Università degli Studi di Genova

Università di Genova | Digital Humanities PhD

Dottorato in Digital Humanities

Curriculum: Linguistica, linguistica applicata, onomastica

XXXVIII ciclo

Serious Games to facilitate personalized learning enhanced by technologies, to promote the development of problem-solving skills within shared training paths in the UNITA Alliance Digital Learning Environment

Tutor

Marina Marchisio Conte
Sergio Rabellino

Candidata

Valeria Fradiante

A.A. 2024/25

SETTORE SCIENTIFICO-DISCIPLINARE DI AFFERENZA:

MATH-01/B, INFO-01/A Informatica

“*That's what games are, in the end. Teachers. Fun is just another word for learning*”

(Koster, 2013)

Abstract

The use of (digital) game-based learning, gamification, and serious games in education is increasingly widespread, providing innovative tools to engage students and enhance their learning experience. These approaches can be effectively supported by digital technologies, which enable more interactive, motivating, learning environments. This thesis presents a set of research activities employing digital, game-based methods to teach and learn Mathematics in a digital learning environment. Mathematics is the subject in which students at all levels of education experience the most difficulty. Previous studies have shown that these difficulties may be caused by a lack of mathematical attitude, which can be influenced by factors such as motivation, interest and anxiety about Mathematics. Regarding the use of game-based learning and serious games for teaching and learning Mathematics, the state of the art reveals a limited number of studies focusing on higher educational levels. Furthermore, the literature highlights the limited availability of validated questionnaires for evaluating such tools. In this study, two serious games were developed: "The murderer at the High-Tech Institute", focusing on the functions of a single real variable, and "Don't open that spacecraft", addressing more specific topics related to probability and statistics. These activities were experimented with approximately 360 university students enrolled in non-mathematics degree programmes and 120 upper secondary school students in their penultimate or final year and their teachers, during the 2023-24 and 2024-25 academic years. The research methodology involved collecting and analysing multiple types of data using both qualitative (pre- and post-questionnaires and post-game interviews) and quantitative (game session scores and exams results) approaches. The results show that the digital game-based activities contributed positively to learning outcomes and developed transversal skills for university students. Consistent findings were observed across the different student samples considered, despite differences emerging between degree programmes. Students did not find the use of these technologies difficult and rated usability as the most positive aspect of the experience. However, less satisfactory results emerged with respect to engagement, game appeal and perceived usefulness, indicating room for improvement in future research. Secondary school teachers gave the proposed methodologies a highly positive evaluation, highlighting their positive impact on learning and perceiving the tools as highly usable. However, acceptance of technology did not reach equally high levels. Notable results were also identified in terms of a reduction in math anxiety, suggesting that students experienced lower levels of math anxiety when the discipline is taught in a game-based setting rather than in a traditional instructional context. Furthermore, the study highlighted the potential benefits of emerging Generative Artificial Intelligence tools for teaching and learning in game-based environments. A model for designing digital game-based activities in a digital learning environment was developed based on the research conducted, and it is intended to be shared as an open educational resource. The research also addresses needs that have emerged significantly within the UNITA European Alliance, where the use of digital learning environments plays a central role and the sharing of engaging online learning pathways and open educational resources represents one of the strategic priorities. This model includes a reliable questionnaire designed to measure the effectiveness and impact of digital game-based tools in terms of technical and gaming aspects, as well as discipline-specific elements. The topics emerging from this research remain open to further development, particularly given the ongoing expansion and rapid evolution of digital technologies.

Preface

From the beginning of the PhD programme, the candidate's training has been embedded in a research project focused on developing innovative teaching methods for learning Mathematics. The candidate brought prior experience from her academic background and involvement in Italian and European research projects. The PhD programme was structured to provide the candidate with the necessary mathematical, methodological and digital knowledge to carry out the research project effectively.

The following publications present part of the research work:

1. Barana, A., Conte, A., Fradiante, V., Marchisio Conte, M., & Rabellino, R. (2025). Serious game for higher mathematics education: Evaluation of the learning tool. In *Proceedings of the 2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 2367–2376). IEEE. https://doi.org/10.1109/COMPSAC65507.2025.00333

2. Fissore, C., Floris, F., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (2024). Involving teachers in gamified learning activities using generative artificial intelligence tools. In A. Schönbohm, F. Bellotti, A. Bucchiarone, F. de Rosa, M. Ninaus, A. Wang, V. Wanick, & P. Dondio (Eds.), *Games and Learning Alliance: 13th International Conference, GALA 2024, Berlin, Germany, November 20–22, 2024, Proceedings* (pp. 36–46). Springer. https://dx.doi.org/10.1007/978-3-031-78269-5_4

3. Fissore, C., Floris, F., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (2024). From theory to training: Exploring teachers' attitudes towards artificial intelligence in education. In *Proceedings of the 16th International Conference on Computer Supported Education* (Vol. 2, pp. 118–127). SCITEPRESS–Science and Technology Publications. https://dx.doi.org/10.5220/0012734700003693

4. Floris, F., Fradiante, V., Marchisio Conte, M., & Rabellino, S. (2024). Strategie di gamification con Moodle per lo sviluppo di competenze di problem solving e per uno sviluppo sostenibile. *Rivista Bricks*, *2*, 25–34.

5. Fissore, C., Fradiante, V., Marchisio, M., & Pardini, C. (2023). Teachers' strategies and difficulties in designing gamification activities. *IADIS International Journal*, *21*(2), 86–100.

6. Floris, F., Fradiante, V., Marchisio Conte, M., & Rabellino, S. (2023). Design gamification strategies in a digital learning environment: The impact on students. In P. Dondio, M. Rocha, A. Brennan, A. Schönbohm, F. de Rosa, A. Koskinen, & F. Bellotti (Eds.), *Games and Learning Alliance: 12th International Conference, GALA 2023, Dublin, Ireland, November 29–December 1, 2023, Proceedings* (pp. 464–469). Springer. https://doi.org/10.1007/978-3-031-49065-1

7. Fissore, C., Fradiante, V., Marchisio, M., & Pardini, C. (2023). Design didactic activities using gamification: The perspective of teachers. In M. B. Nunes, P. Isaías, T. Issa, & T. Issa (Eds.), *Proceedings of the 17th E-Learning and Digital Learning Conference 2023 (EDL2023)* (pp. 11–18). IADIS Press. https://doi.org/10.33965/EL_STE_2023 *(Outstanding Paper Award).*

8. Fissore, C., Floris, F., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (in press). *Applications of Artificial Intelligence and gamification for teaching and learning purposes*.

9. Barana, A., Fissore, C., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (2026). Serious game per supportare l'apprendimento della matematica a livello universitario. In C. Bassi, D. Brunetto, M. Conti, M. G. Fiorentino, & A. Miranda (Eds.), *Book of Abstracts – Workshop 2025 DigiMath: Il ruolo delle tecnologie digitali a supporto della didattica della matematica: esperienze di buone pratiche a livello universitario (*pp. 8–10). ISBN 978-88-6493-1265. https://hdl.handle.net/11311/1306287

10. Fissore, F., Fradiante, V., Marchisio Conte, M., Pardini, C., & Sacchet, M. (2026). *Moodle come ecosistema per l'integrazione di intelligenza artificiale e gamification: progettazione e sperimentazione nel modulo formativo della rete Problem Posing and Solving*. In *Atti del MoodleMoot Italia 2025*, pp. 133-141. ISBN 979-12-985195-1-0

Sommario

1. Introduction

The introductory chapter provides an overview of the structure and contents of the thesis, clearly defining the research objectives and motivations behind the study. It also contextualises the research, providing the necessary background information to frame the work and support the subsequent chapters.

1.1 Research Background

Students often perceive Mathematics as a difficult and unappealing subject. This perception can shape their overall attitude towards the discipline, negatively influencing their confidence, motivation and long-term relationship with it. This makes it a challenging subject to teach and to learn, especially for students with a weaker mathematical background. Some studies report on the difficulties university students, in particular, non-mathematicians, face in Mathematics, such as in mathematical modelling (Marchisio et al., 2021; Fahlgren et al., 2024). The adoption of innovative approaches and methodologies, including problem-solving alongside the use of modern technologies, can significantly enhance Mathematics learning for non-mathematician university students. They contribute to the development of more effective study strategies, and they foster more positive attitudes and beliefs towards the discipline (Marchisio et al., 2020). Recent years have seen several international reports on Mathematics that are not encouraging. Firstly, the "PISA (Programme for International Student Assessment) 2022 Technical Report" shows a decline in average scores in Mathematics for OECD (Organisation for Economic Co-operation and Development) countries. This report highlights the difficulties faced by secondary school students and it provides a useful framework for understanding the causes of Mathematics difficulties that students may encounter later at university. The drop in the OECD average for Mathematics performance between the 2018 and 2022 PISA reports was 17 points and Italy's results in Mathematics are in line with the average for OECD countries. This decline is probably due to the effects of the covid-19 pandemic, which had a significant impact on the education system. Furthermore, Italian students are more affected by math anxiety than their peers in the rest of Europe. Poor performance in Mathematics can increase anxiety, causing students to avoid the subject and preventing them from developing and improving their skills, thus perpetuating the negative cycle (Chen et al., 2023). Secondly, the Survey of Adult Skills 2023 assessed the literacy, numeracy and adaptive problem-solving skills of adults aged 16-65. Italy scored below the OECD average, particularly in numeracy, where 35% of adults scored at or below Level 1. At Level 1 of 5, they can do basic arithmetic with whole numbers. Similarly, in adaptive problem solving, 46% of adults performed at Level 1 or below. Adults at Level 1 of 5 can solve simple problems with few variables and little irrelevant information.

One of the most significant challenges teachers face today is effectively engaging students, motivating them to learn and encouraging them to play an active role in their learning process (Pais et al., 2023). This is particularly difficult with subjects like Mathematics, which are often poorly appreciated by students. The challenge is even greater with older students who have spent years being exposed to traditional learning methods and may now be resistant to new, more active and different approaches. Moreover, education system is currently facing numerous challenges: traditional teaching methods often fail to develop mathematical skills effectively, highlighting the need for new approaches (Crnković et al., 2022; Hussein et al., 2022). Moreover, the shift to distance learning was accelerated by the pandemic (St Omer et al., 2025; Fante et al., 2024). This led to technology being adopted more widely as an educational tool, and the variety of instructional methods used during this transition allowed teachers and students to experiment with new approaches (Chen et al., 2023). One of the major challenges in Mathematics education is overcoming students' lack of motivation and interest in the subject (Pais et al., 2023). Involving students and using resources that make Mathematics lessons more appealing and engaging are essential for fostering meaningful learning experiences. The use of game-based learning (GBL) and gamification in higher and secondary education is a response to several educational challenges, such as the need to increase student engagement, improve academic success rates, and promote more sustained and meaningful learning. GBL is considered an educational approach that utilizes games (digital or otherwise) as the primary tool to facilitate the learning process (Plass et al., 2016; Perrotta et al., 2013). Gamification, while sharing some similarities with GBL, refers to the use of the typical mechanisms of the game, such as the challenge, the use of points, levels and prizes, in a context that is essentially not a game

(Deterding et al., 2011). In recent years, numerous experiences have been developed and implemented in higher education to enhance learning using game mechanics (Duggal et al., 2021). Although research on this topic is still developing, in Mathematics there is some evidence to suggest that integrating a game-based approach into traditional teaching methods can help make learning a more motivating, interactive and engaging experience (Zabala-Vargas et al., 2019).
In the current context of widespread technology use in education following the Covid-19 pandemic, this thesis aims to develop a digital game-based model to facilitate and enhance teaching and learning. This model is intended to encourage the creation of game-based open educational resources integrated within Digital Learning Environments (DLEs).
In recent years, tools such as serious games have become widely adopted. As shown in Chapter 4, these games are not only designed for entertainment; they are also explicitly intended to foster learning and support the development of specific skills (Chen & Michael, 2005). Two serious games were developed and experimented to investigate their potential to enhance Mathematics learning. These games are integrated into the learning tools interoperability framework within the Moodle learning management system. Unlike traditional games, serious games are designed not only for entertainment, but also to achieve specific learning goals and develop targeted skills (Chen & Michael, 2006). The two serious games primarily aim to facilitate and enhance students' approach to Mathematics by stimulating their problem-solving and critical thinking skills. They offer an engaging and interactive approach to reinforcing learning and applying concepts learnt in class. The first game is called "The murderer at the High-Tech Institute", further indicated as the "Murderer game". Players must identify the murderer of Teresa, a scientist at the High-Tech Institute, using the mathematical clues provided, which refer to the path taken by the murderer. The second game is called "Don't open that spacecraft", further indicated as the "Alien game". This game covers probability and statistics topics, including descriptive and inferential aspects. Its main goal is to read and correctly interpret real-world data. Players are asked to help Allan-Meta, an alien from Metaland, to correctly interpret data and information about our planet. The games were proposed to upper secondary school and university students from different degree programmes. Ethical approval for the collection of this dataset was granted by the Bioethics Committee of the University of Turin. The tools are accessible to students anytime, anywhere via any Moodle platform, and are intended as learning support tools. The editor used to develop the game transmits game session data to Moodle, enabling teachers to monitor students' progress.

1.2 Motivation and Objectives of the Research

This research project is motivated by the desire to understand how it is possible to facilitate students' approach to Mathematics, making it more engaging and motivating. The research project aims to explore digital game-based strategies and design solutions by encouraging active participation and fostering a positive attitude towards the subject, to improve the learning experience and reduce math anxiety (Weir, 2023).

The first objectives of the research are to:

- stimulate the development of disciplinary and transversal skills.
- promote personalised and adaptive learning that can effectively respond to students' needs.
- foster inclusive learning.
- encourage interaction, motivation and engagement among students through experiential, hands-on learning rather than purely theoretical study without real-world application, thus facilitating their engagement with the discipline.
- simplify and enhance teaching activities by providing teachers with effective tools to monitor students' learning processes.

One of the central challenges of the research has been to identify the most appropriate tools with which to evaluate the effectiveness and impact of serious games. To assess whether and how serious games achieve their intended outcomes, reliable and valid instruments are needed to capture the quantitative and qualitative dimensions of learning and engagement.

In the current post-pandemic context of widespread technology use, the final objective of the research project is to develop a digital game-based learning model that can be used to facilitate and enhance learning. The model was iteratively defined and developed based on the experience gained from conducting this research project in the field of Mathematics, but it is designed to be adopted for use in other disciplines. The research also addresses needs that have emerged significantly within the UNITA European Alliance, where the use of digital learning environments plays a central role and the sharing of engaging online learning pathways and open educational resources represents one of the strategic priorities.

1.3 Contents and Structure of the Work

The introductory chapter outlines the contents of the thesis, research objectives, motivations and context. The second chapter is dedicated to the state of the art. It first examines the institutional references, followed by the theoretical framework of the research and a literature review of the main research topics. This review was conducted primarily during the first year of the PhD, with further development in the second and third years. The main methodologies adopted for developing and evaluating (digital) game-based resources within the theoretical framework are examined. Additionally, issues related to students' attitudes and perceptions of Mathematics, such as math anxiety (Weir, 2023), are explored. Chapter 3 defines the context, objectives and research questions. Chapter 4 describes the design and development process of the two serious games related to Mathematics, which were integrated within a DLE. These games were experimented with students from different degree programmes at university level and with upper secondary school students and their teachers. Chapter 5 shows the details of the didactical experimentation, together with the methods and tools used during the experimental phase. Chapter 6 presents the results of the analysis of the data collected during the experimentations of the two serious games, in both qualitative and quantitative terms. The analysis of the results contributes to define a digital game-based learning model that is presented in Chapter 7. The thesis concludes with Chapters 8 and 9, which summarise the results and implications of the research project. It also considers the innovative nature of the research and its potential impact, while exploring possible future developments.

2. State of the Art

The state-of-the-art chapter outlines the theoretical framework that supports the research and provides an extensive literature review related to the study's central topics, which include concepts related to game-based learning but also attitudes towards mathematics. The theoretical framework examines also Italian national as well as European guidelines to situate the research within broader educational policies, particularly in relation to wider international contexts, such as the UNITA European Alliance. Within this framework, a dedicated section focuses on the theoretical and practical tools used to develop a model for designing digital game-based activities. During the development of the theoretical framework, a systematic literature review was conducted to provide a solid, up-to-date foundation for the research. This review specifically addressed game-based learning in Mathematics education, enabling a clear picture of the current state of the art. The resulting theoretical framework formed the basis for the practical development of the serious games, as well as for the design and implementation of the experimental phase, drawing on existing evidence from the literature. Particular attention was also devoted to evaluating serious games using reliable and valid assessment instruments, informing the evaluation approach adopted for those developed in this study.

