Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 5 n.4 (2025)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Generative Artificial Intelligence and Videogame Analysis for Game-Based Learning: A Comparison between Automated Assessment and Expert Evaluation

Francesco Claudio Ugolini[1,*], Francesca Fallucchi[1], Michele Sardo[1] and Domenico Morreale[1]

[1] Università degli Studi Guglielmo Marconi; f.ugolini@unimarconi.it; f.fallucchi@unimarconi.it; m.sardo@unimarconi.it; d.morreale@unimarconi.it;
* Correspondence: f.ugolini@unimarconi.it;

Abstract: This study focuses specifically on Game-Based Learning (GBL), with the aim of examining how generative AI can assist teachers in designing educational activities that incorporate the pedagogical use of videogames not originally intended for educational purposes. Within a more general model which includes a game analysis from a pedagogical point of view and, subsequently, the design of learning activities, in this paper we compare the results of a human-based analysis whit an AI-based one, both referring to a specific framework involving three main dimensions: directivity, sociality and modifiability. We chose 8 games varying by notoriety, type and recency; we analyzed them according to our framework and then made them analyze by generative AI, trained with the same framework, and then compared them one with each other. We found that in most renowned games, the AI adopts a more technical or slang language. AI-generated answers also tend to go beyond the focus led by the prompt giving suggestion about instructional design. The study also helped to refine our theoretical framework by comparing its human interpretation with another one, AI-generated, namely about the dimension of modifiability.

Keywords: Game-based learning; generative AI; videogames

1. Introduction

This paper aims to examine how generative Artificial Intelligence (AI) can help teachers in designing Game-based learning activities, namely when involving videogames not originally intended for educational purposes.

AI in Education (AIED) opened promising scenarios, yet still methodologically and design-wise underexplored. We focus here on how AI can support teachers in activities without direct interaction with students, namely in instructional design, in this case including a game-based activity within a wider learning unit.

Our approach to Game-based learning actually involves integrating games with activities "next to the game" (Andreoletti & Tinterri, 2023), designed to be effective

[1] The whole research has been designed and realized by the four authors together; to the extent of this paper, Domenico Morreale is the author of par. 2.2, 3.3 and 3.4; Michele Sardo is the author of par. 3.1, 3.7 and 3.8. Francesca Fallucchi wrote par. 5. The remaining of the paper has been written by Francesco C. Ugolini.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

together in pursuing curricular objectives. We therefore proposed a theoretical framework (Ugolini & Morreale, 2023) with the aim of helping teachers to:

- choose and properly evaluate a game to be included in a learning activity from an educational and pedagogical perspective;
- design the learning activities built around one or more games.

The framework is based on the following premises:

- the instructional architectures (Clark, 2000) in the field of instructional design and the need of a continuous regulation of the openness of the problemic space (Calvani, 2011) by the teacher;
- the increasing introduction of the social dimension in many games, that now provide for interaction between several players, thus opening up cooperation/collaboration also in the educational field;
- the spread of games based on open worlds, characterized by a marked photorealism or the possibility of interacting in multiplayer mode with explorable environments and spaces;
- the renewed interest in constructionism (Papert, 1993), as we can found in games like Roblox or Minecraft (Benassi, 2021);
- the concept of interactivity, as it has been stated by Jens F. Jensen (1998; 2008), namely regarding conversational, consultational and registrational interactivity;
- the concept of scenario competence stated by Hanghøj (2013), involving three different aspects: problem scenarios, social scenarios and identity scenarios.

Moving from these premises, our framework includes the following three main dimensions:

- Directivity: the extent to which a game allows the player freedom in determining their actions, offering multiple alternative paths to problem-solving.
- Sociality: the extent to which a game fosters interaction and relationships, typically through online multiplayer features, up to the creation of genuine communities;
- Modifiability: the extent to which a game enables users to modify the environment or game components, either by creating entirely new content and objects or by creatively reworking existing ones.

According to our overall model, these dimensions aim to help teachers both in evaluating games and in designing learning activities next to them, within a logic of balance; for instance, a game with a high level of directivity should be balanced with activities with a lower level of directivity.

In this paper we aim to evaluate the reliability of generative AI in conducting this kind of analysis in order to support the teacher in designing game-based learning activities. We can identify three more specific research questions:

RQ1: Is the knowledge base adequate to make generative AI express a reliable evaluation of games according to our framework?

