10, which captures the same relative response as the original 1-7 TAM scale [PC00] [LGCM08]. The list of questionnaire items corresponding to the PU and PEOU can be viewed in Appendix A.

4 Results & Analysis

The following chapter details the results and opinions of the focus group after they had interacted with the prototype Text Adventure game. This includes the answers given to the in person interview questions, as well as their answers to the questionnaire that concerned the Perceived Usefulness (PU) and Perceived Ease of Use (PEOU) of the game. One participant did not complete the survey and will therefore be omitted from the analysis.

4.1 Cronbach's Alpha

Fred Davis used Cronbach's alpha [Cro51] as a means to calculate the internal consistency of for the PU and PEOU items. This is done by examining the degree to which participants answers are consistent across questions. Higher alpha values correspond to a greater internal consistency, which justifies combining items into a single score for further analysis. In this study, the PU items produced an alpha value of 0.64 and the PEOU produced an alpha value of 0.88. Whilst an alpha value of 0.7 is often cited as the acceptable threshold, Taber (2018) noted [Tab18] that lower alpha values are common in small-sample contexts and do not indicate poor measurement quality.

4.2 Means from statistical analysis

Further statistical analysis was conducted on the participant PU and PEOU scores. The participant means were used instead of the individual question means as these provide observable estimates. The individual question means are only indicators of the latent variables which the Technology Acceptance model seeks to estimate. This resulted in the following average scores per user for both the PU and PEOU:

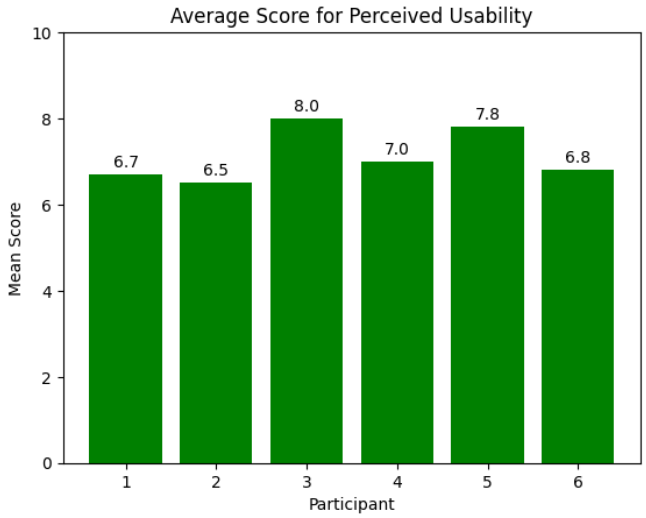

Figure 7: The Perceived Usefulness bar graph, with the average rating from the participants for the PU questions, assuming a 1-10 scale. The mean PU = 7.13

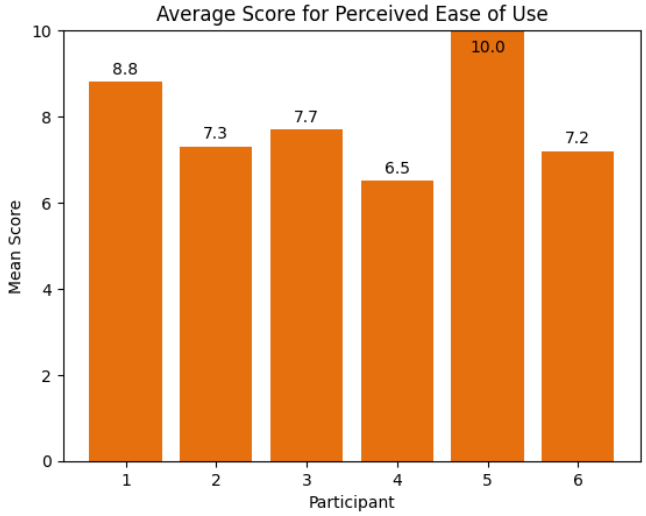

Figure 8: The Perceived Ease of Use bar graph, with the average rating from the participants for the PEOU questions, assuming a 1-10 scale. The mean PEOU = 7.92