2.1 Institutional Framework

Children learn through games from an early age. This natural and engaging method of learning can be effectively incorporated into the formal education system (Crnković et al., 2022). Through games, children observe, ask questions, test ideas, justify their choices and interpret outcomes in an informal and natural context (Casasso et al., 2022). The importance of games in the educational process from kindergarten onwards is also emphasised in the Italian National Guidelines for Kindergarten and the First Cycle of Education (Ministero dell'Istruzione e del Merito, 2025). Specifically, the document states that the Italian school system recognises the fundamental importance of game in its heuristic value as a tool for learning and personal development. This is achieved through the adoption of educational and teaching methodologies that enhance its educational, emotional, relational, expressive, aesthetic and other purposes. It also emphasises how game can provide "augmented spaces" for learning, discovering new realities and broadening personal experiences. While games in kindergarten act a specific role and focus primarily on exploratory and relational learning, national guidelines section for primary and lower secondary schools generally limit the context of games to physical education and art disciplines. Similarly, in the 2010 National Guidelines for High Schools (Ministero dell'Istruzione, dell'Università e della Ricerca, 2010a), games are emphasised as a means of developing physical skills, promoting an active lifestyle, and fostering the values of collaboration and fair play, as well as encouraging social interaction between students. Instead, the 2010 Guidelines for Technical Institutes (corresponding to the National Guidelines for high schools, but specifically for technical institutes in Italy) recommend adopting teaching methodologies that consistently integrate technology to design learning activities that focus on experience, such as educational games. These methods enable students to learn by validating their knowledge in "interactive learning environments that simulate real-life contexts, utilising either discovery learning or programmed instruction" (Ministero dell'Istruzione, dell'Università e della Ricerca, 2010b). This approach connects with students through familiar methods, providing a safe environment where they can experiment, make mistakes, and learn, much like in a game. The Guidelines for Professional Institutes also recognise the importance of games in the learning process. This is highlighted, for example, in competence 8, which refers to the ability to design educational, playful resources that are appropriate for different contexts and needs (Ministero dell'Istruzione, dell'Università e della Ricerca, 2018). It is a way to connect with students by using familiar approaches, such as play, to encourage hands-on learning. The greater reference to games in these two types of Italian upper secondary schools is probably linked to their more practice-oriented nature compared to Italian high schools.

Looking at the European reference frameworks, game is increasingly recognised as an important tool for developing transversal skills and encouraging student engagement. The use of digital games for teaching purposes is mentioned in the European Framework for the Digital Competence of Educators (DigCompEdu) (Redecker, 2017), a European framework that provides a common reference for understanding and assessing teachers' digital competence. Developed by the European Commission, it is intended to help teachers and trainers effectively integrate digital technologies into their professional practice. According to DigCompEdu, technologies such as games can be used to create educational resources that facilitate learning and provide students with engaging and motivating experiences. Furthermore, digital games offer a dynamic and interactive way of monitoring learning progress from a formative assessment perspective. They also support personalised learning paths, enabling students to progress at their own rhythm, providing different levels of difficulty and revisit activities as required.

The Digital Competence Framework for Citizens (DigComp) 2.2 (Vuorikari, 2022) is the European framework that defines and promotes citizens' digital competences. It outlines the key areas of the digital competence and defines the necessary skills for both schools, social or work contexts. It is also intended as a reference for the development of actions or initiatives aimed at fostering this competence in students and the broader population. Firstly, it is important to define what is meant by "digital competence". According to the Council Recommendation of 2018 (European Commission, 2019), it is defined as the "the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society". Digital competence encompasses more than just using digital tools; it also involves knowledge and awareness of areas such as information and data literacy, communication, digital safety, and intellectual property issues. The latest released version of the document, DigComp 3.0 (Cosgrove & Cachia, 2025), updates the reference framework for citizens by addressing, among other aspects, the use and comprehension of Artificial Intelligence and the fight against disinformation, integrating these emerging challenges into the competence areas already identified in the previous framework and its earlier versions. It highlights the importance of the critical use of information and the development of AI-related knowledge and skills in a transversal way across all competence areas. Within both the DigComp 2.2 and DigComp 3.0 digital games are mentioned as a type of digital content comparable to any other software or digital resources. Furthermore, digital games are referenced specifically as practical case studies that illustrate how digital competencies can be applied in real-world work contexts. Moreover, young people are often exposed to gamification mechanisms in their daily lives, for example through various communication services, which are sometimes used to influence user behaviour. Integrating this approach positively into learning is an important way to educate students and raise their awareness, helping them to recognise and evaluate such techniques in a critical way (Vuorikari, 2022). Considering the European reference frameworks DigComp 2.2 and DigCompEdu, technologies can be used to develop innovative educational resources, such as games, to facilitate the learning and teaching process.

In the digital age, the development of problem-solving and critical thinking skills is fundamental to all areas of the curriculum, from kindergarten to university. All Italian institutional theoretical frameworks for upper secondary schools recognise the importance of problem-solving as a fundamental competence, both transversal and disciplinary. The 2010 Guidelines for Technical Institutes recommend adopting a problem-solving approach to teaching, given that the world of work and technology are constantly evolving and require the ability to solve unexpected problems (Ministero dell'Istruzione, dell'Università e della Ricerca, 2010b). The ability to identify appropriate strategies for solving problems is not only a skill demanded in Mathematics but a transversal competence applicable across all disciplines. Similarly, the Guidelines for Professional Institutes (Ministero dell'Istruzione, dell'Università e della Ricerca, 2018) promote a problem-solving approach with increasing levels of autonomy and responsibility. It states that critical thinking and problem-solving are foundational to all key competences, providing an important reference for schools when developing teaching proposals. Even in the 2010 National Guidelines for High Schools (Ministero dell'Istruzione, dell'Università e della Ricerca, 2010a), problem-solving plays a central role, both as a transversal competence - emphasising the importance of solving problems and acquiring critical awareness of one's own actions - and from a disciplinary perspective, particularly in Mathematics. Specifically, students at the end of the path are expected to be able to identify and solve problems, as well as find possible solutions.

Another important point to note is that the concept of digital competence is closely associated with problem solving and critical thinking. The cases studies in DigComp 2.2 clearly demonstrate how digital competencies can be applied in professional contexts, particularly in Competence Area 5, which focuses on problem solving. The examples demonstrate the strong connection between games, digital competencies and problem solving, showing how game-based activities can provide a powerful context for developing and applying digital skills, encouraging critical and creative thinking to solve problems. In this sense, the game provides students with an opportunity to solve problems that require the use of a variety of skills. This encourages them to make observations, formulate questions and possible solutions, and ensures their involvement, enthusiasm, motivation, competitiveness and respect for the rules.

2.2 The Digital Learning Environment

The term Digital Learning Environment (DLE) refers to an ecosystem designed to support teaching, learning and competence development in classroom-based, online or blended settings (Barana & Marchisio, 2021). A DLE incorporates a human element, a technological element, and the interactions between the two. The human element consists of one or more learning communities, whose members can include teachers or tutors, students or learners, and their peers, as well as the administrators who manage the online environment. The technological element includes all kinds of technological devices and how they are used together. It includes a Learning Management System (LMS), as well as software, tools, and integrations designed for specific learning purposes, such as web conferencing and assessment tools and sector-specific software. An LMS is an online platform used to create and distribute course content. It delivers static or interactive activities and resources, which can be used in either synchronous or asynchronous modes. LMS features enable teachers to track student submissions and activities, in particular allowing students to access resources and teachers to monitor their activities on the platform (Marchisio et al., 2022). The interrelations between the two elements include the interactions and learning processes that occur within the community and through the use of technology, such as communication among members of the learning community and human-technology interactions, as well as the pedagogies and methodologies used to design the learning environment (Barana & Marchisio, 2021).

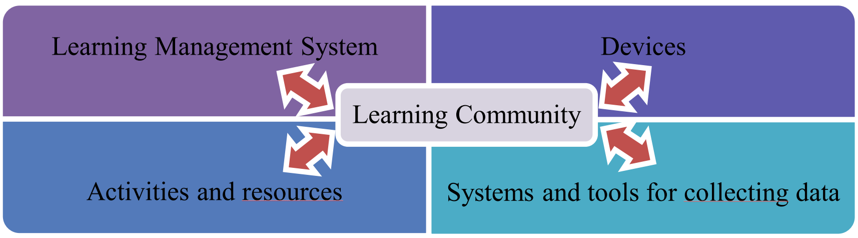

Figure 1 - The components of a DLE.

Taking inspiration from the description and graphical representation in Barana and Marchisio's (2021) study, Figure 1 illustrates the components of a DLE. At the centre of Figure 1 there is the learning community, representing a human-cantered approach to learning. The technological component surrounds the learning community representing the physical environment where it lives. The double arrows linking the community and the technologies represent the reciprocal relationships between the two elements: the learning processes within the community and the methodologies used to design learning materials and interpret data from the digital environment (Barana & Marchisio, 2021). In this model, the digital learning environment is not limited to technological artefacts, despite their crucial role, because the learning community has a prominent place. Regardless of whether digital learning environments are based on a web platform, teaching and learning activities can take place in one of the following modes:

- in-person, in a classroom or computer lab, with students working independently or in groups using digital devices or pen and paper while viewing the task on an interactive whiteboard;
- fully online: using the DLE as the sole learning environment in online courses or MOOCs;
- in a blended approach involves using platform-based activities to complement classroom work, for example by assigning them as homework.

These three modalities can be adapted to suit different situations, levels, objectives and requirements (Barana & Marchisio, 2021). In a DLE, students can access multiple resources, including interactive materials, links, videos, presentations, further reading, and both synchronous and asynchronous activities. Technologies such as web-based platforms and assessment tools allow learning data to be collected, recorded and utilised.

Elaborating on this data within the DLE makes it possible to provide useful information to inform decision-making and action (Barana & Marchisio, 2021).
The Digital Learning Environment can support the following processes:

- the analysis of interactions within activities that provide an assessment;
- the collection of quantitative and qualitative data relating to students' actions, use of materials and participation in the activities;
- feedback on activities, both for teachers regarding students' impressions and progress, and for students to help them progress in their learning path;
- sharing and collaboration within the learning and practice community.

The DLE enables teachers to monitor all students' activity and track which learning objectives have been achieved (Fissore et al., 2021). Combining game components with a DLE can support teachers or trainers in teaching and monitoring students' development of specific skills, allowing them to act as supervisors of the learning process. Conversely, game components could empower students to take ownership of their learning, allowing them to play an active role in the learning process and maintain a high level of engagement and interaction (Barana et al., 2022). Games offer potential for learning because the criteria for an engaging game are similar to those for a motivating learning environment - provided that learning and entertainment objectives are balanced. Consequently, learning games could promote an inclusive culture by providing students with an engaging, adaptive learning experience and enabling teachers to monitor both results and the learning process (Bente & Breuer, 2010). Educational games, including serious games, can be included within a DLE. When embedded within a DLE, serious games become part of a broader ecosystem. The LMS collects information about students' activities. This information is useful for learning analytics, as data generated during gameplay, such as completion times and error rates, can be collected and analysed to understand and improve teaching and learning processes (Zaki et al., 2020).
The following section explores in more detail gamification, game-based learning, and serious games.

2.3 Gamification, Game-Based Learning and Serious Games towards a Definition

The proliferation of technological and IT tools over the last few decades, such as computers, smartphones, tablets, which make it possible to play anywhere, has contributed to the demand for new methods, approaches and tools, such as games, in the field of education. At the beginning there was a strong scepticism towards the use of games in education, since it was believed that games had the only purpose of entertaining the participants (De Freitas, 2006). Today, however, gaming and gamification represent new forms of interaction, learning and exploration, contributing to the achievement of a defined purpose that is different from pure entertainment (Gounaridou et al., 2021).
In recent years, there has been a significant and growing interest in the fields of gamification and game-based learning (Zabala-Vargas et al., 2021). Researchers and educators are exploring how these innovative approaches can enhance the learning process, particularly by improving student motivation and engagement, and creating more dynamic and enjoyable learning experiences (Hussein et al., 2022; Aragonez et al., 2021). By incorporating game elements or utilising actual games in educational settings, educators seek to promote deeper learning and maintain students' interest over time. As highlighted in the literature review by Aragonez et al. (2021), the three approaches most commonly employed by researchers in this field are gamification, game-based learning, and the use of tools such as serious games.
Deterding et al. in 2011 provided a widely accepted academic definition of the term gamification and emphasised its potential as a promising area of future research. Gamification refers to the use of the typical mechanisms of the game, such as the challenge, the use of points, levels and prizes, in a context that is essentially not a game (Deterding et al., 2011). The term game refers to any form of structured play involving rules, goals and challenges, undertaken for the purposes of diversion or entertainment (Cheng et al., 2015). Games are linked to enjoyable problem-solving activities and function as self-contained systems that are structured by game mechanics, or rules (de Carvalho & Coelho, 2022). As McGonigal (2011) highlighted, games are delivered in a variety of forms, platforms, tools and genres today, but they all share four "defining traits":

- A goal provides players with a specific outcome to work towards, giving their actions purpose and direction.
- Rules place limitations on how players can achieve the goal, encouraging creativity and strategic thinking.
- A feedback system provides players with information about their progress towards the goal, motivating continued engagement.
- Voluntary participation ensures that all players willingly accept the rules and the goal, creating a safe and enjoyable space for play.

In contrast to entertainment games, gamification is characterised by its serious purpose, as Krath et al. (2021) highlighted. The core idea behind gamification is to leverage the most engaging and enjoyable elements of games in non-game products or contexts to maintain high levels of participation, involvement, and motivation. There are multiple definitions of gamification in the literature. These definitions tend to focus either on game elements and mechanics or on gaming processes and experiences outside of games context (Krath et al., 2021). The definition proposed by Kapp (2012) focuses on learning since it describes gamification as the use of "game-based mechanics, aesthetics, and game-thinking to engage people, motivate action, promote learning, and solve problems". Similarly, in Zichermann and Cunningham's (2011) definition, the use of aspects related to game experience to solve problems is also emphasised. In this view, gamification refers to "the process of game-thinking and game mechanics to engage and solve problems". Nevertheless, the concept of gamification has been adapted and applied in various fields, such as education, industry, business, and healthcare (Krath et al., 2021).

In education, in particular, gamification aims to foster positive attitudes that support and enhance the learning process (Deterding et al., 2011), by incorporating elements such as points, badges, leaderboards and storylines, turning ordinary tasks into engaging challenges (Qasim, 2024). This definition clearly delineates the area of research and distinguishes gamification from related concepts such as serious games, which will be explored later. The authors emphasise that the term refers specifically to game elements, such as rules, goals and competition, rather than to "play" in its broader sense. The authors also highlight that gamification is not limited to digital contexts. It can be applied in various non-entertainment domains, including education, health, journalism and marketing. The authors suggest that the term should not be restricted to a single field; the proposed definition is intentionally broad to encompass the wide range of gamification applications.

The specific features of games are essential to shaping the gamification experience (Fissore et al., 2023b, Huang et al., 2013):

- Involvement: players like to feel they have an active role in the game and they are therefore encouraged to participate dynamically. The challenge is one of the key elements to engage users within the game.
- Control: players like to feel they have power and control over their actions.
- Rewards: prizes, even small ones (e.g. badges or achievements) encourage players to keep going, especially when the rewards are regular.
- Fun: users enjoy the game because it is interesting and they will continue to play as long as there is interest.
- Competition: design challenges and tasks to motivate users, foster a sense of competition and guide them towards achieving specific objectives.
- Achievement: players are stimulated to continue the game if they have the impression that they are moving forward, getting better, and gaining skills. The use of levels in the game is an incentive factor for the user who can be motivated by the progression of the challenge. Another element used in games is the leaderboard to stimulate users to play and get better results.
- Accumulation: those who play appreciate the possibility of being able to accumulate rewards (e.g., money, treasure, points).
- Personalization: players like to customize the game, for example by choosing an avatar to develop their own identity.
- Adaptability: players like the possibility of making the game vary depending on their decisions, whose path is the result of their own choices.