RQ2: Are there some biases due to the prompt formulation or to the training of AI?

RQ3: What information can we obtain about the very validity of the framework, by comparing the human analysis with the AI one?

2. Materials and Methods

2.1. Selection of the games

In this research we identified 8 games to be evaluated by both human researchers and generative AI, according to our framework.

These games have been chosen according to the following two major criteria:

- Notoriety and recency: we may assume that AI can rely on a wider knowledge base if the game is well known and consolidated (see RQ1);
- Game type: the framework has been built with in mind some paradigmatic games, typically based on large open worlds to explore (typical example of low directivity) or of a constructionist approach (typical example of high modifiability); we therefore are testing both the framework (see RQ3) and the AI asking this latter to evaluate games of different types.

The list of games is reported in Table 1.

Table 1. List of the game been evaluated

Game Title	Game Type	Recency and notoriety
Mission 2030	Interactive story	Little known and recent
Caesar IV	Strategy simulation game	Known and not recent
Brookhaven	Role-Playing game	Known and recent
Among us	Social roleplay	Known and recent
World of Warcraft	Massive Multiplayer Online Role-Playing Game	Known and not recent
TLOZ Tears of the Kingdom	Open World with construc-tionist elements	Well Known and recent
Assassin's Creed – Discovery Tour	Open World	Well Known (paradigmatic) and not recent
Minecraft	Sandbox	Well Known (paradigmatic) and fairly recent

2.2. AI settings

To the extent of this research, we adopted ChatGPT (OpenAI), model GPT-5 Thinking, via the ChatGPT Plus subscription (web interface).

Generative AI was trained to perform the analysis through a sequence of prompts. The first prompt provided the text of the scientific article outlining the framework (Ugolini & Morreale, 2023), which the AI then used to structure its evaluations for each game.

The following prompt was used for each title:

Analyse the videogame "[TITLE]" using the Ugolini–Morreale framework, which includes three dimensions: directivity, modifiability, and sociality. For each

dimension, indicate the level (high, medium, low) and justify your evaluation with concrete examples.

Both human and AI analysis were conducted in October 2025. Deeper analyses of *Assassin's Creed - Discovery Tour* and *Minecraft* in the light of our framework were also formerly done by the researchers (Sardo et al., 2024).

3. Results

We report here, game by game, a brief result of the comparison of the human analysis and the AI one, in relation to the three dimensions of our framework.

3.1. Mission 2030

Mission 2030 is a gamified interactive story aimed at making the goals of the 2030 Agenda known; it is based mainly on different quiz- or puzzle-based activities, present in each "mission" which is related to each 2030 Agenda goal.

Both human and AI agree that directivity is high, modifiability is low (there are no authoring functions nor deep personalization) and sociality is low (it's a single-player game).

We note that AI also suggests that sociality and modifiability can be improved by activities around the game.

3.2. Caesar IV

Caesar IV is a strategy simulation game; it simulates a social life in the ancient Rome. It has been chosen in this research as it represents the classical GBL approach, referring to the situated guided discovery architecture (Clark, 2000) as the learner is confronted with a model.

Directivity is medium, as the player is free to act within the constraints of the model. AI is more precise, stating that the objectives are set while there are different means to achieve them. For both human and AI, modifiability is medium as it is linked to the "editor" functions, and sociality is low as the multiplayer mode has been shut down in 2012.

3.3. Brookhaven

Brookhaven is a Roblox game and we can see it as a little collaborative metaverse, a digital life simulator and a role-playing game. Players can freely explore and socialize in this digital world interacting and collaborating with other players.

Both human and AI agree that directivity is low and sociality is high. There are some smooth differences regarding modifiability. Human states that it is "medium/high" while AI evaluates it simply as "medium". Both argues that the game allows an extensive customization of objects and spaces, but AI points out that Brookhaven does not allow the player to modify the game rules.

3.4. Among Us

Among Us is a multiplayer online video game in which participants play crewmates or impostors inside a spaceship or in other scenarios. Crewmates must complete a series of tasks and find out who, among them, is the impostor, calling meetings to discuss and vote on the expulsion of suspects, while impostors try to eliminate the crewmates and to sabotage the mission without being discovered.

Regarding directivity, the human stated it was "medium/high", as he emphasized the repetition of roles, turns and actions, though players can decide about timing and mode of execution; AI evaluated it as "medium", giving much importance to the "highly emerging" way to play, and to the possibility of the host to customize the game rules.