4.3 One sample T-test against a neutral midpoint ($H_0 = 5.5$)

One sample T-tests were conducted for the means of the PU (7.13) and PEOU (7.92) to see if these were significantly different when compared to a neutral mid point. The neutral midpoint in this case represented the null hypothesis, the assumption that students hold neither positive nor negative attitude towards the text adventure game with AI generated content. The neutral

midpoint was thus defined as 5.5, the central value of the 1-10 scale. The standard deviations, based on the participants scores, were 0.619 (rounded up) for the PU and 1.270 for the PEOU. The final values of the t-statistic, based on the 6 participants sample, were 6.468 (rounded up) for the PU and 4.660 (rounded up) for the PEOU. Both results were statistical significant at the ($p < 0.05$) threshold, with p-values being calculated for PU ($p = 0.001$) and PEOU ($p = 0.006$). The positive t-statistics showed that the mean of both the PU and PEOU were significantly higher than the neutral midpoint, indicating that participants held a positive attitude towards the game's usefulness and ease of use.

4.4 Confidence Interval against a neutral midpoint ($H_0 = 5.5$)

95% Confidence Intervals were calculated for the mean PU and PEOU in order to visualize the differences between them and the neutral midpoint, when accounting for uncertainty. Confidence Intervals give a range of population mean values based on the observed data, which makes it possible to quantify the uncertainty due the focus group's small sample size of six people. By compering the intervals to the neutral midpoint, it becomes possible to infer whether the midpoint is a plausible population mean and whether the null hypothesis holds.

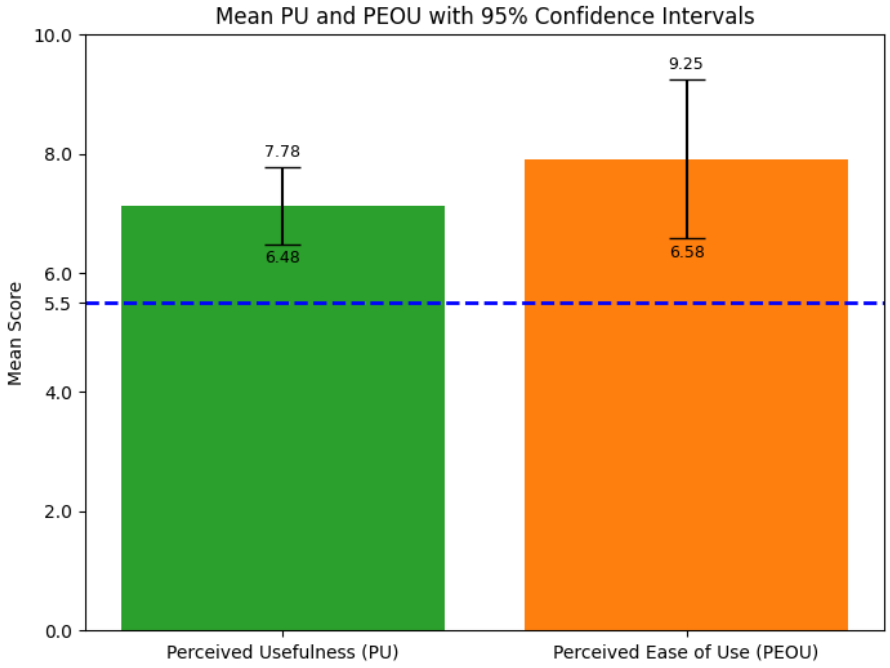

Figure 9: The Confidence Intervals for the mean PU (7.13) and PEOU (7.92). The neutral midpoint (5.5) falls below both Confidence Intervals of [6.48, 7.78] and [6.58, 9.25].