By combining different tools such as points, challenges, leaderboards, levels, and badges, it is possible to create multiple and different game systems (Sümer & Aydın, 2022).
Different views exist on the framework for game elements and gamification (Kim et al., 2018). One of the simplest gamification frameworks was developed by Bunchball (2016). In this model, gamification is broken down into two fundamental components: dynamics and mechanics. Mechanics refer to the "various actions, behaviours, and control mechanisms that are used to "gamify" an activity - the aspects that, taken together, create a compelling, engaging user experience". Therefore, dynamics are the "compelling, motivational nature of this experience is, in turn, the result of desires and motivations". In this sense, game dynamics include rewards, achievement, and competition, while game mechanics involve the specific tools and processes, such as points, challenges, and leaderboards, that influence user behaviour and interaction within the system. The MDA Framework, developed by Hunicke, LeBlanc and Zubek (2004), is a foundational model in game design. It creates a bridge between game design and the player's game experience. The MDA framework consists of mechanics, dynamics, and aesthetics. Mechanics refers to the basic components of the game - essentially, how the game works at a technical level. Dynamics arises from players' interactions with the mechanics, shaping the game's behaviour and flow in real time. Finally, aesthetics represents the emotional responses and experiences that games evoke in players when they interact with the game system. The framework emphasises the importance of designing with the player's experience in mind, stressing that "the content of a game is its behaviour, not the media that streams out of it towards the player". In fact, the game is conceived more like an artefact than a medium. Another influential perspective on the gamification framework was proposed by Werbach and Hunter (2012). In their model, gamification elements are categorised into three hierarchical levels: components, mechanics, and dynamics - forming a pyramid structure. At the top of the pyramid there are the dynamics, which represent the broad, conceptual aspects that shape the overall user experience, such as progression, relationships, and emotions. Mechanics, which sit at the middle level and comprise the fundamental processes that drive user engagement, including challenges, feedback, and rewards. At the base of the pyramid, the components are the specific, concrete elements that implement the mechanics and dynamics, such as points, badges, story, leaderboards, levels, and avatars. The authors identify five principal game dynamics, ten important game mechanics and fifteen components, but suggest that generally no gamification project include all these elements together. One of the key recognised elements within the dynamics is narrative. This refers to the story or progression that gives meaning and context to the user's actions within the system. Integrating narrative - whether in the form of an event, myth, legend or mission - makes gamified experiences more engaging and emotionally resonant, as users feel they are part of a larger purpose or unfolding story (Chorianopoulos & Giannakos, 2014).
Another comprehensive model for understanding and designing gamification is the Octalysis Framework, developed by Yu-kai Chou (2018). This model divides human motivation into eight core drives, based on various psychological theories. Each core drive plays a significant role in motivating players through game mechanics. These drives are organized into a visual 'octagon', with each one representing a different psychological and motivational factor that influence human behaviour and engagement. The eight factors include: epic meaning, accomplishment, empowerment, ownership, social influence, scarcity, unpredictability and avoidance. For example, accomplishment can be driven by progression and skill acquisition, leveraging the individual's motivation to improve and grow. On the other hand, scarcity can be addressed by introducing game mechanics that encourage discovery and exploration. Unlike frameworks that primarily focus on game mechanics or components, Octalysis emphasizes the psychological factors driving user engagement. The Octalysis framework incorporates both positive and negative motivators for analysing user interaction within a game. By balancing these core drives, designers can create more compelling and meaningful gamified experiences that resonate on a deeper emotional level.
Gamification has close relations with two other concepts: game-based learning (GBL) and serious games (Krath et al., 2021). There are many definitions of GBL in literature. It is widely considered an educational approach that utilizes games (digital or otherwise) as the primary tool to facilitate the learning process (Plass et al., 2016). Perrotta et al. (2013) intend GBL as a form of experiential engagement where people learn through trial and error, approaching a topic not as content but as a set of rules or a system of choices and consequences. According to Qian and Clark (2016), GBL is an environment in which game content and play enhance knowledge and skill acquisition. There are definitions of GBL that place a stronger emphasis on its connection

to teaching and learning. For instance, Gris and Bengtson (2021) define GBL as "the use of games to support teaching and learning processes", thereby emphasising its educational role. Several studies have highlighted a strong link between GBL and the development of transversal skills: students actively participate in games designed to stimulate critical thinking, problem-solving, and teamwork, as well as the acquisition of knowledge and skills across various subject areas (Perrotta et al., 2013; Qian & Clark 2016; Ting et al., 2019; Qasim, 2024).

GBL may utilize serious games, which are tools designed not only for entertainment but also with the specific aim of learning and developing specific skills (Chen & Michael, 2005). These tools are created with the intention of, for example, educating, training, or improving skills (Chatzea et al., 2024; Qasim, 2024). Al Fatta et al. (2028) intend serious games as part of the digital gaming world, where the "serious or hidden purposes are well blended inside the digital application " and where the entertainment and enjoyment is not the primary objective of the application. Serious games aim to integrate learning objectives and real-world scenarios into their gameplay, allowing users to engage in interactive experiences that enhance their knowledge and problem-solving skills (Becke, 2021). Moreover, these tools enable the user to experience situations that would otherwise not be accessible and provides a safe place where multiple trials can be performed. For example, it is used in the health and safety field to educate people to behave responsibly (Gounaridou et al., 2021). Education is a field that offers a greater number of successful examples of the use of serious games (de Carvalho & Coelho, 2022).

The main difference between serious games and game-based learning, on the one hand, and gamification, on the other, is that the former two involve the use of complete games that are specifically designed for educational or training purposes (Deterding et al., 2011), whereas gamification is a broader concept involving the application of game features to non-game contexts (Qasim, 2024; Krath et al., 2021; Al Fatta et al., 2028). In game-based learning, tasks are redesigned to make them more engaging and interesting than they would be with non-game or gamified approaches (Plass et al., 2019). Becke (2021) provided a useful summary in the form of a table outlining the key characteristics of various game-related terms. The table clarifies definitions, highlights intended purposes and distinguishes between concepts such as gamification, serious games and game-based learning. All these concepts share the idea of using positive, gameful experiences for purposes other than entertainment, such as education or behaviour change (Krath et al., 2021).

As technology has advanced, games have evolved into video games, making it essential to take the digital dimension into account today. The term "Digital Game-Based Learning" (DGBL) is used when games take a digital form. According to Becke (2021), DGBL involves "learning of some knowledge, skills, attitudes that happens with the deliberate use of digital games". Hussein et al. (2022) developed another definition of DGBL based on well-established literature (Prensky, 2001; Qian, 2026). This definition emphasizes the interactive dimension associated with the use of digital games: "DGBL refers to a student-centered approach in which educational objectives and materials are embedded within gaming activities to motivate students to learn and enhance their skills and knowledge by providing them with an enjoyable and interactive learning environment". To begin with, it is important to understand the difference between digital and non-digital games. Non-digital games are those played without electronic devices, such as puzzles, card games, board games, or other physical activities (Pais & Hall, 2024). Jaipal-Jamani and Figg (2018), on the other hand, define digital games as rule-based interactive systems played on digital devices involving one or more players, which generate variable or quantifiable outcomes that can be ranked. The number of games played on electronic media is growing. With the advancement of technology, it can be said that gaming has largely shifted to digital and online platforms, as reflected by the growing number of studies focused on this field (Barboros, 2024).

Prensky (2001) identified 12 key characteristics that make computer and video games engaging for millions of people. First, games are enjoyable because they provide fun and pleasure, while also offering structured play through rules and goals that motivate players. Their interactive and adaptive nature keeps players actively involved, creating a sense of flow and continuous engagement. Games also provide feedback and clear outcomes, which promote learning and personal achievement. Moreover, the presence of competition, challenge, and problem-solving stimulates excitement and creativity. Finally, the social interactions and storytelling elements found in many games evoke emotional connections and a sense of belonging among players. Altogether, these twelve features explain why digital games captivate such a broad and diverse audience.

The relationships between game feature design, motivational aspects, cognitive learning levels and final satisfaction with the learning process in DGBL were investigated by Huang et al. (2013). Based on the responses of 264 university students after completing a single player open online serious game, the results of the exploratory factor analysis reveal three converging factors of DGBL features: game structure, game involvement, game appeal. Game structure includes several components such as rules, goals and explanations of game tasks. Game involvement refers to the features that encourage players to participate and immerse themselves in the gameplay. Game appeal encompasses the sound, graphics, and animation of the game, which can help engage learners cognitively and emotionally. The authors reassembled the various game features from prior studies and, through an exploratory factor analysis, identified only three game features. The authors suggest that the design complexity and development costs of DGBL could be significantly reduced.

Feedback is another essential feature to consider when designing game-based learning experiences. To ensure that DGBL environments effectively support learner engagement, it is crucial to include clear and timely feedback mechanisms throughout the learning activities (Velasco-Hernández, 2024). Similarly, Zabala-Vargas et al. (2021) identify the following as essential elements of a game: goals and objectives; narrative; rules; freedom to choose; freedom to make mistakes; rewards; feedback; visible status; cooperation and competition; time restrictions; progress; and surprise. These components must be structured cohesively through a compelling narrative and continuous feedback to sustain the learner's sense of progression and engagement. Among the factors that serious game designers should keep in mind to achieve students' learning success, Zhonggen (2019) highlights feedback and debriefing as factors that serious game designers should consider to support students' learning success. Debriefing consists of "communication sessions where information is shared and examined after gameplay". Ease of use and clear instructions were also demonstrated as impacting factors in game design.

Furthermore, instructional designers intending to integrate games into learning environments must carefully align game elements, such as mechanics and dynamics, with the intended educational objectives. However, as highlighted by Zabala-Vargas et al. (2022), a broader level of categorization must be considered at the outset of the design process. When designing or selecting games for educational purposes, it is essential to account for the variables that define the nature of the game itself. These include the game's primary purpose (its original intention), its digital or non-digital format, genre (e.g., action, puzzle, or role-playing), platform (PC, console, or mobile), and the expected outcomes—such as learning and behavioural changes, knowledge acquisition, content understanding, and the development of perceptual, cognitive, and motor skills. Establishing a coherent alignment between these overarching categories and the core game elements ensures that the gaming experience effectively supports the intended instructional objectives and learning outcomes.

2.4 Gamification, Game-Based Learning and Serious Games in Education

In recent years, research in the field of gamification and GBL has grown significantly (Barboros, 2024; Hussein et al., 2022) within the domain of education. The use of game elements and digital games for learning purposes has attracted increasing attention from researchers and educators alike, as it offers innovative ways to motivate students, enhance engagement and foster deeper learning experiences. However, the effects of gamification on learning vary greatly depending on factors such as game design, context, learner characteristics and instructional integration (Bertram, 2020). For example, Clark et al. (2015) conducted a systematic review and meta-analysis examining the influence of digital games on the learning of K–16 students (from primary school to college level). The authors found that digital games significantly improved learning outcomes compared to traditional methods, and that they can meaningfully enhance cognitive and intrapersonal learning outcomes such as motivation, work ethic and conscientiousness. Hanus and Fox (2015) conducted a longitudinal study which found different results. The authors examined the effects of gamification in higher education. They investigated whether integrating common game mechanics, such as badges, leaderboards and reward systems, could enhance or undermine students' motivation, satisfaction, effort, sense of empowerment and academic performance. Their findings showed that, compared to students in the non-gamified class, those in the gamified

class demonstrated lower levels of intrinsic motivation, satisfaction, and empowerment. Using leaderboards increased social comparisons and the perception of competition towards the end of the semester. Notably, final exam performance was lower in the gamified group, which was mediated by reduced intrinsic motivation. The authors highlight that gamification mechanisms can have negative effects on motivation and learning if not carefully designed and recommend avoiding mandatory reward systems.
Plass et al. (2016; 2019) identified the fundamental elements of game design for learning. These elements are:

- Game mechanics, which define the core actions through which learning and assessment occur. Game mechanics represent the central link between player actions and learning. Game mechanics can function as "learning mechanics", directly supporting learning objectives, or "assessment mechanics", which are used to evaluate learners' knowledge and skills.
- Visual aesthetic design includes the game's look and feels, characters, and the way key information is represented.
- Narrative structures refer to the game's storyline, which unfolds through cutscenes, dialogue, in-game events and player choices.
- Incentive systems regulate effort and persistence by offering intrinsic and extrinsic rewards.
- Musical score provides auditory feedback that directs attention to important events and signals success, failure, danger or opportunity, shaping the player's emotional experience.
- Learning content refers to the subject matter and competencies that the game is designed to teach. This influences the design of all other game elements, including mechanics, visuals, narrative, incentives and sound, as it determines how each design component is implemented.

One of the key challenges in educational game design is striking a balance between the playful nature of games and the learning objectives of the curriculum. As Qasim (2024) highlights, educational games must carefully integrate engaging gameplay with instructional content to ensure that entertainment does not overshadow learning goals. Curriculum alignment and collaboration between developers and educators are essential for effective design, ensuring content accuracy and methodological relevance. Similarly, Albano et al. (2020) emphasize this balance in their project. Their design addresses the challenge of creating games that resonate with students' gaming culture and are relevant to real-world problem solving. Furthermore, as Harrington and Mellors (2021) point out, successful gamification requires attention not only to game mechanics - such as points, badges, and narratives - but also to the learning objectives that underpin them. Gamification should thus be driven by clearly defined learning goals rather than by game elements alone. According to Kapp et al. (2014), gamification can be distinguished into structural gamification and content gamification. Structural gamification refers to the application of game mechanics - such as points, badges, levels, leaderboards, or progress bars - to an existing learning structure without altering the instructional content itself. Content gamification, on the other hand, involves turning learning content into a game. This approach incorporates educational material into narratives, challenges, quests, characters or simulations so that learning takes place through the gameplay itself. In this case, the content is redesigned to align with game mechanics. Both structural and content-based gamification should be combined to maximise engagement and learning outcomes (Filatro & Cavalcanti, 2016).
Vidakis et al. (2015) highlight the different roles involved in creating an educational game and emphasise the importance of cooperation between these roles. These roles include the educational expert, responsible for defining the educational theories and pedagogical principles on which the game is based; the game designer, who develops the game mechanics and materials in accordance with these principles; the instructor, who integrates and adapts the game for use in the classroom; the learner, who uses the game to enhance his/her learning experience.
GBL and gamification are not only valuable tools for student learning but also serve as effective strategies for teachers to monitor and track student progress (Ting et al., 2019) even in online settings (Di Biasi et al., 2022). Formative assessment can be effectively embedded within a game-based context to enhance engagement and learning outcomes. GBL enables learners to receive feedback and track their progress and level of participation in the learning path. This approach fosters a more personalised learning experience, allowing students to progress into a learning path while staying motivated through gameplay (Di Biasi et al., 2022).
Adopting GBL and gamification approaches also makes educational environments more inclusive. As highlighted by (Qasim, 2024), embracing these innovative approaches is essential for creating learning

environments that are more engaging, effective and inclusive. By incorporating playful and interactive elements, GBL and gamification can support different learning styles and address diverse learning needs, thereby encouraging participation among students who might otherwise feel disengaged or excluded in traditional settings. Thus, these approaches have the potential to promote a more democratic approach to learning, making education more accessible and meaningful for all learners. In this regard, both games and gamification should be accessible and flexible across different educational contexts, as well as sustainable in the long term. According to Harrington and Mellors (2021), gamification is scalable: key strategies and activities can be initially integrated into specific aspects of a teaching session and later expanded to cover larger portions of the courseware. Gamification is also flexible because it taps into deeply intrinsic human drives, such as competitive behaviour and the desire to excel. This allows almost any element of a course activity to be adapted to these drives through practical motivational triggers, including external feedback and rewards. Zhonggen (2019) also highlights that, compared to traditional learning, serious games are more flexible. For example, they may offer the option to play them whenever possible.

Several studies have emphasised the importance of providing teachers with specific training to help them overcome the barriers and challenges they face when implementing gamification and GBL in the classroom. Teachers' perceptions are crucial to the adoption of these approaches. Sánchez-Mena and Martí-Parreño (2017) highlighted the main factors encouraging teachers working in higher education institutions to use gamification in their courses: its ability to attract students' attention, its entertaining nature and its contribution to more interactive and engaging learning. Common obstacles include a lack of resources, limited preparation time, mixed student interest and uncertainty about the suitability of gamification for specific subjects. In addition to the practical challenges of implementing gamification in the classroom, teachers often feel unprepared to design compelling gamified experiences during the planning phase. Teachers may not fully understand the potential of gamification in education, believing that it can only be applied to a limited number of subjects. Research in teacher education confirms the importance of targeted professional development. As González-Fernández et al. (2023) demonstrated in their systematic review of gamification in teacher training, there is a widespread lack of understanding of gamification processes and instructional design models among teachers. This often results in game elements being included randomly, without methodological coherence or clear learning objectives. A similar conclusion emerges from the study of Zabala-Vargas et al. (2021) which emphasises the need for structured training programmes to help educators integrate gamification more consciously and effectively into their teaching practice. In their study, Fissore et al. (2023a) presents the results of a workshop on gamification and education that involved 54 Italian teachers of different levels, many of whom had limited prior exposure to gamification. They found that the sample considered is composed by teachers who are willing to actively involve and motivate their students in the learning process and that they were open to discover and learn new teaching methodologies to use in their teaching practices for these purposes. Teachers collaboratively developed gamified lesson plans through structured guidance, demonstrating that, with the right support, they can successfully integrate gamification even without advanced digital skills. This fact highlights also the need to train teachers who are very often attracted by innovative methodologies such as gamification, but they do not know how to put them into practice in their daily teaching practices. Malvasi et al. (2022) in their research found that “unconscious gamification” was observed in Italian high schools: elements of gamification were being applied intuitively, despite teachers having not received any formal training in the field. As pointed out by González-Fernández et al., there is a general lack of knowledge about the process of gamification systems, which leads to the introduction of gamification elements without a specific criterion. Therefore, there is a need to introduce gamified practices in teacher training, providing experiential learning that allows them to apply this methodology to their professional life in a relevant way (Fissore et al., in press).