About sociality, AI claimed it was high as the social deduction multiplayer was the "heart of the game". Human nuanced this evaluation ("medium/high") as the social interactions by chat happen only during the emergency meetings.

We noticed major differences regarding modifiability: the human considered it "medium/low", as the player can customize his/her avatar and choosing between several scenarios; AI considered it low, arguing that such customizations do not amount to content construction/reprocessing.

3.5. World of Warcraft

We decide to include World of Warcraft in this research as it represents the typical Massively Multiplayer Online Role-Playing Game (MMORPG).

As a premise, we must note that AI took only 33 seconds to give its answer, and it adopts a technical slang. We may deduce that the knowledge base is very large, especially in social informal contexts.

Both human and AI found that directivity is medium as the goals are fixed but the way to achieve them may vary, and that sociality is high.

Modifiability is seen to be medium by the human due to its RPG features, that help the players in expressing themselves to some extent inside the game; AI argues that players cannot shape the world and this lowers the modifiability from a constructionist perspective, although some small customization were possible.

3.6. The Legend of Zelda – Tears of the Kingdom

We included here this game because it includes inside the same adventure game both exploration of a large open world and some constructionist elements thanks to the "ultrahand" power. Researchers therefore consider it with very low directivity and high modifiability. Sociality is low, as the game is single-player, but there are several communities around the game, by which player creations can be shared.

AI confirms this analysis regarding modifiability and sociality. It worth to be noted that AI also stated that any social practices are external to the game.

Although Tears of the Kingdom focuses on exploration, AI consider it has medium-low directivity because some segments of the game (typically linked to the main quest) had tighter restraints.

3.7. Assassin's Creed Odyssey – Discovery Tour

Assassin's Creed is a paradigmatic digital environment from a game-design point of view, as it is an expression of an exploratory play paradigm. The Discovery Tour aims to teach knowledge about classical Greece, in so limiting interaction inside the game[2].

Human and machine agree therefore about a medium-high directivity. The analysis of AI highlights how there are different modes/modes of "gaming": one related to the free-roam of the digital environment, which tends to decrease the directivity of the game and the other related to the pursuit of the tour of discovery, which increases the directivity of the gaming experience.

They also agree about a low modifiability, even if human highlights the parkour as an element for the players to express themselves while AI only points out the limited possibilities to customize characters or mounts.

Both human and AI agree about a low sociality as the game is not multiplayer. It has to be noted that, as it has been trained by the paper (Ugolini & Morreale, 2023) in its entirety, AI suggests some solutions around the game in order to improve sociality.

3.8. Minecraft – Education Edition

This game is the pedagogical declination of the famous videogame environment, a sandbox procedural environment based on construction, recalling the principles of constructionism.

Human and AI fundamentally agree about all three dimensions: directivity is medium-low (as it is generally low, but some possibilities are more directive), modifiability and sociality are high, as players can shape the game world and interact one with each other.

This is not surprising because this game and its constructionist principles has been one of the framework's basic inspirations.

4. Discussion

Regarding RQ1, overall, we may say that the synthetic evaluations are more or less accurate; the difference between human and AI are not substantial. Maybe the case of Among Us shows little differences in all three dimensions.

We can note that the analysis of the most renowned games, which is based on a larger knowledge base, adopt a more informal language (often slang or technical language). The answer on its entirety, may present some inaccuracy, but it still can give some interesting information based on our framework to a human designer.

About RQ2, prompt and training show two major limits; as we trained the AI with a scientific paper, and not with a focused document. This has been done on purpose in order to explore the possibilities of the generative AI, making the analyses interesting in the eyes of the researchers' but introducing at the same time more inaccuracies in the additional unsolicited information. The model should therefore be refined in the final version in order to make the analysis more focused. Regarding the

[2] We already studied the specificity of this game in (Sardo et al., 2024; Sardo & Thibault, 2024).

prompt, we noticed that AI adopt a very rigid scale, not including the "absent" evaluation as well (namely regarding sociality in single players games). Prompt can be refined being more precise about the scale.

Regarding RQ3, we probably have the most interesting information.

In general, sociality has been well interpreted by AI.