The Confidence Intervals for the PU and PEOU are shown in figure 9. Both intervals of PU [6.48, 7.78] and PEOU [6.58, 9.25] appear above the neutral midpoint of 5.5, which suggests that this point is not a plausible population mean. It shows that, even when considering the sample size uncertainty, the true mean must lie somewhere in the positive range rather than near the neutral midpoint. It also suggests that students would, in general, rate the game as more useful and easier

to use than a neutral stance would suggest. This further strengthens the evidence against the null hypothesis drawn from the t-test, and reinforces the believe that students attitudes towards the game were meaningfully above the neutral.

4.5 Results from the empirical analysis

The game was overall positively received by the participants, where many stated that the tool showed potential to be used as an engaging way to learn their school content. Many appreciated the text adventure elements, such as the score counter and the direct involved due to the narrative immersion of the user into the story. They also valued the detailed feedback received at the end of their gameplay session, noting that it addressed specific points of improvement rather than offering a generic summary.

At the same time however, participants also noted the various problems they had with the game. Some noted that the AI would often repeat itself by asking the same question multiple times when given minimal information or sometimes even at random. Moreover, all participants mentioned that the application was slow to produce stories and analyze results, a problem which was most likely rooted in the way that API calls were implemented. Finally, while the students were open to using the tool for learning school material, they were hesitant to use the tool for its entertainment value alone, suggesting that its primary strength lies in its educational value rather than its use as a recreational tool.

5 Conclusions & Further Research

This research has shown that there is genuine interest in educational serious games which makes use of an AI content generator. Various frameworks for educational games, as well as the integration of AI have been shown, with the final prototype game being based on the text adventure game framework. Generative AI could be used to generate stories, questions, scene imagery and general performance feedback for games which aids in creating a personalized experience for the user. The prototype text adventure game provided a practical example of a serious game, which was tested on a focus group for its Perceived Usefulness and Perceived Ease of Use. The statistical analysis showed that students on average rate the educational game above a neutral point, suggesting that the game was perceived to be useful and easy to engage with.

5.1 Limitations

Whilst this project provided meaningful discoveries about the perception towards eductional game, several limitations should be acknowledged. Future research into this topic should thus ideally aim to adress the following issues:

5.1.1 Small sample size

The research project focused on the results of six (orginally intended to be seven) Dutch secondary school students. The small heterogeneous group made it easy to base the design of the game on their preferences, whilst also facilitating a deeper analysis of their opinions on the game through the one-on-one interviews. However, while this resulted in a detailed empirical study, involving a borader target group could provide a clearer understanding of how to design a more generalisble educational game. On top of that, a larger sample size would be more suitable for an effect study, in which the impact of the game is examined by comparing the test scores between students who did and did not use it.

5.1.2 The flaws of the game AI

Whilst the overall reception to the final prototype was positive, all of the participants still voiced their problems they had with the game. Students noted how the game was slow to generate stories and images, which negatively impacted their overall experience. In addition, several stories included parts that closely resembled previously generated content and several questions would repeated multiple times. The perceived slowness may have been caused by communication and request-handling between the python application and the GPT-5 Nano model. The repetition of parts on the other hand may have resulted from the poor prompt structures. Future research should there for aim to avoid these issues so that participants' opinions focus more on the educational effectiveness of the prototype, rather than the technical shortcomings.

5.1.3 Testing of more genres

The focus of this project was on single game genre for the educational game, however future research should aim to create and evaluate games across a broader range of genres. This may include the frameworks proposed in this project, or entirely new framework designs for genres which were not covered here. Expanding the scope in this way would allow for the evaluation of a wider range of individual genres in addition to enabling comparative studies, examining the difference between genres a creating a clearer understanding of which genres are more suitable to educational game design with generative AI integration.

References

[AIA18] Abejide Ade-Ibijola and Kehinde Aruleba. Interactive learning of factual contents using a game-like quiz. In *Annual Conference of the Southern African Computer Lecturers' Association*, pages 285–298. Springer, 2018.

[Ami25] Narmin Amirova. Traditional vs. non-traditional teaching in secondary education: A comparative analysis. *Porta Universorum*, 1(3):101–109, 2025.