While much research has examined the challenges teachers face when incorporating gamification into their classrooms, it is also worth examining the most prevalent gamification strategies used in educational settings. By analysing the guided forms completed by 54 Italian teachers during the design stage of a workshop on gamification and education (Fissore et al., 2023a), it was possible to identify the most common strategies teachers used to create activities that enable students to develop skills and knowledge in their subject areas. All the didactic activities entail collaboration among students, for instance, to accomplish a shared objective. The most frequent gaming factors are involvement, rewards, personalisation and progress. These have mainly

been implemented through challenges, levels, leaderboards, and points, coins, or treasure. Moreover, the initial questionnaire and the guided design forms revealed that teachers struggle to make learning adaptable.
As Tomczyk and Teckchandani (2023) pointed out, a potential danger of using gamification strategies is that educators may focus too much on designing a game or making the experience enjoyable, resulting in the educational element being overshadowed or lost in the fun. Alternatively, an educator may lean too heavily on the conventional approach, resulting in a lesson that closely mirrors a traditional classroom setting. In order to guide teachers in the design phase, Tomczyk and Teckchandani (2023) identify some steps to gamify a lesson: (1) determine the necessary concepts to be covered in the experience, (2) identify learning outcomes, decide how much time one has to create, (3) prepare, explain, run, and debrief the experience, (4) select and use gamification strategies to build the experience. Some advice is given on the last steps: encouraging students' curiosity, allowing them to make meaningful choices so that they are not just passive players, and using game-based learning activities to provide a shared experience and encourage collaboration between students.
While on the one hand there are still barriers to the introduction of game elements in education, on the other hand, current technologies increasingly support teachers by providing accessible tools that enable the development and integration of game features into existing digital platforms and everyday teaching practices, without requiring extensive digital skills from educators. For instance, gamification can be seamlessly incorporated into learning management systems by leveraging elements such as points, badges, and certificates. These features not only provide students with immediate recognition for their achievements but also foster a sense of progress and accomplishment, encouraging them to complete courses, actively engage in discussions, and participate more consistently in learning activities. Moreover, they offer teachers detailed analytics on student performance, allowing for more informed and targeted instructional interventions (Qasim, 2024). For example, based on the effective gamification strategies identified in the literature on distance learning (Sümer & Aydın, 2022), Floris et al. (2023; 2024) implemented these strategies within a DLE in a manner similar to that employed by Lee et al. (2023). The aim was to reinforce the existing gamification elements and boost participant engagement, particularly regarding developing disciplinary competences in Mathematics, as well as problem-solving skills and competences in education for sustainable development. The gamification strategies in the DLE were implemented using the commercial Moodle plugin, Level Up XP+. This allowed students to earn points and level up based on their participation in course activities. In the storytelling element, each student receives a mission linked to achieving the sustainable development goals of the 2030 Agenda adopted by all United Nations (General Assembly, 2015). Students could automatically accumulate custom points called Digital Math Equos (DMEs) through forums, online meetings, submitting problems and answering questionnaires. Progression involves each level corresponding to a specific sustainable development goal, with an additional "Super Sustainability" level for exceptional performance. When users reach a higher level, a motivational message was displayed alongside the badge for that level. The challenge was not only to win the final prize, but also to earn the most DMEs, level up, and achieve the greatest number of the sustainable development goals of the 2030 Agenda. This would ensure a fair society. A personalised leaderboard showed students' positions relative to their peers, and a personal report shows learners' progress and allows the course teacher to see how many points each student has earned, how they earned them, and what level they have reached. Students received different types of feedback: weekly through problem evaluation, instantaneously through online tutoring or from Level Up when they reached a new level, and daily through forums where tutors are available to answer students' questions and queries. Users also receive a notification when they earn points, explaining the number of points awarded and the activity for which they have been awarded them. The study found that there was a general increase in interaction within the forums and that students with the highest motivation levels expressed greater appreciation for the gamification strategies implemented in the DLE. Similar findings were reported by Yong et al. (2016), who found that students who were more interested in Mathematics were more likely to support the use of digital games for learning purposes. On the other hand, Li et al. (2024), based on a meta-analysis involving 2,500 students, found that gamification is an effective educational strategy for enhancing motivation among secondary school students, regardless of their initial motivation levels, although the effect size was small.
In their systematic literature review, Hussein et al. (2022) highlight that, regarding the design features and underlying theories of DGBL, less than half of the studies considered provide background information on the design features and theoretical foundations of the DGBL applications employed in their learning interventions.

Despite the growing interest in the field of GBL, the design of serious games is still an unstructured process. Gentile et al. (2014) also emphasized that there are no proper guidelines based on a validated methodological approach.

The models identified in the literature on game design, which are explained below, offer various approaches and frameworks for designing games and gamified learning experiences. According to Garris et al. (2002), GBL can be understood through an input–process–output model (Figure 2). In this framework, input refers to the instructional contents and design features of the game, such as its rules, challenges, feedback and goals, which are intentionally structured to stimulate learning. The process involves the learner interacting with the game, which activates engagement, motivation, behaviors and cognitive processes through gameplay. The output represents the learning outcomes and behavioral changes resulting from this interaction, such as improved knowledge, skills or attitudes. This model emphasizes that the effectiveness of GBL depends on how well the game design elements (input) elicit the desired processes that ultimately lead to meaningful learning outcomes. Games should be designed to sustain motivation and encourage learners to engage in repeated cycles of play. The game cycle is an iterative process involving continuous loops of judgement, behavior and feedback. During gameplay, users experience reactions such as increased interest, enjoyment or confidence, which motivate behaviors such as persistence and effort. These behaviors generate system feedback that informs future actions, thereby keeping players engaged in repeated cycles of play.

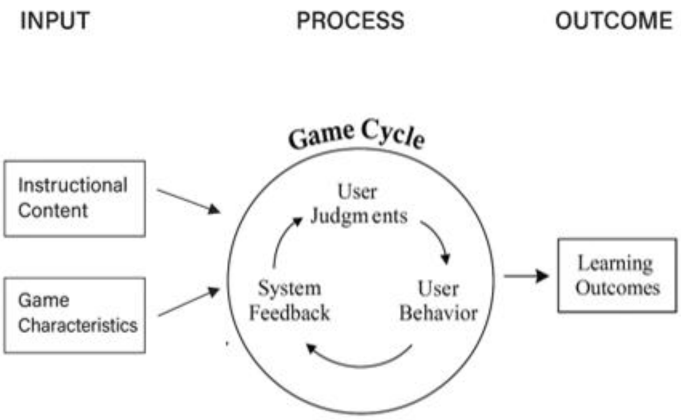

Figure 2 – Model of game-based learning (Garris et al., 2022).

Lee at al. (2023) designed and implemented a gamified Mathematics course, based on the Analysis, Design, Development, Implementation, and Evaluation (ADDIE) model, which is a widely used framework for the design and development of online and blended courses, often used in the field of distance education and e-learning. Analysis, Design, Development, Implementation, and Evaluation are used to systematically and cyclically approach to instructional design, allowing for continuous improvement and refinement of instructional materials over time. They designed an ADDIE-based model to study fundamental Mathematics. Within this framework, an initial analysis of the students involved was conducted, after which mathematical content was explored through gamification components integrated into the LMS (Design). Corresponding teaching materials were then created to reflect these components (Development), after which the course was piloted (Implementation), followed by an evaluation to verify its effectiveness (Evaluation). Results showed that students with stronger math performance responded more positively, while lower-achieving students benefited less. The study suggests that more complex motivational strategies are needed to support struggling college students better.

2.4.1 Game-Based Learning in Mathematics

Students often perceive Mathematics as a complex and abstract subject (OECD, 2024; Christopoulos, 2024). This perceived difficulty can lead to frustration and a general lack of interest or enjoyment in the subject and many students suffer from math anxiety, which is characterised by feelings of fear when faced with mathematical tasks or think about numbers (Chen et al., 2023; Weir, 2023). The following section will provide a detailed discussion of attitudes towards Mathematics, paying particular attention to aspects such as math anxiety and its impact on students' motivation, performance and overall engagement with the subject.

In this context, it is important to encourage students to learn Mathematics and to use resources that make maths lessons more engaging and to promote meaningful learning experiences (Pais et al., 2023). One of the greatest challenges teachers face today is effectively engaging students, motivating them to learn and encouraging them to play an active role in their learning process (Pais et al., 2023). This is particularly difficult with subjects like Mathematics, which are often poorly appreciated by students. The challenge is even greater with older students who have spent years being exposed to traditional learning methods and may now be resistant to new, more active and different approaches. Qasim (2024) emphasises that games often pose challenges that require students to think critically and develop solutions, thereby enhancing their problem-solving abilities. This section presents some case studies that illustrate the use of game-based learning and serious games in Mathematics education, particularly in upper secondary and higher education contexts. A systematic literature review will explore the topic of GBL and gamification for Mathematics education in upper secondary school and higher education in greater depth in a later section.

There is evidence that GBL and gamification have a positive impact on students' motivation and attitudes towards learning Mathematics (Zabala-Vargas, 2020; Zabala-Vargas, 2022; García-López, 2023). Studies conducted by Zabala-Vargas et al. (2019) and Clarke et al. (2016) have shown that the use of serious games in Mathematics education can have a positive impact on university students' motivation and attitudes towards learning Mathematics, and it can also help to reduce math anxiety (Piñero et al., 2024).

Positive results have also been obtained in the high school setting by Barbieri et al. (2021), who found that serious games have a positive impact by fostering better emotional attitudes towards Mathematics, making the subject more engaging and enjoyable. They also state that serious games enhance the teaching and learning process in Mathematics by providing a more interactive and motivating experience and that offered teachers the opportunity to reflect on their teaching methods, prompting them to adjust their strategies to make Mathematics teaching more effective. Another benefit of serious games is that they allow the behaviour of users to be measured and data to be collected based on their actions within the game (Barbieri et al., 2021).

Cruz et al. (2023) describe the design, development and evaluation of a serious game called ORUN-VR2, which aims to enhance students' understanding of projectile motion (kinematics). The study, which involved over 130 high school students, compared the effectiveness of a serious game to that of traditional instruction. On average, students who played ORUN-VR2 demonstrated a learning gain of 52%, compared to the control group. Overall engagement was rated as moderate, with strong results for immersion and high usability scores. No substantial gender differences were found, both male and female students achieved similar learning outcomes using ORUN-VR. The study also offers also design guidelines. After choosing the topics, the game design phase incorporated mechanics, narrative and user requirements based on feedback. Following pilot testing of an initial version, the final version was released. There are experiences that demonstrate how serious games can be used to address the difficulties that students often encounter when learning Mathematics. For example, the study of Correia et al. (2023) presents a serious game called "Dyscalc Game", designed to support students with learning difficulties in Mathematics, including dyscalculia. The study involved participants aged 15–17. The serious game combines educational and entertainment elements to boost motivation and support the teaching and learning process. The results were positive in terms of engagement, demonstrating the game's strong potential as a tool to support students with maths learning difficulties. Several studies (Nautiyal et al., 2024; Chorianopoulos & Giannakos, 2014) suggest beginning the design phase with popular and simple games to minimise the effort and time required for student orientation, emphasising learning objectives over game mechanics. Nautiyal et al. (2024) propose a five-step framework to help teachers integrate board games into their classrooms for problem-solving. The five phases comprise:

- Adopt: choose a familiar commercial game to reduce learning time.
- Enhance: modify rules to introduce strategic interaction and reduce luck-based outcomes.
- Align: integrate curriculum content (e.g., math problems) into gameplay.
- Implement: conduct student-led GBL sessions, with the teacher observing engagement.
- Repurpose: adapt and reuse games for new lessons or subjects, encouraging teacher collaboration to design new games.

The first three phases of the framework, Adopt, Enhance, and Align, primarily focus on the design stage, where teachers select, modify, and adapt games to fit learning objectives. The last two phases, Implement and Repurpose, concern the implementation in the classroom. The model outlined by the authors is also based on ease of use for teachers, without requiring any advanced digital or technical skills, for those who are not experts in game design. The authors applied the framework to design two educational board games and implement a two-week GBL-based Mathematics instruction in a high school in the Philippines. The qualitative analysis of student feedback revealed that the design elements integrated into the framework positively impacted the student experience. The authors also highlight the importance of including student feedback during the classroom integration phase, underscoring its significance for educators in implementing the framework.

2.5 Attitude towards Mathematics: Motivation, Perception, Math Anxiety

In the literature on gamification and game-based learning, considerable attention is devoted to the theme of motivation as it is widely recognised as a key factor influencing students' engagement with, attitudes towards, and performance in the subject. As Zabala-Vargas et al. (2022) point out, Keller's (2010) definition of motivation is "what people want to do, what they decide to do, and what they commit to doing". Another definition of motivation is the "impulse that human beings have to satisfy their needs in different dimensions" (Maslow, 1943). Keller (2010) agrees that motivation in a learning setting has both internal and external dimensions. Intrinsic motivation refers to performing a task for its own sake, because the learner finds it interesting, meaningful or satisfying. In contrast, extrinsic motivation refers to performing a task for the sake of an external consequence or reward (or to avoid a negative outcome). For instance, a learner may engage in an activity in the hope of receiving a grade, certificate, recognition or other external reward. Krath et al. (2021) conducted a systematic review motivated by the absence of a comprehensive overview and analysis of the theoretical foundations of gamification and GBL in the existing literature. The systematic review revealed that research on gamification, serious games and GBL has drawn on a wide range of theoretical frameworks from diverse perspectives to design and assess gamified interventions, as well as to explain the mechanisms through which gamification works and it produces effects on learning, including theories related to motivation, affect, behaviour, and learning. They also highlight a positive relationship between the use of serious game or gamification and motivational outcomes (Krath et al., 2021).

Among the instruments used to assess students' motivation and learning strategies, there is the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991). This self-report questionnaire is designed to analyse students' motivational orientations and learning strategies. Students rate themselves on a seven-point Likert scale ranging from "not at all true of me" to "very true of me". The complete version of the MSLQ includes 81 items organised into two major sections. The first one is the Motivational Section that consists of 31 items assessing students' goals and beliefs, for example about the perceptions for a course, their self-efficacy for success, and their test anxiety. It includes six subscales: Intrinsic Goal Orientation (reflecting intrinsic motivation and interest in learning), Extrinsic Goal Orientation (measuring external motivation, such as grades or rewards), Task Value (evaluating the perceived importance and usefulness of the task), Control of Learning Beliefs (referring to the belief that outcomes depend on one's own effort), Self-Efficacy for Learning and Performance (capturing confidence in one's learning abilities) and Test Anxiety (assessing anxiety related to examinations and assessments). The Learning Strategies Section comprises 31 items concerning students' use of various cognitive and metacognitive strategies and a further 19 items related to the management of learning resources. This section comprises nine subscales: Rehearsal (involving repetition and memorisation strategies), Elaboration (connecting new information to prior knowledge), Organisation (structuring and summarising learning materials), Critical thinking (applying critical and analytical reasoning), Metacognitive self-regulation (planning, monitoring and evaluating one's learning process), Time and study environment management (managing time and study settings effectively). The authors reported good levels of internal reliability for most of the subscales. Reliability refers to the degree to which a test or a scale consistently and accurately measures a given variable. One method to assess reliability is to evaluate internal consistency, i.e., how well the different items in a questionnaire measure the same construct (Poels et al., 2007; Ranganathan et al., 2024). A common statistic used to assess internal consistency is Cronbach's alpha, with values closer to 1 indicating higher reliability. Cronbach's alpha scores are generally considered acceptable to good when they are 0.7 or higher (Poels et al., 2007). In the original study by Pintrich et al. (1991), the Cronbach's alpha coefficients for the various subscales generally ranged from 0.52 to 0.93. The Self-Efficacy for Learning and Performance, Task Value and Metacognitive Self-Regulation subscales showed α values

greater than 0.80, indicating a good internal reliability. The shorter subscales, such as Rehearsal and Help Seeking, demonstrated lower values.
In the field of Mathematics education, students' attitudes and perceptions towards Mathematics have been widely investigated, as they play a crucial role in shaping learning behaviours and achievements with the subject. The PISA 2012 Technical Report (OECD, 2014) devotes considerable space to this theme, including various scales to measure emotional and motivational factors related to the subject. This large-scale international study assesses the mathematical, reading, scientific and problem-solving skills of 15-year-old students from 65 countries. The proposed surveys are designed to collect information on students' backgrounds, learning environments, and their attitudes, motivations, and perceptions towards learning, providing valuable insights into the relationship between emotional and attitudinal factors and academic performance. The scales and constructs of the questionnaires proposed in the OECD's (2014) PISA 2012 Technical Report are derived from a combination of recognised theoretical models, established academic research, and indicators already used in previous PISA surveys. Within the PISA 2012 Student Questionnaire, attitudes and perseverance towards the subject received considerable attention, and there were also items measuring math anxiety (MA) that were included and used in the Main Survey of PISA 2012. Within the field of Mathematics education, the issue of MA has received significant attention. Math anxiety is defined as a negative "feeling of tension, apprehension, or fear that interferes with math performance" (Ashcraft, 2002). Richardson and Suinn (1972) further specified that MA is characterised by feelings of tension and apprehension that interfere with the ability to work with numbers and solve mathematical problems in everyday and academic contexts. MA can affect students' performance, for example by preventing them from passing Mathematics classes or pursuing advanced Mathematics courses. Recent research has shown that MA tends to increase with age and can persist throughout secondary and post-secondary education, extending into adulthood (Maisey et al., 2022; Dowker et al., 2016). Notably, numerous studies have consistently demonstrated a negative correlation between MA and mathematical achievement. For example, Maisey et al. (2022) reported that higher levels of MA are associated with lower Mathematics performance, while Wen and Dubé (2022), in their systematic review, found a clear negative linear relationship, indicating that as MA increases, performance tends to decrease. These findings emphasise the long-lasting impact of MA on students' ability to succeed in Mathematics and highlight its significant detrimental effect.