Directivity is often well interpreted by AI, with some little differences. AI evaluated twice a medium directivity when goals were fixed and means to achieve them were free. This was not initially described by the model, but can be an interesting criterion to refine this concept. Both in "Among Us" and in "Tears of the Kingdom", human and AI diverged in assessing the overall directivity of the game, giving a different importance to existing elements: for instance, "Tears of the Kingdom" is renowned for its open world which fosters exploration, so that directivity should be considered as very low, but actually AI is not totally wrong in pointing out some more directive segments related namely to the main quest. The balancing between rigid routines and space of action in "Among Us" has also been differently interpreted, with the human highlighting the repetitiveness of roles, turns and actions and AI focusing more on the emerging way to play.

The most interesting information were about modifiability that probably needs further precision. Actually, this dimension was conceived thinking of constructionism and, therefore, having games like Minecraft in mind, that have been studied by scientific literature (Benassi, 2021). So the technical features of a game to allow authoring overlapped a more deep meaning regarding the player self-expression, which can be linked to the identity scenario (Hanghøj, 2013), emerging from the human answers. Those of AI, on the contrary, see customization as content creation, and often adopt the phrase "shaping the world"; in some case, they refer to the possibility to modify the very rules of the game. Probably the model should be improved giving to these dimension a more deep meaning, changing its name in case.

5. Conclusions

Our research raises relevant epistemological questions about the nature of the knowledge produced by AI: whereas the human expert integrates embodied, context-sensitive experience within the ludic-educational environment, AI constructs its analyses from a wide array of heterogeneous sources, filtered and aggregated through algorithmic logic. Consequently, there is a fundamental difference between the teacher's tacit knowledge— based on observation, direct interaction, and pedagogical reflection—and the generative AI's knowledge, trained on vast datasets that primarily reflect representations and discourses prevalent in the public sphere and within relevant digital communities.

From this perspective, the study also explores how AI can help make game assessment more accessible. It is neither realistic nor desirable to expect videogame-specific expertise to become a stable component of every teacher's skill set. In this sense, AI has the potential to support even those approaching the pedagogical use of videogames for the first time, offering a replicable and adaptable tool for different contexts.

Ultimately, the teacher retains the responsibility to contextualize the game within their specific learning environment: an algorithmic evaluation— despite being rapid and scalable—may overlook the complexity of relational dynamics and contextual

variables that emerge only through direct experience and in-depth disciplinary knowledge. Moreover, the tool must be trained and used through prompts designed to minimize semantic misinterpretations that could undermine the validity of results.

References

Andreoletti, M. & Tinterri, A. (2023). *Apprendere con i giochi. Esperienze di progettazione ludica*. Carocci.

Ugolini, F.C. & Morreale, D. (2023). Ambienti videoludici per l'apprendimento: un framework per la progettazione di attività didattiche. *I problemi della pedagogia*. a. LXIX (2), 359-386.

Clark, R. (2000). Four Architectures of Instruction. *Performance Improvement*. 39 (10), 31-38.

Calvani, A. (2011). *Principi dell'istruzione e strategie per insegnare. Criteri per una didattica efficace*. Carocci.

Papert, S. (1993). *The Children's Machine. Rethinking School in the Age of the Computer*. Basic Books.

Benassi, A. (Ed.) (2021). *A scuola con Minecraft. Progettare un mondo a cubetti*. Giunti Scuola.

Jensen, J.F. (1998). Interactivity: Tracking a New Concept in Media and Communication Studies. *NORDICOM Review*, 19 (1).

Jensen, J.F. (2008). The Concept of Interactivity – Revisited: Four New Typologies for a New Media Landscape'. *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video UXTV* '08, ACM.

Hanghøj, T. (2013). Game-based teaching: Practices, roles, and pedagogies. In *New Pedagogical Approaches in Game-enhanced Learning: Curriculum integration* (pp. 81-101). IGI Global.

Sardo, M., Morreale, D., & Ugolini, F. C. (2024). Directivity and Modifiability in video-game environments for learning. Two paradigmatic cases compared. *Journal of Inclusive Methodology and Technology in Learning and Teaching*, 4(3). https://www.inclusiveteaching.it/index.php/inclusiveteaching/article/view/236

Sardo, M., & Thibault, M. (2024). An educational leap of faith: Exploring the connections between instructional architectures and digital game spaces. In P. DSouza, M. Buji, N. Xi, & J. Hamari (Eds.), *Proceedings of the 8th International GamiFIN Conference* (pp. 68-78). (Ceur Workshop proceedings; Vol. 3669). CEUR Workshop Proceedings. https://ceur-ws.org/Vol-3669/paper7.pdf