[Aun19] Khin Moe Myint Aung. *Comparison of levenshtein distance algorithm and needleman-wunsch distance algorithm for string matching*. PhD thesis, MERAL Portal, 2019.

[BD18] A Brennan and Mary Dempsey. The student voice: the students own views on smartphone usage and impact on their academic performance. In *INTED2018 Proceedings*, pages 7709–7718. IATED, 2018.

[BdAA+18] Thiago A Beppe, Ítalo Linhares de Araújo, Bruno Sabóia Aragão, Ismayle de Sousa Santos, Davi Ximenes, and Rossana M Castro Andrade. Greatest: A card game to motivate the software testing learning. In *Proceedings of the XXXII Brazilian Symposium on Software Engineering*, pages 298–307, 2018.

[Bul23] Duygu Bulut. The association between attention impairments and the internet and social media usage among adolescents and young adults with potential consequences: A review of literature. *Psychology*, 14:1310–1321, 2023.

[CK15] Edward Castronova and Isaac T Knowles. A model of climate policy using board game mechanics. *International Journal of Serious Games*, 2(3), 2015.

[Cro51] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

[D+89] Fred D Davis et al. Technology acceptance model: Tam. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, 205(219):5, 1989.

[DCVM+23] Luigi Di Caro, Laura Ventrice, Rachele Mignone, Stefano Locci, et al. Semantic doppelgängers: How llms replicate lexical knowledge. In *CEUR WORKSHOP PROCEEDINGS*, volume 3571, pages 12–18. CEUR-WS, 2023.

[DPBG23] Nixon Duarte, Yury Montoya Pérez, Adán Beltrán, and Matilde Bolaño García. Use of artificial intelligence in education: A systematic.. 2023.

[DWB+24] Jacob Doughty, Zipiao Wan, Anishka Bompelli, Jubahed Qayum, Taozhi Wang, Juran Zhang, Yujia Zheng, Aidan Doyle, Pragnya Sridhar, Arav Agarwal, et al. A comparative study of ai-generated (gpt-4) and human-crafted mcqs in programming education. In *Proceedings of the 26th Australasian Computing Education Conference*, pages 114–123, 2024.

[EAD+11] Henrik Engström, Erik Ambring, Carl-Johan Dahlin, Erik Sjöstrand, and Peter Håkansson. Making a game of the old testament balancing authenticity, education and entertainment. *IADIS International Journal on WWW/Internet*, 9(1):1–17, 2011.

[Fer25] Cátia Ferreira. Genre, bias, and narrative logic in ai dungeon: Generative ai as a game-based storytelling engine. *Hipertext. net*, (31):77–89, 2025.

[GAD02] Rosemary Garris, Robert Ahlers, and James E Driskell. Games, motivation, and learning: A research and practice model. *Simulation & gaming*, 33(4):441–467, 2002.

[GM17] William Goddard and Alexander Muscat. Towards genre as a game design research approach. In *Proceedings of DiGRA 2017 Conference*, 2017.

[IGK+25] Calvin Isley, Joshua Gilbert, Evangelos Kassos, Michaela Kocher, Allen Nie, Emma Brunskill, Ben Domingue, Jake Hofman, Joscha Legewie, Teddy Svoronos, et al. Assessing the quality of ai-generated exams: A large-scale field study. *arXiv preprint arXiv:2508.08314*, 2025.

[JOMJ18] Luke Conrad Jackson, Joanne O'Mara, Julianne Moss, and Alun C Jackson. A critical review of the effectiveness of narrative-driven digital educational games. *International Journal of Game-Based Learning (IJGBL)*, 8(4):32–49, 2018.

[Kat25] Menelaos N Katsantonis. From pandemic legacy to serious games: A systematic review of cooperative board games under the educational perspective. *European Journal of Education*, 60(1):e70048, 2025.