2.6 Evaluation of the Impact of Game-Based Learning and Serious Games

Recent studies indicate that the discussion on assessing the impact of game-based learning in education is ongoing, reflecting both the growing interest in game-based learning and the complexity of measuring its educational outcomes (Barana et al., 2025). In their systematic review, Gris and Bengtson (2021) examined the assessment of GBL effectiveness, considering the dimensions of learning, engagement, and usability. In the studies they reviewed, they found that learning aspects are evaluated much more than engagement and usability features. Furthermore, in contrast to learning, engagement and usability are almost exclusively assessed through indirect measures, which are related to thoughts and perceptions about the learning experience. The assessment of engagement and usability shows a lack of measurement in terms of established reliability and validity. Tools used to assess GBL include instruments that have adapted validated metrics such as the System Usability Scale (SUS) (Brooke, 1996; Lewis, 2018) and the Technology Acceptance Model (TAM) (Davis,1989). The SUS is a ten-item Likert scale questionnaire designed to assess the usability of a system, product or interface. It is widely used and valued for its flexibility, ease of implementation and ability to provide a quick measure of usability in many areas. Developed by Fred Davis in 1989, the TAM is a theoretical framework that aims to explain and predict how users come to accept and use new technologies. The model identifies two key factors that influence a user's decision to accept and use a technology, each measured through a six-item Likert scale questionnaire. The first is "Perceived Usefulness", which refers to the degree to which a person believes that using a particular technology will improve their job performance or overall experience. The second is "Perceived Ease of Use", which refers to the degree to which a person believes that using a technology will be effortless. Several studies have utilized TAM to evaluate game-based learning (Gris & Bengtson, 2021). The Perceived Usefulness and Perceived Ease of Use scales of the TAM are reliable and validated measures (Davis, 1989). The SUS has also been recognized for its high reliability and robustness (Brooke, 1986).

Valencia and Duque (2023) adapted the TAM to analyze the perception and intention of teachers to use serious games as teaching-learning strategies in higher education. Factor analysis confirmed five constructs in their model: attitude, ease of use, confidence, intention to use, and perceived usefulness, which were examined to understand their interrelationships. As a result, attitudes towards using the serious game are mainly explained by perceived confidence and ease of use. Furthermore, additional results were obtained by Chauhan et al. (2021), who built a model based on the TAM to analyze the most important factors that encourage the use of games. In their analysis, they identified eight factors that are important when considering online games: social interaction, context of use, perceived ease of use, perceived enjoyment, perceived usefulness, attitude, flow and behavioral intention. It emerged that perceived usefulness, attitude and flow (intended as involvement in the game) are the most important factors influencing the model and should therefore be considered for the development of games. In this way, the study provides a better understanding of adoption behavior and helps game developers to improve their services. SUS has been applied in several contexts, including in DGBL research (Marques & Pombo, 2019; Perini et al., 2018). Morais Marques and Pombo (2019) used SUS to assess teachers' readiness to adopt mobile augmented reality technologies in their teaching, bridging the gap between educational practice and educational research. The average scores were very high, achieving excellent usability for teachers. In addition, Perini et al. (2018) included the SUS in a usability evaluation questionnaire completed by 62 university students. Their study shows that a higher level of usability and enjoyment is essential to promoting student involvement in the learning process.

Another important factor to consider in the evaluation of a serious game is the overall gameplay experience. The Game Experience Questionnaire (GexpQ) is one of the tools identified in the literature for evaluating video games. The GexpQ was developed by IJsselsteijn, et al. (2007; 2013) as a tool for evaluating players' subjective experiences during and after gameplay. Designed within the framework of the Fun of Gaming project, it provides a standardized method of measuring the various aspects of the gaming experience using self-report data. The questionnaire has a modular structure (IJsselsteijn, et al., 2013) comprising three main components: the Core Questionnaire, the Social Presence and Post-Game modules. The first module has 47 questions, while the last two have 17 each. Each component is scored as the mean of its corresponding items, rated on a five-point Likert scale ranging from "not at all" (0) to "extremely" (4). The Core Questionnaire Module consists of seven components: Competence, Sensory and Imaginative Immersion, Flow, Tension/Annoyance, Challenge, Negative affect and Positive affect. The Social Presence Module consists of three components: Empathy, Negative Feelings and Behavioural Involvement. The post-game Module consists of four components: Positive Experience, Negative experience, Tiredness and Returning to Reality. Empirical analyses conducted to assess the reliability of the GexpQ (et al., 2007) demonstrated satisfactory internal consistency across its main components. These results confirm that the questionnaire exhibits good reliability, and it can be considered a reliable tool for assessing players' multidimensional game experiences.

Another commonly used tool to evaluate levels of engagement specifically in the context of video games is the Game Engagement Questionnaire (GengQ) (Cruz et al., 2023). The tool was developed by Brockmyer et al. (2009). It is a self-report measure that assesses player engagement and consists of 19 items rated on a Likert scale, which can vary from simple yes–no options to more detailed multi-point scales. The items are organized into four dimensions: Absorption, Flow, Presence, and Immersion. Absorption refers to the ability to maintain attention and develop consciousness. Flow refers to the "feelings of enjoyment that occur when a balance between skill and challenge is achieved in the process of performing an intrinsically rewarding activity". Presence refers to the feeling of being in a normal state of consciousness while situated within a virtual environment. Finally, Immersion is defined as the ability of a game to make players feel as if they are part of the game environment (Cruz et al., 2023). Unlike the GexpQ, the GengQ focuses specifically on violent video games. In fact, the motivation behind the GengQ was to develop a measure of engagement with video games that could be used to assess their potential impact, particularly that of violent games (Norman, 2013). Although both instruments examine the subjective nature of gameplay, the GengQ examines the mechanisms of engagement in more detail, while the GexpQ provides a multifaceted overview of players' emotional, cognitive and social experiences (Norman, 2013). In terms of reliability, the authors (Poels et al., 2007) report good results for the GengQ, with a Cronbach's alpha of 0.85. In their study, Cruz et al. (2023) adopted the GexpQ to assess participants' engagement with a serious game specifically designed to study the projectile kinematics. The overall average engagement score was 3.0 out of 5 on the Likert scale, indicating a moderate level of engagement across all dimensions. The highest results were achieved with items in Immersion, while the lowest scores were achieved with items in Flow.

2.7 Integrating GenAI into Game Features

According to the OECD definition: Artificial Intelligence (AI) refers to "a machine-based system that for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment." (OECD, 2024C). AI systems interact with us and act on our environment, either directly or indirectly. They often appear to operate autonomously and can adapt their behaviour when provided with additional context. This is just one definition of AI, but it seems relevant due to its future-proof, human-oriented and data-dependent nature (Holmes et al., 2022). According to the DigComp 2.2 and DigComp 3.0 framework (Vuorikari, 2022), educators should develop age-appropriate tools and materials to ensure that students at all levels can access basic AI concepts and identify, evaluate, and navigate the complex information landscape in the AI age. Achieving these goals is essential to understanding the importance of AI from both a theoretical and technical perspective. The use of AI in education reveals a notable discrepancy between the user-friendliness of AI-based educational tools and the complexity of the underlying mechanisms and technologies, which teachers often struggle to explain. Teachers must be trained in both the theoretical content related to these topics and in planning teaching activities based on innovative methodologies (Fissore et al., 2022). A national survey of 284 Italian teachers from all levels of education revealed a strong need for greater knowledge and training on AI and gamification in Italian education. At the same time, teachers expressed a strong interest in using these tools in their teaching (Fissore et al., 2024a). The findings also suggest that AI could support teachers in creating increasingly engaging, motivating and personalized didactic activities for their students. AI can also help teachers integrate specific gamification strategies that they consider essential for their students' motivation and engagement (Fissore et al., in press). AI can transform teaching and learning at all levels of education and in different fields, for example GenAI can support collaborative learning and problem solving (Barana et al., 2023).
The connection between AI and education involves three areas:

- learning with AI (e.g., the use of AI-powered tools in classrooms);
- learning about AI (its technologies and techniques);
- preparing for AI (e.g. enabling all citizens to better understand the potential impact of AI on human lives).

Generative artificial intelligence (GenAI) refers to artificial intelligence systems that can generate text, images, videos and other materials using deep learning models, in response to complex and varied prompts, such as languages, instructions or questions (Lim et al., 2023). Building on survey findings involving Italian teachers, subsequent research (Fissore et al., 2024a; 2024b) investigated whether Italian lower secondary school teachers could make significant use of both gamification GenAI in designing didactic activities. The study examined the most used gamification elements and GenAI tools, as well as their added value. The results show that gamification and GBL could benefit from new GenAI tools. These technologies can be used to create materials that are more personalised, dynamic and engaging. GenAI allowed teachers to quickly create engaging and motivating educational and narrative content, which is useful for designing game-based activities. This is an enormous advantage for teachers who wish to incorporate gamification experiences into their teaching practice but feel discouraged by the challenges associated with the time required to produce student materials. According to teachers, GenAI may also encourage students to become critical and active users of AI tools. Another strength of GenAI that emerged in the study was the ability to create innovative and original materials, especially for those who struggle to create imaginative materials. Regarding the potential of gamification strategies, the greatest benefits were seen in making learning more engaging and fun. Compared to previous studies, teachers' use of storytelling in designing didactic activities has significantly increased with the advent of GenAI (Fissore et al., 2023a; Fissore et al., 2023b). The most widely used GenAI tools among teachers include chatbots, as well as tools for creating pictures, concept maps and stories (Fissore et al., 2024b).

2.8 Game-Based Learning, Gamification and Serious Game to Learn and Teach Mathematics: a Systematic Literature Review

During the development of the theoretical framework for this thesis, it became necessary to conduct a systematic literature review to provide a solid, up-to-date foundation for the research. The review focused particularly on game-based learning in Mathematics education, providing an overview of the current state of research and identifying both well-established and emerging trends. This activity began in the first year of the PhD programme, but significant progress was made in the third year thanks to the acquisition of more refined methodological tools, which enabled a more in-depth, critical evaluation of the review findings.

In recent years, research on game-based learning has gained increasing attention, driven in part by technological advancements. This systematic review aims to understand the current status and potential of GBL and serious games research in Mathematics. The idea is to expand the research including more recent articles, incorporating both well-established, peer-reviewed works and more recent, less formally validated studies. Given the dynamic and fast-evolving nature of the game-based field, where emerging innovations often precede formal academic validation, an appropriate approach was needed to capture current trends. For this reason, both Google Scholar and Scopus were chosen as the databases for this review. The research was narrowed down to papers in English that focused on gamified content for Mathematics learning in upper secondary and university education. Although the thesis aims to develop a digital gamified model, broader contributions related to game-based learning and serious games were considered to capture the fundamental methodologies, principles, strategies, and engagement mechanisms that are common to both digital and non-digital contexts. This was undertaken to ensure that the model was based on well-established educational theory, rather than being limited to the latest technological tools.

The initial research question that this review aimed to answer was RQ0: "What is the current state of the art in the literature on the use of game-based learning and serious games for teaching and learning Mathematics at upper secondary school and university level?

To answer the RQ0, a systematic literature review was conducted on the current state of using game-based learning to teach and learn Mathematics at upper secondary and university level. A specific timeline and method were outlined. The PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) guidelines guided the selection of the studies (https://www.prisma-statement.org/). The PRISMA flow diagram has already been adopted in previous studies within game-based learning domain to guide the article selection process (Gris & Bengtson, 2021; Ilić et al., 2024; Gusti, 2025). The PRISMA flow diagram consists of four steps:

1. Identification: perform a search of databases and other sources for relevant studies, recording the total number found.
2. Screening: review the titles and abstracts of the studies, removing those that are irrelevant or incoherent.
3. Eligibility: read the full texts of the selected studies and exclude those that do not fit the fixed criteria.
4. Inclusion: the final group of studies to be included in the literature review is established.

During the identification phase in July 2025, the keywords for the research were selected, as well as the way in which they should be combined. The first search was carried out in Google Scholar using a combination of keywords and Boolean operators (see S0 string in Table 1. This query yielded 17,900 results in Google Scholar and 316 results in Scopus; both of which were considered too many for the purposes of this review. To generate an acceptable number of results and align with the objectives of the thesis, the research question needed to be more specific. The research question was divided into more dimensions aligned with the aims of this thesis, resulting in:

($RQ0_{\_A}$) How does the use of game-based learning and serious games in Mathematics education affect student motivation at upper secondary and university level?

($RQ0_{\_B}$) To what extent do serious games and game-based learning enhance student engagement in learning Mathematics at upper secondary and university levels?

($RQ0_{\_C}$) What impact do serious games and game-based learning have on students' math anxiety in upper secondary and university Mathematics education?

($RQ0_{\_D}$) How are serious games or any other game-based tools for Mathematics learning evaluated in the literature in terms of usability?
($RQ0_{\_E}$) What evidence is there that serious games and game-based learning approaches improve learning outcomes in Mathematics at upper secondary and university level?
Therefore, in the second stage of the research, additional keywords were added that incorporated the concepts and outcomes of the theoretical framework and the research questions. The additional words were motivation, engagement, math anxiety and usability. Initially, the idea was to add each of these terms separately as variables to the original query string (S0) to examine each dimension individually. Restricting the search terms to the title, abstract and keywords of the article produced an acceptable number of results in Scopus, while this approach yielded an excessive number of results in Google Scholar. The S1 string, which combined all the new terms with the S0 string, was used as a solution for the Google Scholar database. This produced 136 results when restricted to scientific articles. In Scopus, however, each dimension was investigated using each word and combined it with the S0 string to produce the S2, S3, S4 and S5 strings (see Table 1, which shows the number of results found in the second stage of the research).

Database	**Code of the string**	**Search items**	**Number of results found**	**Number of results removing duplicates and not open access**
Google Scholar	S0	(“serious game” OR “game based learning”) AND Mathematics AND education AND (“upper secondary school” OR “high school” OR “university” OR “college”)	17,900	Not considered
Scopus	S0	(“serious game” OR “game based learning”) AND Mathematics AND education AND (“upper secondary school” OR “high school” OR “university” OR “college”)	316	Not considered
Google Scholar	S1	(“serious game” OR “game based learning") AND Mathematics AND education AND engagement AND motivation AND ("math anxiety" OR "Mathematics anxiety") AND ("upper secondary school" OR "high school" OR "university" OR "college")	136	99
Scopus	S2	TITLE-ABS-KEY (("serious game" OR "game-based learning") AND (Mathematics OR math) AND ("math anxiety" OR "Mathematics anxiety") AND ("secondary school" OR "high school" OR "upper secondary" OR "post-secondary" OR university OR college OR "higher education"))	4	4
Scopus	S3	TITLE-ABS-KEY (("serious game" OR "game-based learning") AND (Mathematics OR math) AND (motivation) AND ("secondary school" OR "high school" OR "upper secondary" OR "post-secondary" OR university OR college OR "higher education"))	66	51
Scopus	S4	TITLE-ABS-KEY (("serious game" OR "game-based learning") AND (Mathematics OR math) AND (engagement) AND ("secondary school" OR "high school" OR "upper secondary" OR "post-secondary" OR university OR college OR "higher education"))	47	37
Scopus	S5	TITLE-ABS-KEY (("serious game" OR "game-based learning") AND (Mathematics OR math) AND (usability) AND ("secondary school" OR "high school" OR "upper secondary" OR "post-secondary" OR university OR college OR "higher education"))	20	15

Table 1 - Strings used for database searches.