[KBCVdC20] David Kristan, Pedro Bessa, Ricardo Costa, and Carlos Vaz de Carvalho. Creating competitive opponents for serious games through dynamic difficulty adjustment. *Information*, 11(3):156, 2020.

[KGML+21] Isabelle Kniestedt, Marcello A Gómez Maureira, Iulia Lefter, Stephan Lukosch, and Frances M Brazier. Dive deeper: empirical analysis of game mechanics and perceived value in serious games. *Proceedings of the ACM on Human-Computer Interaction*, 5(CHI PLAY):1–25, 2021.

[Kho17] Solmaz Khodaeifaal. Student engagement: Enhancing students' appreciation for learning and their achievement in high schools. 2017.

[LCS24] Matthias Lehmann, Philipp B Cornelius, and Fabian J Sting. Ai meets the classroom: When does chatgpt harm learning. *arXiv preprint arXiv:2409.09047*, 2024.

[LGCM08] Luis M Lozano, Eduardo García-Cueto, and José Muñiz. Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2):73–79, 2008.

[MSCS24] Uday Mittal, Siva Sai, Vinay Chamola, and Devika Sangwan. A comprehensive review on generative ai for education. *Ieee Access*, 12:142733–142759, 2024.

[PC00] Carolyn C Preston and Andrew M Colman. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104(1):1–15, 2000.

[PCB92] Charles F Petranek, Susan Corey, and Rebecca Black. Three levels of learning in simulations: Participating, debriefing, and journal writing. *Simulation & gaming*, 23(2):174–185, 1992.

[PSL23] Muh Putra Pratama, Rigel Sampelolo, and Hans Lura. Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. *Klasikal: Journal of education, language teaching and science*, 5(2):350–357, 2023.

[Rom12] Margarida Romero. Learner engagement in the use of individual and collaborative serious games. In *Increasing student engagement and retention using immersive interfaces: Virtual worlds, gaming, and simulation*, pages 15–34. Emerald Group Publishing Limited, 2012.

[Sch08] Jesse Schell. *The Art of Game Design: A book of lenses*. CRC press, 2008.

[Tab18] Keith S Taber. The use of cronbach's alpha when developing and reporting research instruments in science education. *Research in science education*, 48(6):1273–1296, 2018.

[WVNVOVDS13] Pieter Wouters, Christof Van Nimwegen, Herre Van Oostendorp, and Erik D Van Der Spek. A meta-analysis of the cognitive and motivational effects of serious games. *Journal of educational psychology*, 105(2):249, 2013.

[WW24] Tyler B Wright and Jennifer L Weible. Current attitudes on digital interactive fiction and text adventure games within learning contexts: A systematic literature review. *Journal of Research in Science, Mathematics and Technology Education*, 7(1):23–51, 2024.

[Zyd05] Michael Zyda. From visual simulation to virtual reality to games. *Computer*, 38(9):25–32, 2005.

A Questionnaire items

Perceived Usefulness and Perceived Ease of Use questionnaire items were to be rated on a scale from 1-10, alongside an explanation for the given answer.

Perceived Usefulness:

1. The webapp would make it so I learn faster.
2. The webapp would cause me to better remember my learning objectives.
3. The webapp would increase my productivity of learning.
4. The webapp would make it so I learn extra details about the material.
5. The webapp would make it easier to learn.
6. I would find the webapp useful for learning.

Perceived Ease of Use:

1. Learning to use the webapp was easy for me.
2. It was easy to make the webapp do what I wanted.
3. The interaction with the webapp was clear and understandable.
4. I found the webapp flexible to use.
5. It would be easy for me to become skillful at using the webapp.
6. The system was easy to use.

Open Ended Questions

1. List the aspects of the game you were most positive about.
2. List the aspects of the game you were most negative about.
3. What was your opinion on the AI feedback summary?
4. What did you think of the AI in general? Would the interaction with the AI bother you or motivate you should you use the tool for educational purposes?
5. Would you be open to AI-assistance in general?
6. Would you use this tool for recreational purposes, such as in your free time?