The only remaining dimension to be examined was the impact of GBL on learning outcomes. However, when the keywords “learning” or “learning outcome” were included, the number of results in both databases was again excessive. Therefore, the evaluation began with the papers identified through the previous searches, which could also provide insights into learning outcomes. A more detailed study of the learning outcomes dimension was deferred and would only be considered if the results from the previously identified papers were unsatisfactory. All the documents obtained were consolidated into an Excel matrix containing their title, author, year of publication and access status. The matrix was then refined to evaluate the quality and relevance of the studies to the objectives of this thesis. The Excel matrix also included other papers from the bibliography of those that were chosen for their relevance, and these papers were also examined. The selection of relevant papers was conducted after researching them online to determine their accessibility. Before screening phase, duplicates and papers that were unavailable online, not available as open access were removed and excluded (N=37 for Scholar and N=30 for Scopus, see Fig.3). Studies that could not be found online in their entirety, or to which we did not have access even after contacting the author, were excluded.
To ensure the retrieved studies are relevant to the scope of the review, all the following inclusion criteria must be satisfied:

- scientific article published in English;
- articles concern the use of game-based learning, serious games, or gamification in educational settings;
- each study is related to Mathematics learning and includes participants from upper secondary school to higher education;
- the articles address the impact on motivation, engagement, math anxiety, or learning outcomes, or they deal with the evaluation of serious games and game-based learning tools, especially in terms of usability.

In the screening phase, the titles and abstracts of the studies were reviewed, and those that were irrelevant or inconsistent with the purpose of the literature were removed. The screening phase yielded 31 results from Google Scholar and 34 results from Scopus for eligibility (see Fig.3). The main reasons for exclusion in screening phase were:

- papers focusing on primary or lower secondary school grades;
- papers not aligned with the aims of the literature review, e.g. topics not strictly related to game-based learning or gamification, or not specifically related to Mathematics;
- results obtained from the different Scopus searches (S2, S3, S4 and S5) included duplicates, for example, some papers appeared in both the S2 and S3 results;
- papers written in languages other than English.

Although the term “gamification” was not included in the search terms, as it already yielded many results, studies addressing gamification were also considered, given its close relationship with GBL and the shared characteristics of the two concepts.
Finally, the last step was the eligibility phase, which involved reviewing the available articles by reading them in full and applying the mentioned inclusion criteria. Based on these criteria, 5 articles were included from Google Scholar and 23 from Scopus (Fig. 3). Eight more papers were added to these 28 results: some were identified through the bibliographies of the selected studies, while others were already known on the topic. The PRISMA process was also followed for these latest articles added.

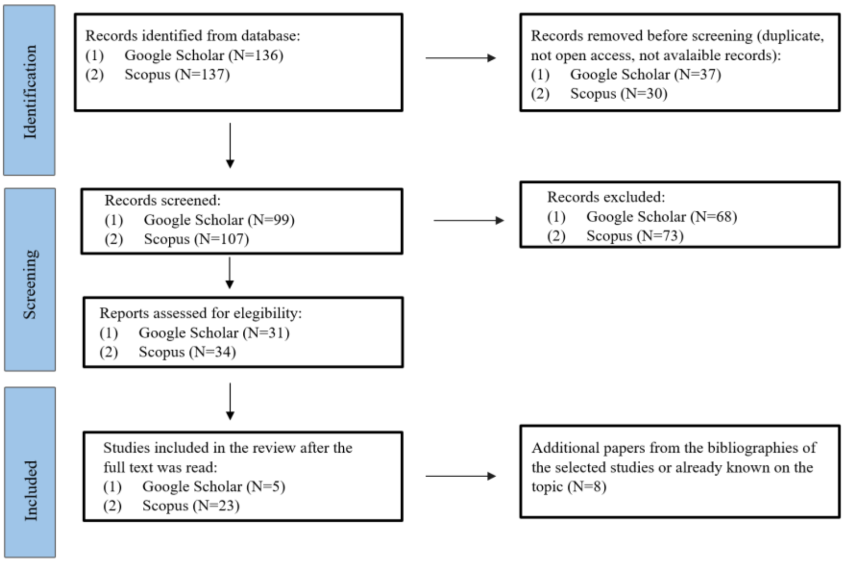

Figure 3 - Flow diagram of the studies selection process.

The reasons for exclusion in the eligibility phase are similar to those identified during the screening one. Initially, these reasons were not evident from reading the papers and their abstracts. Papers were excluded in eligibility phase for the following reasons:

- papers focusing on primary or lower secondary schools;
- papers not aligned with the aims of the literature review, e.g. topics not strictly related to game-based learning or gamification, or not specifically related to Mathematics;
- the results are unclear;
- papers do not focus specifically on GBL, but rather on ICT in general.

In some cases, the papers did not focus exclusively on upper secondary schools, but rather on K–12 students in general. These papers were included, with particular attention given to the results concerning upper secondary school students. No time restrictions were applied to the publications considered in the database search. Articles included in the sample were published from 2010 to July 2025. However, as highlighted in previous studies (Hussein et al., 2022; Aragonez et al., 2021), most of the selected papers (37 out of 45) were published between 2021 and 2025. This was not because there was an intentional focus on more recent works, but because the pool of papers from the last five years was considerably larger than those from earlier periods. This trend indicates a rapid increase in the number of publications, reflecting growing interest in the topic (Figure 4).

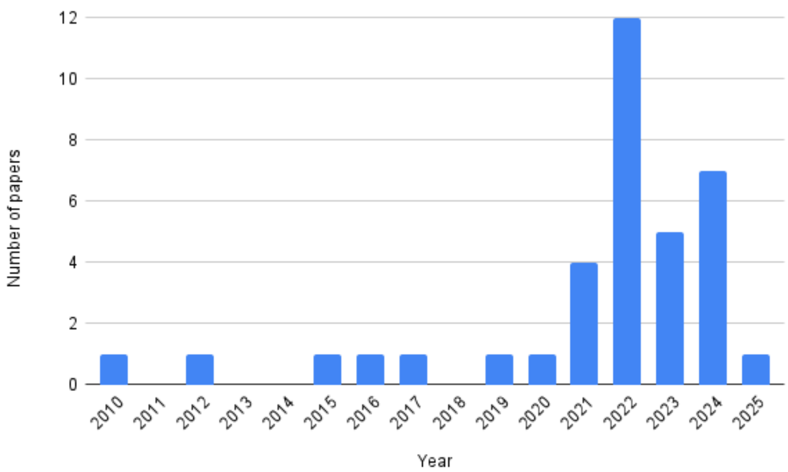

Figure 4 - Distribution of selected papers by publication year.

After the identification of the 36 relevant studies, an analysis of them was carried out, focusing on the results and content of the identified papers. The following table summarizes the key findings of the 36 papers selected for the literature review, with a focus on those related to the objectives of this thesis. The table also includes five columns (RQ0_A, RQ0_B, RQ0_C, RQ0_D and RQ0_E), which indicate how aligned each paper is with the research questions, thereby justifying its inclusion in the review.

Year	Authors	Target	Relevant findings	RQ0 _A	RQ0 _B	RQ0 _C	RQ0 _D	RQ0 _E
2021	Fadda et al.	K-12 students	Most of the selected studies focus on primary and lower secondary education. No significant differences were found in the effect on student motivation at different school levels. This study is a starting point, as the findings suggest that digital games could be an effective teaching tool for motivating students compared with conventional instructional practices. High-quality studies that use reliable and validated measures are needed to assess the long-term impact of digital games. Adequate teacher training is important to maximise the benefits of using digital games to support educational outcomes.	✓			✓	✓
2021	Chen et al.	K-12 students	The review highlights that most studies in GBL in Mathematics and science disciplines have focused on primary school (40%). Over the past decade, considerable attention has been given to the collaborative learning and problem-solving skills of learners in the context of GBL. Although the results are not always clearly categorised by school level, it has been shown that games can effectively enhance the collaboration, problem-solving skills, attitudes, motivation and engagement in primary and secondary school students.	✓	✓			✓
2022	Hussein et al.	K-12 students	The number of studies focusing on DGBL increased significantly between 2014 and 2021, indicating a growing interest in using digital games in K–12 Mathematics education. The review reveals that few studies have been conducted at the upper secondary level, with only 8 out of 43 focusing on this stage of education. Several reasons for the limited use of DGBL in high schools are identified. Firstly, games are generally perceived as joyful and pleasing, particularly for young learners. Secondly, teachers often encounter barriers when attempting to use DGBL with older students, such as a mismatch between the knowledge and skills embedded in the games and those outlined in the official curriculum, a lack of ICT skills, and negative attitudes towards the use of games in formal educational settings. Furthermore, primary school educators tend to have greater flexibility in their teaching methods and are therefore more inclined to experiment with serious games; high school teachers, on the other hand, may feel more constrained. Finally, younger students are generally easier to engage and less demanding, whereas high school students tend to expect more sophisticated and complex DGBL applications. Regarding high school settings specifically, teachers recognise the potential benefits of DGBL in enhancing student learning and engagement. However, one study notes that the actual implementation of DGBL at the high school level remains limited compared to its broader use in primary education. The review highlights that few studies provide detailed background information on the design features or the theoretical frameworks underlying the DGBL applications used in educational interventions.		✓			✓
2022	Pan et al.	K-12 students	Most of the studies (35 out of 43) examined the effects of maths learning games on primary and lower secondary school students, while five focused on upper secondary school students. One possible reason for this is that Mathematics taught at lower grade levels is generally easier to integrate into games. The most frequently studied variables are knowledge and skill acquisition, engagement, and students' motivation, attitudes, and self-efficacy towards maths learning. Findings suggest positive effects of gameplay on students' maths learning performance, even though there is insufficient evidence to suggest benefits for metacognitive skills and knowledge transfer. Gameplay has the most significant positive	✓		✓		✓

			impact on self-efficacy, though its effect on reducing math anxiety is less significant. In terms of game design, evidence suggests that play mode and feedback components positively influence learning outcomes, whereas aesthetic features have a comparatively smaller effect. The most adopted model was the individual model. In terms of how games were used, they were primarily employed to consolidate previously learnt content or to support the application of acquired knowledge. More than half of the studies focused on arithmetic topics, while geometry was the least explored mathematical domain. This suggests that game developers should invest more effort in designing DGBL applications for underrepresented topics, especially those taught at higher grade levels.					
2025	Gusti	Secondary school students	17 of the 22 studies reported improvements in students' conceptual understanding of algebra, particularly regarding topics such as algebraic expressions, linear equations, and factorisation. 10 studies also reveal improvements in problem-solving ability. GBL was found to consistently enhance student participation and motivation. The potential of GBL to improve algebra instruction in Indonesia is especially highlighted, particularly when aligned with curricular goals and supported by solid pedagogical frameworks. The effectiveness of GBL depends not only on the design of the game itself, but also on how well it is implemented and integrated into lesson plans.	✓	✓			✓
2022	Maisey et al.	Participants aged 18-25	The study compares the effects of game- and non-game-based versions of a number line estimation task on participants aged 18–25. The results showed no significant difference in performance between the two versions, and the game version used in the current study did not influence participants' anxiety levels. Therefore, these results did not support the hypothesis that the game-based version of the task would reduce reported math anxiety more than the non-game-based version. It should be noted that levels of math anxiety were generally low in the sample of students involved. Furthermore, controlling for the influence of gender did not alter the results, but it is worth noting that the three participants who scored highly for math anxiety were all female.			✓		✓
2022	Zabala-Vargas et al.	University students	This study investigates the relationship between the use of gamification/GBL strategies and student motivation in a course with a high failure rate. The ARCS model by Keller (2010) was applied to evaluate its effectiveness in terms of motivation. The findings suggest that GBL has a positive effect on motivation. Furthermore, students demonstrate greater attention and engagement, which facilitates the learning process. Participants perceive the game-based approach as an innovative strategy that improves the learning experience. The use of narrative elements was particularly appreciated, as it encouraged both competition and collaboration.	✓	✓			
2024	Ilić et al.	K-12 and university students	It focuses on the use of digital games in STEM education (not just in Mathematics). Most studies (75%) reports that DGBL increases student motivation at both K-12 and university levels. Motivational impact varies by educational level: K-12 students show significantly higher motivation (67.8%) compared to university students (32.2%). Motivation also depends on lesson content—students are more motivated when engaging with simpler, more familiar topics than with abstract or unfamiliar ones. The influence of prior knowledge on motivation is mixed. While students with prior knowledge of a STEM topic tended to be more motivated, several studies argue that such background knowledge is not essential. K-12 students are generally more enthusiastic about learning through digital games than university students, and older K-12 students show the strongest preference for this method. The review also found that motivation, academic achievement, and engagement are the most frequently studied variables in DGBL research, with academic motivation being the most explored.	✓				

2024	Fante et al.	Secondary school students	It focuses on the use of digital games in STEM education (not just in Mathematics). Recent studies further confirm that integrating GBL with curricular goals enhances both engagement and learning outcomes. Game-based environments help make complex STEM content more accessible and enjoyable through interactive and gamified elements and also stimulate problem-solving skills. GBL's potential to foster holistic scientific understanding and creative problem-solving.		✓			✓
2022	Sánchez-Ruiz et al.	University students	A digital escape room was designed and tested as an educational tool, achieving highly positive results. Students found the activity enjoyable and beneficial, and reported increased motivation, interest and engagement. It contributed to their learning and helped them to view Mathematics more positively. Post-test results and comparisons with students from previous academic years revealed a notable rise in the proportion of students achieving passing grades, alongside higher average marks.	✓	✓			✓
2021	Zabala-Vargas et al.	University students	Gamification and GBL demonstrate significant potential for improving educational processes, offering highly encouraging results in terms of motivation. The ARCS model by Keller (2010) was applied to evaluate its effectiveness in terms of motivation. Early empirical evidence suggests that implementing educational games is linked to a reduction in dropout rates. GBL activities, which are often problem-solving focused, also support the development of the scaffolding needed to tackle mathematical problems. Students who engage with game-based strategies tend to exhibit high levels of motivation. Successful implementation requires careful planning, clearly defined rules, high-quality materials and effective teacher feedback. Additionally, activities with engaging narratives that connect to students' interests are crucial for designing impactful pedagogical strategies.	✓	✓			✓
2022	Malvasi et al.	Secondary school students and their teachers	The authors developed a questionnaire to assess the level of gamification in Italian secondary schools. The results showed moderate internal reliability. Despite teachers having not received training in gamification, we observed "unconscious gamification" in high schools. The findings suggest the need to reconsider teaching methods in Mathematics. For Italian secondary school teachers, gamification deserves more attention due to the benefits it brings, particularly for subjects that intimidate students and require greater motivation.					✓
2022	Crnković et al.	Upper secondary school students	This case study was conducted with students who had no prior knowledge of cryptographic concepts. The results confirm the effectiveness of the proposed model. Students demonstrated a measurable increase in familiarity with the terminology and knowledge of the complex new topic, as well as the ability to apply their knowledge. The survey results also show a very positive response in terms of motivation to learn new concepts.	✓				✓
2021	Jiménez-Hernández	University students	A serious game was developed to support learning about tree traversals in discrete Mathematics. A pre- and post-test study involving a control group demonstrated that students who used the game performed better in the final exam than those who practised using traditional paper-and-pencil methods. The ARCS model by Keller (2010) was applied to evaluate its effectiveness in terms of motivation. The results indicated that the game significantly increased students' motivation compared to traditional reinforcement methods.	✓				✓
2023	Pais et al.	University students	An escape room activity was tested for teaching Mathematics to Life Science students enrolled in a Mathematical Analysis course. Teachers observed that students found the experience highly enjoyable and engaging. The escape room provided a more positive and interesting approach to Mathematics, with most students reporting that it enhanced both their learning process—by making it more interactive, challenging, and stimulating—and their motivation to learn math.	✓	✓			✓

2024	Pais & Hall	University students	This study investigated whether integrating game-based learning strategies could enhance student motivation and engagement when studying Mathematics as part of the Tourism programme. Despite the small sample size, the results demonstrate that incorporating game-based learning activities in maths classes enhances engagement and motivation by making learning enjoyable.	✓	✓			
2023	Chen et al.	Upper secondary school students	The innovative aspect of this study lies in combining multi-representational scaffolding with gamification elements. The ARCS model by Keller (2010) was used to assess motivational impact. While no significant differences were found in learning outcomes - either within the experimental group before and after the intervention or between the experimental and control groups - the study revealed that students in the experimental group showed significantly higher motivation after the gamified activity. Additionally, students demonstrated strong focus and engagement, along with lower levels of anxiety.	✓	✓	✓		✓
2023	Lee et al.	University students	The study examined a gamified Mathematics course designed for liberal arts students with weaker math backgrounds, evaluating its effectiveness in improving engagement, motivation, and learning outcomes. The ADDIE model was used to design the gamified elements. Given the limited research on gamification in Mathematics education—especially at the higher education level—the authors also reviewed findings from other educational stages to assess broader applicability. The missions and badges were considered the most effective, while the leaderboards and progress bars were rated lower. Results showed that students' prior math skills and self-motivation influenced their perception of the course: those with stronger math performance responded more positively, while lower-achieving students benefited less. The study suggests that more complex motivational strategies are needed to support struggling college students better.	✓	✓			
2023	Magat	University teachers and IT professionals	A gamified mobile app for learning Probability and Statistics was tested with teachers of the subject, Mathematics experts, and IT professionals. Math teachers and experts evaluated the app's content, learning objects, and educational value, while IT professionals assessed its engagement, functionality, aesthetics, and informational quality. The results were largely positive, especially regarding the learning objects, indicating that the app was seen as effective and engaging in delivering its educational goals. As content quality, engagement, and learning assessments affirm that the developed gamified mobile app can be an effective instructional material for distance learning, particularly during asynchronous learning modalities. However, the evaluation also highlighted areas for improvement, particularly in usability, including the user interface, user experience design, user control, interaction flexibility, data quality, reliability, and privacy.				✓	✓
2017	Gil-Doménech	University students	To overcome the negative attitude of non-STEM students towards Mathematics-related courses, GBL has been used to improve the confidence and interest in learning of Business Administration students. After implementing GBL, students gradually improved their academic grades. The positive results obtained were not only in the acquisition of knowledge regarding derivatives and integrals, but also in terms of interest and motivation expressed by students. The study remarks that not all teachers have the appropriate skills for designing gamified activities. Therefore, it is necessary for universities to take a leading role in this process by promoting courses and training programmes that focus on putting innovative methodologies into practice.	✓				✓
2024	Christopoulos et al.	Upper secondary school students	The aim of the study is to verify whether immersion is associated with an improvement in mathematical skills when using a GBL intervention. The immersion and engagement items of the User Experience in Immersive Virtual Environment Scale (Tcha-Tokey et al., 2016) was used at this scope. An exploratory factor analysis of the engagement questionnaire showed that the scales had good internal		✓			✓

			consistency. The results confirmed that mathematical skills improved through GBL. However, immersion did not predict improvements in test scores, suggesting that other factors explain the effectiveness of GBL. Additionally, no correlation was found between improvement and initial (pre-test) scores, indicating that GBL benefits students regardless of their starting level. It is hypothesised that GBL enhances learning by providing students with additional learning opportunities. Finally, no gender differences were observed in the outcomes.					
2024	Nautiyal et al.	K-12 students	The authors propose a five-step framework to help educators integrate problem-solving, game-based learning sessions into their classrooms. The framework incorporates the design elements and key features of board games, emphasising ease of design and use, and requiring no specialist skills from teachers. To assess students' perceptions of the GBL sessions, a set of ten questions was used, including seven Likert-scale items. Additionally, the Playful Learning Observation Tool developed by Algayres et al. (2022), was employed to monitor student engagement and observe classroom dynamics during the GBL sessions. Most students liked the GBL activities, and adapting the games to their preferences could make them even better, especially in maths. Teachers also reported high levels of student engagement during gameplay. The games captured the attention of all students, including those with lower academic performance in Mathematics.		✓			
2022	Di Blasi et al.	Upper secondary school students	This study focuses on the use of Kahoot! in online learning to engage students in Physics and Mathematics, covering topics that are not usually taught in schools or providing a university-level perspective on familiar topics. Several benefits emerged: Kahoot! helped students to maintain their focus, and the presence of a final ranking system boosted their attention, motivation and participation. The study highlights gamification can be an effective learning tool for teachers to monitor student progress, for both remote and traditional in-person learning.	✓	✓			
2016	Crocco et al.	University students	The study provided professors with training in GBL and then implemented game-based lessons in their own subject areas. These lessons were analysed using Bloom's Taxonomy, which covers cognitive processes ranging from simple to complex. There were no significant differences in overall learning outcomes between the control group and the group that experienced GBL. However, a strong positive linear correlation was observed in both the experimental and control groups between students' enjoyment of the lesson or game and their performance on deep learning quiz questions, but not for surface learning. This correlation did not apply to surface learning. Therefore, enhancing student enjoyment through games can be an effective way to support deeper learning.		✓			✓
2019	Ting et al.	University students	This is a preliminary exploration of active learning through problem-based, collaborative games in a 13-week , first-year "Applied Mathematics" course. Topics covered included functions, limits, continuity and differentiation. The study involved an experimental and a control group. The study found a positive correlation between students' perceptions of their active engagement with their academic performance. Specifically, students' perceived level of active engagement was a significant predictor of their academic achievement.		✓			✓
2024	Ardani et al.	Secondary school students and their teachers, pre-service teachers	This study investigates the practicality of using edutainment games in a classroom setting. Edutainment involves integrating video games into teaching and learning. After the game-based media had been implemented, users were asked to evaluate its effectiveness. This focused on two key aspects -usability and ease of use -based on an adaptation of the Technology Acceptance Model. The findings suggest that the media are considered highly practical and flexible, with the potential to enhance student performance.				✓	

2022	Estrada-Molina et al.	Secondary and higher education	This review aims to identify the most applied international standards and usability norms in educational technology for teaching Mathematics at secondary and higher education levels. The analysis focuses on two specific contexts: digital educational resources and virtual learning environments and GBL. The usability criteria used to evaluate GBL tools are primarily based on international standards such as ISO/IEC 9126-1:2004, ISO 9241-11:2018, which ensure the integration of efficiency, effectiveness and ease of use. Furthermore, the importance of using standardised instruments such as the System Usability Scale and the Game User Experience Satisfaction Scale is emphasised, as is the recommendation to take a user-centred approach to evaluation.				✓	
2022	Wiboonsin & Kasemsukpipat	Upper secondary school students	This study focuses on developing a board game to introduce concepts related to wealth creation for enhancing the financial literacy of high school students. Particular attention is given to the design of the game and the development of game mechanics that align with core financial concepts. The game was tested to evaluate both gameplay functionality and content accuracy. Students were invited to play the game and provide feedback to assess how well it communicated its intended content. The results showed that students could demonstrate their financial knowledge through gameplay. The Serious Game Usability Evaluator (SeGUE) method was applied to evaluate the game.				✓	✓
2023	Dondio et al.	K-12 and university students	This study presents a meta-analysis of 16 experimental studies examining the effectiveness of game-based interventions in alleviating math anxiety in students. Most of the studies were conducted in primary schools, with one in a secondary school and one at university level. Participants were randomly selected, making the results more generalisable. The findings showed that non-digital collaborative games were more effective in reducing math anxiety than digital games, which had a smaller impact. Longer interventions yielded better outcomes. Game-based interventions were particularly effective when they encouraged social and collaborative interactions. As all the digital games analysed were single player, it remains unclear whether the collaborative nature or the non-digital format itself was the key factor in reducing math anxiety.			✓		
2024	Ayyasy & Asrul	University students	This study presents the results of an experiment in which participants played a game called "Math Bingo". A questionnaire and descriptive learning outcome test questions was used to evaluate the game's impact. The questionnaire was tested for validity and reliability, and it was found to be both. Math Bingo increases student motivation and engagement and provides immediate feedback and social support that helped reduce anxiety and improve learning outcomes. Math anxiety decreased significantly from the initial to the final questionnaire. Also, the results achieved on the learning outcomes were remarkable, with significant progress being made from the pre-test to the post-test.	✓	✓	✓	✓	✓
2022	Wijaya et al.	Secondary school teachers	A questionnaire based on the Technology Acceptance Model was developed to predict and identify the factors influencing Mathematics teachers' behavioural intentions and actual use of microgames. The aim was also to determine the most significant positive factor affecting these intentions and behaviours. The model demonstrated good content validity, and its reliability was considered satisfactory. The results revealed that perceived ease of use had the strongest positive impact on teachers' behavioural intention to use microgames. This suggests that developers should prioritise ensuring that microgames are user-friendly for teaching Mathematics. However, perceived usefulness did not have a significant effect on teachers' intention to use these games in Mathematics education.				✓	
2010	Kebritchi et al.	Upper secondary school students	This study examined the effect that playing a computer game had on students' mathematical achievement and motivation in a varied urban secondary school setting. According to most interviewed teachers, as well as pre- and post-test results, there was an increase in maths skills and learning outcomes.	✓				✓

		and their teachers	Teachers attributed this to the experiential nature of the games and their role as alternative teaching tools, which gave students an opportunity to practise maths and improve the relation towards Mathematics. Students appreciated the combination of fun and learning, and the challenges offered by the games. However, the study found no significant effect on motivation, and prior maths knowledge or computer skills did not significantly influence outcomes.					
2012	Afari et al.	Upper secondary school students	The study examined whether incorporating games into college-level Mathematics classes improves students' perceptions of learning and their attitudes towards Mathematics. To this end, the 'What Is Happening In This Class?' (WIHIC) questionnaire was adapted for the purposes of the research. The scale demonstrated factorial validity and internal consistency reliability. The findings provide a solid foundation for using educational games to enhance students' attitudes towards Mathematics. Improvements were particularly evident in terms of student involvement and enjoyment. These results suggest that integrating games into Mathematics education could improve students' attitudes towards the subject.				✓	✓
2020	Albano et al.	Upper secondary school students	The study examined how digital games can support mathematical learning. The authors proposed two key strategies to ensure games are engaging and aligned with learning goals: designing games that resonate with youth gaming culture, and creating game experiences relevant to real-world problem solving. A final questionnaire revealed positive learning outcomes: students reported that the game stimulated their creativity and offered a new way of thinking about Mathematics, strengthening their confidence in problem solving. The case study showed that the game encouraged anticipatory thinking, making conjectures, developing strategies. This helped students to identify mathematical properties with increasing levels of formality. Changes were also observed in the language used by the students and in their approach to problem solving within the game, shifting from imprecise, practical descriptions to more general, abstract formulations.		✓			✓
2020	Bakri et al.	University students	An innovative board game called Graph Puzzle was developed to help students overcome their difficulties in sketching the graphs of mathematical functions using derivatives. Designed to improve students' understanding of the interconnected algebraic, symbolic and graphical representations of functions in the context of derivatives, the game supports the development of both procedural and conceptual knowledge.To evaluate its effectiveness, students completed pre- and post-tests before and after a learning session involving the board game. The treatment group achieved significantly higher post-test scores than the control group for questions on polynomial functions. However, both groups performed similarly on rational function questions, showing no statistically significant difference. These results suggest that Graph Puzzle could be an effective tool for overcoming teaching and learning challenges related to graph sketching in calculus.					✓
2015	Verkijika & De Wet	Participants aged 9–16	This study investigated whether using a brain-computer interface could effectively reduce students' math anxiety. The results revealed a significant positive correlation between age and overall math anxiety scores, suggesting that math anxiety tends to increase with age. This trend may be explained by the fact that older students are generally exposed to more complex mathematical problems than younger students. In contrast, there was no significant relationship between math anxiety and gender or grade level, which aligns with previous research findings. Furthermore, the study found a significant negative correlation between math anxiety and maths performance, confirming existing evidence that higher levels of math anxiety are associated with lower achievement in maths.			✓		

Table 2 – Summary of articles included in the literature review.

It is worth noting that the studies identified through Google Scholar focus on the entire K–12 education continuum and they provide interesting statistics on GBL and gamification usage in Mathematics education. For example, they reveal that it is easier to find papers on GBL in the context of primary education than in higher school grades. Hussein et al. (2022) attempted to explain the limited use of DGBL in secondary schools by identifying several contributing factors. First, games are often considered more appropriate for younger learners as they are associated with enjoyment and play. Moreover, teachers often encounter obstacles when integrating DGBL with older students, such as discrepancies with the official curriculum, limited ICT proficiency, and negative attitudes towards games in formal education. Infact, younger students are generally easier to engage, whereas high school students expect more advanced and complex DGBL applications. Furthermore, primary school teachers tend to have more freedom to experiment with innovative methods than high school teachers. Another notable finding is the significant rise in publications on this topic in recent years, especially since 2020. This increase may be attributed to the wider uptake of online learning during the pandemic. However, it should be noted that in these studies there is not always a clear link between the results and the students' grade level. These papers were included among the selected articles for the literature review because they provided useful statistics on the state of the art in the field of GBL.
The review also revealed interesting findings that extend beyond the specific research questions yet are still valuable for this thesis. For instance, several studies emphasised the importance of teacher training in using gamified approaches to enhance the effectiveness of these innovative learning methodologies (Fadda et al., 2021; Malvasi et al., 2022; Gil-Doménech, 2017; Crocco et al., 2016). The study by Malvasi et al. (2022) also revealed that, despite not having received training in gamification, teachers showed an "unconscious gamification" in high schools.
Furthermore, the literature review highlights emerging models and design characteristics in the development of educational games. It emerged that the effectiveness of GBL depends not only on good game design -for example, clearly defined rules -but also on how well it is implemented and integrated into lesson plans (Gusti, 2025; Zabala-Vargas et al., 2021). Pan et al. (2022) found that play mode and feedback have a positive influence on learning outcomes, while aesthetics matter less. It also emerged that most games feature single-player modes, and that they are used to consolidate previously learnt content or support the application of new content. Research has mainly focused on arithmetic and indicates that there is a need for more DGBL applications at higher levels.
In the context of this thesis, it is interesting to note that the literature review revealed examples of GBL being used in mathematics classes aimed at non-mathematics students.
The subsequent aim is to answer the research questions that have emerged from the literature review.

(RQ0_A) How does the use of game-based learning and serious games in mathematics education affect student motivation at upper secondary and university level?
Motivation is one of the most frequently studied variables in the literature on GBL in Mathematics, alongside engagement and learning outcomes (Pan et al., 2022; Ilić et al., 2024). A review of the selected articles reveals that the ARCS model, developed by Keller (2010), is a frequently used framework for assessing the effectiveness of GBL in terms of motivation (Zabala-Vargas et al., 2021; Zabala-Vargas et al., 2022; Jiménez-Hernández, 2021). Most findings regarding the effectiveness of GBL and gamification in terms of motivation in Mathematics are positive. The study of Fadda et al. (2021) serves as a starting point, as its results suggest that digital games are an effective teaching tool for motivating students to learn Mathematics compared to conventional instructional practices. The majority of the reviewed studies (14 out of 17) that examined the impact of GBL and gamification on motivation found that they can effectively enhance student motivation to learn Mathematics (Chen et al., 2021; Gusti, 2025; Zabala-Vargas et al., 2022; Ilić et al., 2024; Sánchez-Ruiz et al., 2022; Zabala-Vargas et al., 2021; Crnković et al., 2022; Jiménez-Hernández, 2021; Pais et al., 2023; Pais & Hall, 2024; Chen et al., 2023; Gil-Doménech, 2017; Di Blasi et al., 2022; Ayyasy & Asrul, 2024). Conversely, two studies found no significant effect on motivation (Lee et al., 2023; Kebritchi et al., 2010). Lee (2023) suggested that more complex motivational strategies are required to support struggling college students better, highlighting that motivation may depend strongly on prior mathematical skills and levels of self-motivation. This last result aligns with those of the study by Floris et al. (2023). Furthermore, Ilić et al. (2024) showed that motivation in STEM disciplines also depends on lesson content, with students reporting greater

motivation when engaging with simpler, more familiar topics compared to abstract or unfamiliar ones. In addition, Fadda et al. (2021) found no significant differences in the effect of GBL on student motivation in Mathematics across school levels. In contrast, Ilić et al. (2024) reported that the motivational impact varies by educational level.

(RQ0_B) To what extent do serious games and game-based learning enhance student engagement in learning Mathematics at upper secondary and university levels?

As highlighted in the study by Pan et al. (2022), engagement is one of the variables most frequently examined in the context of GBL and gamification. Of the 36 selected articles for this literature review, 18 specifically address this topic. Most of these studies report positive effects of GBL on student engagement (Chen et al., 2021; Gusti, 2025; Fante et al., 2024; Sánchez-Ruiz et al., 2022; Pais et al., 2023; Pais & Hall, 2024; Chen et al., 2023; Di Blasi et al., 2022; Ayyasy & Asrul, 2024; Zabala-Vargas et al., 2022). In the study by Nautiyal et al. (2024), the engagement of students in gameplay was positively evaluated by their teachers. Another important finding reported by Zabala-Vargas et al. (2021) is that students who engage with game-based strategies tend to display higher levels of motivation.

Regarding the most appreciated and effective gamification strategies for student engagement, Lee et al. (2023) found that missions and badges were considered the most effective elements, while leaderboards and progress bars were rated less favourably. Contrary, Di Blasi et al. (2022) emphasize that the presence of a final ranking system increases students' attention and participation in Physics and Mathematics lessons.

Christopoulos et al. (2024) investigated whether immersion, a construct closely related to engagement, was associated with improved mathematical skills in a GBL intervention. Their results show that immersion does not predict improvements in test scores, suggesting that additional factors may contribute to the effectiveness of GBL in improving Mathematics learning outcomes. Conversely, Crocco et al. (2016) reported a strong positive linear correlation between students' enjoyment of the game and their performance in deep learning, even if not surface learning. This suggests that engaging students in the learning process can lead to better outcomes for higher-order learning. Similar results were found by Ting et al. (2019), who showed that students' perceived level of active engagement is a significant predictor of academic achievement. With regard to balancing engagement and learning outcomes in educational game design, Albano et al. (2020) highlighted two key strategies: (i) designing games that resonate with youth gaming culture and (ii) creating game experiences that are relevant to real-world problem solving.

Unlike what was observed for motivation, no specific shared framework has been identified for measuring engagement in the selected articles.

(RQ0_C) What impact do serious games and game-based learning have on students' math anxiety in upper secondary and university mathematics education?

Of the 36 articles reviewed, 6 specifically deal with math anxiety and results are rather heterogeneous. In their systematic review involving K-12 students, Pan et al. (2022) reported that gameplay has only a minor positive impact on reducing math anxiety compared to other factors, such as self-efficacy. In their study, Maisey et al. found that participants aged 18-25 with low initial levels of math anxiety did not experience a reduction when using the game-based version of the task compared to the non-game-based version. However, other studies report a reduction in math anxiety as a result of GBL among upper secondary school students (Chen et al., 2023; Dondio et al. 2023) and university students (Ayyasy & Asrul, 2024).

It is worth noting that no significant relationship between math anxiety and gender was identified in the studies by Maisey et al. (2022) and Verkijika and De Wet (2015). The latter study also found that math anxiety tends to increase with age, which may be explained by the fact that older students are generally exposed to more complex mathematical problems than younger students. Furthermore, their study revealed a significant negative correlation between math anxiety and performance, corroborating existing hypothesis that higher anxiety levels are associated with lower achievement.

As previously noted for engagement, no shared and validated instrument for measuring math anxiety has emerged.

(RQ0_D) How are serious games or any other game-based tools for Mathematics learning evaluated in the literature in terms of usability?
Fadda et al. (2021) emphasise the need for high-quality studies using reliable and validated measures to assess the long-term impact of digital games. Of the 36 selected studies, 8 reported on the evaluation of game-based tools in the literature. In some cases (Magat, 2023; Ayyasy & Asrul, 2025), the quality and impact of games on students were assessed using self-developed questionnaires that were then analysed for validity and reliability. In other studies (e.g. Ardani et al., 2024; Wijaya et al., 2022; Afari et al., 2013), users were asked to evaluate the effectiveness of the games using instruments adapted from existing models such as the Technology Acceptance Model. Further evaluation standards frequently used are reported in Estrada-Molina et al. (2022), including ISO/IEC 9126-1:2004, ISO 9241-11:2018, the System Usability Scale and the Game User Experience Satisfaction Scale. This study also highlights the recommendation to adopt a user-centred approach to evaluation. Wiboonsin and Kasemsukpipat (2022) applied another tool, the Serious Game Usability Evaluator method, to assess the game from the students' perspective.

(RQ0_E) What evidence is there that serious games and game-based learning approaches improve learning outcomes in Mathematics at upper secondary and university level?
Of the 36 selected studies, 25 provided evidence of the impact of GBL and gamification on Mathematics learning outcomes. As enough relevant papers on the topic had been identified, it was deemed unnecessary to conduct any further database searches. Most of these studies highlights a positive effect on maths learning. Firstly, it has been emphasised that games can effectively improve problem-solving abilities (Chen et al., 2021; Gusti, 2025). Beyond problem solving, Albano et al. (2020) identified additional mathematical benefits, including anticipatory thinking, making conjectures, developing strategies and using more precise language. Teachers in both high school and university settings recognised the potential benefits of games - both digital and non-digital -in supporting student learning (Hussein et al., 2022; Fante et al., 2024; Pais et al., 2023; Afari et al., 2013). This use of games was often associated with improved learning outcomes (Pan et al., 2022; Gusti, 2025; Crnković et al., 2022; Jiménez-Hernández, 2021; Magat, 2023; Gil-Doménech, 2017; Christopoulos et al., 2024; Wisconsin & Kasemsukpipat, 2022; Ayyasy & Asrul, 2024; Kebritchi et al., 2010). Evidence of learning outcomes is often based on the difference between pre- and post-test results (Sánchez-Ruiz et al., 2022; Jiménez-Hernández, 2021; Ayyasy & Asrul, 2024; Kebritchi et al., 2010). Zabala-Vargas et al. (2021) also reported that there is early empirical evidence linking the use of educational games to a reduction in student dropout rates. The concept of an escape room, a gaming experience where participants must solve challenges to escape from a locked room, has gained popularity in educational contexts. Escape rooms are also widely used in Mathematics teaching and have been shown to improve learning outcomes (Sánchez-Ruiz et al., 2022; Pais et al., 2023). Furthermore, Ting et al. (2019) found that students' perceived level of active engagement was a significant predictor of academic achievement. However, some studies have found no significant difference in maths performance between game-based approaches and traditional methods used in control groups (Maisey et al., 2022; Chen et al., 2023; Crocco et al., 2016). Another important finding is that adequate teacher training is essential to maximise the benefits of digital games in supporting educational outcomes (Fadda et al., 2021; Malvasi et al., 2022; Gil-Doménech, 2017; Crocco et al., 2016).
Among the limitations of this review, it is worth noting that the search was restricted to only two databases, Google Scholar and Scopus. Moreover, the analysis focused on aspects related to the research questions, although results also emerged beyond these aspects. Additionally, to limit the number of results retrieved from Google Scholar, it was necessary to cross the dimensions of the five research questions, so only papers addressing all five dimensions were returned from the database and then considered.

3. Context of the Research

Chapter 3 provides a detailed overview of the research context, clearly defining the setting in which the study was conducted. It also outlines the research objectives and formally articulates the research questions that guided the different phases of the study, which are described in subsequent chapters. In particular, paragraph 3.1 provides a background to the research, situating it within a post-pandemic educational context and addressing the difficulties related to mathematics learning that have emerged from various international studies.

3.1 Research Background

As outlined in the section on the theoretical framework, the adoption of game-based learning and gamification in secondary schools and higher education institutions is a response to various educational challenges. These include the need to boost student engagement, enhance academic success rates and encourage more sustained and meaningful learning. Although research on this topic is still developing, in Mathematics there is some evidence to suggest that integrating a game-based approach into traditional teaching methods can help make learning a more motivating, interactive, and engaging experience. The PISA 2022 Technical Report (OECD, 2024a) shows a decline in average scores in Mathematics for OECD countries. The drop in the OECD average for Mathematics performance between the 2018 and 2022 PISA reports was actually 17 points, and Italy's results in Mathematics are in line with the average for OECD countries. This decline is probably due to the effects of the COVID-19 pandemic, which had a significant impact on the education system. This Report (OECD, 2024a) highlights the difficulties faced by secondary school students and provides a useful framework for understanding the causes of difficulties in Mathematics that students may encounter later at university. Some studies report on the difficulties university students (non-mathematicians) face in Mathematics, such as in mathematical modelling (Marchisio ET AL., 2021; Fahlgren et al., 2024). Moreover, the Survey of Adult Skills 2023 (OECD, 2024b) assessed the literacy, numeracy, and adaptive problem-solving skills of adults aged 16–65. Italy scored below the OECD average, particularly in numeracy, where 35% of adults scored at or below Level 1. At Level 1 of 5, they can do basic arithmetic with whole numbers. In adaptive problem solving, 46% of adults performed at Level 1 or below. Adults at Level 1 of 5 can solve simple problems with few variables and little irrelevant information. In this scenario and in the current post-COVID-19 pandemic context, which is characterised by the widespread use of technology and extensive online and blended learning in education, two serious games for Mathematics learning were integrated into the Learning Tools Interoperability (LTI) framework within the Moodle LMS. The games are accessible to students anytime and anywhere, and they are intended as a supporting tool to learn Mathematics. Their main aim is to facilitate and enhance students' approach to Mathematics, to stimulate problem-solving and critical thinking skills, by providing an engaging and interactive way to reinforce learning and to put into practice the concepts introduced in the classroom. The study involved non-mathematics university students from four different degree programmes during the 2023-24 and five degree programmes in 2024-25 academic years: undergraduate degree course in Medical Radiology, Imaging and Radiotherapy Techniques, undergraduate degree course in Biotechnology, undergraduate degree course in Prevention Techniques in the Environment and Workplaces, and postgraduate degree course in Strategic and Security Sciences. Undergraduate students of Strategic and Security Sciences were also involved in the 2024-25 experimentation. Serious games were also experimented with students and teachers from Italian upper secondary schools during the same period. The serious games were also presented to PhD students on the Doctoral Programme in Digital Humanities at the University of Genoa (in collaboration with the University of Turin) who attended the “Gamification and Serious Games” course during the 2024-25 academic year. Qualitative feedback was collected on these last two occasions through interviews and/or questionnaires.

Although serious games have become popular as an innovative and engaging way to improve learning, it was also important to evaluate their real impact on students' learning and gaming experiences. This involves analysing multiple dimensions beyond simple academic performance. Key aspects to be assessed include:

- Usability: how easily students can navigate and interact with the game environment
- Game appeal: the level of enjoyment and aesthetic satisfaction derived from the gameplay
- Perceived usefulness: how students value the game as meaningful support for their learning process

It was also important to explore how serious games influence student engagement and motivation, as these factors play a central role in fostering participation and deeper learning. As outlined in the theoretical framework, serious games may also contribute to reducing math anxiety. Therefore, the impact of games on math anxiety was also investigated. The methods and results of the study are presented in the following sections.

3.2 Research Objectives and Research Questions

The research follows four main directions:

1. The design and development of innovative, game-based learning activities, specifically two serious games, aimed at improving students' approach to Mathematics. These activities are intended for both upper secondary school and university students from scientific and non-scientific degree programmes. The activities are integrated within a DLE.
2. The experimentation of the two serious games with university and upper secondary school students and teachers to evaluate the impact of the tools and the proposed teaching methodologies on learners.
3. The evaluation of the impact of the proposed teaching methodologies with a focus on developing their disciplinary and transversal skills and the engagement and perception towards the discipline.
4. The development of a digital game-based model that is designed to be integrated into a DLE.

One of the main problematic questions addressed in this research was evaluating the effectiveness and impact of the serious games. Several well-established models and validated instruments, shown in the previous theoretical framework section, were identified and adapted to design an initial and final questionnaire capable of assessing the proposed tools from multiple perspectives, including students' levels of engagement, motivation and attitudes towards Mathematics, as well as their degree of appreciation of the experience.

The research questions identified in this study are the following:
(RQ1) How can the effectiveness and impact of the developed serious games be evaluated, both in terms of technical and game-related aspects (such as usability, game appeal, engagement, and perceived usefulness) and discipline-related dimensions (such as motivation, attitude, and math anxiety)?
(RQ2) Do serious games designed improve students' approach to, and learning outcomes in, Mathematics?
(RQ3) How can game-based activities be designed to be effectively employed by teachers in their instructional practices and by students in their learning process?
(RQ4) What advantages can new GenAI tools offer in terms of developing and using serious games?

4. Design and Development of the Serious Games

Chapter 4 provides a detailed description of the design and development process of the two serious games developed for Mathematics education and integrated within a digital learning environment. The games were implemented and tested in higher education and upper secondary school contexts. Chapters 5 and 6 examine the experimental procedures and findings related to these implementations in depth. Furthermore, designing and developing the two serious games played a central role in shaping the proposed model for developing digital, game-based learning activities, presented in Chapter 7.

4.1 Design of the two serious games

Two serious games were designed and developed. Their main aims are to facilitate and enhance students' approach to Mathematics, by providing an engaging and interactive way to reinforce learning and to put into practice the concepts introduced in the classroom. Rather than being conceived as mere supplementary activities, these games were designed as instructional tools that translate key mathematical concepts into interactive scenarios. The serious games cover topics taught in the classroom and encourage students to apply their knowledge into practice. The games are designed not only to develop disciplinary competencies but also to stimulate transversal skills such as problem-solving and critical thinking. The games are integrated into the Learning Tools Interoperability (LTI) framework within the Moodle Learning Management System (LMS). To develop the serious games, the E-core editor was used, which is characterized by a high level of complexity while ensuring ease of use for non-game designers (Fiorese et al., 2019). The E-core editor is a web-based software (available at https://github.com/EntropyKN/ecore3.0) that can be integrated with any LMS, such as Moodle, that supports the LTI protocols. The editor communicates with the LMS platform by transmitting data related to the game sessions, which is a great advantage as it allows for seamless tracking and analysis of user progress. This integration makes the games accessible to students at any time and from any location, providing flexible learning support that extends beyond the traditional classroom. This also supports a formative assessment approach, enabling learners to practise, reflect on their progress and continuously improve their skills. This integration also enables both instructors and students to access the game-generated scores directly within the LMS. The system centralises performance data on a single platform, providing students with feedback on their progress and allowing teachers to monitor their learning paths based on the performance. This compatibility ensures that the E-core can be easily integrated into a wide range of educational platforms, facilitating the adoption of DGBL and breaking down the barriers associated with the difficulty and cost of adopting digital tools in schools and universities (Huang et al., 2013; Vidakis et al., 2019). The editor allows response options to be inserted in the form of images, text or formulas. Media such as files, images and videos, as well as external links, can also be attached at various stages of the game. In this way, the editor facilitates the incorporation of multiple representations within DGBL as mentioned by Chen et al. (2023), enabling learners to interact with content through different modalities.

The serious game "The murderer at the High-Tech Institute" (further indicated as the "Murderer game") revolves around with the murder of Teresa, a scientist at the High-Tech Institute (Figure 5). The murder investigation format was chosen to leverage a well-established and familiar gaming model. Adopting a recognisable and engaging narrative structure minimises the effort and time required, allowing players to focus more on the learning objectives than on understanding unfamiliar game mechanics (Nautiyal, 2024). Moreover, this format naturally encourages curiosity, problem-solving and active participation, as players are motivated to collect clues, analyse evidence and draw conclusions. The aim of the game is to find the killer by using the mathematical clues given to the player, which refer to the path taken by the murderer. By solving the mathematical tasks, the player can find the murderer. The story opens with Teresa's murder, immediately setting out the central challenge of determining who committed the crime. The first clue suggests that the route taken by the murderer corresponds to the graph of a real-valued function. At this point, Kathy, a mathematician at the High-Tech Institute, enters the scene. Although she was unable to recognise the killer, who was wearing a disguise, she observed several details of the route taken closely. These observations are translated into a sequence of mathematical clues, which are formulated as properties that the function must satisfy for its graph to coincide with the murderer's trajectory. Once all the clues have been analysed, the player must evaluate the possible suspects and select the one whose movements are consistent with the reconstructed path. The game consists of 12 steps, with a maximum of four answer options for each step. Players can earn a maximum of 20 points per step by answering correctly. With a total of 63 points, the game is won by scoring at least 48 points. The first version of the serious game consisted of 43 points, even if several modifications were introduced in the second version, as described in Section 6.2. The second serious game is called "Don't open that spacecraft", further indicated as the "Alien game" (Figure 6). It revolves around Allan-Meta, an alien from the planet Metaland who has just arrived on the Earth. Upon landing, Allan-Meta is confronted with a flood of misleading information and fake news about our planet. Overwhelmed by these alarming claims, he retreats fearfully into his spacecraft. Oliver, a human character, helps Allan-Meta by examining the available evidence, clarifying misconceptions and calmly debunking false claims, to reassure the alien. The player's task is to help Allan-Meta, guided by Oliver's input, to understand Earth more accurately, teaching him to read and interpret data and information correctly. The game is designed to help students apply key concepts related to statistics and probability, particularly descriptive and inferential statistics. Through Allan-Meta's inquiries, players are encouraged to critically evaluate real-world issues, assess the reliability of information and consider how data can be interpreted - or misinterpreted - in everyday contexts. The game consists of 15 steps, each with a maximum of four answer options. Players can earn a maximum of 5 points per step by answering correctly. With a total of 41 points, the game is won by scoring at least 31 points.

Figure 5 - Frame from the serious game "The murderer at the High-Tech Institute".

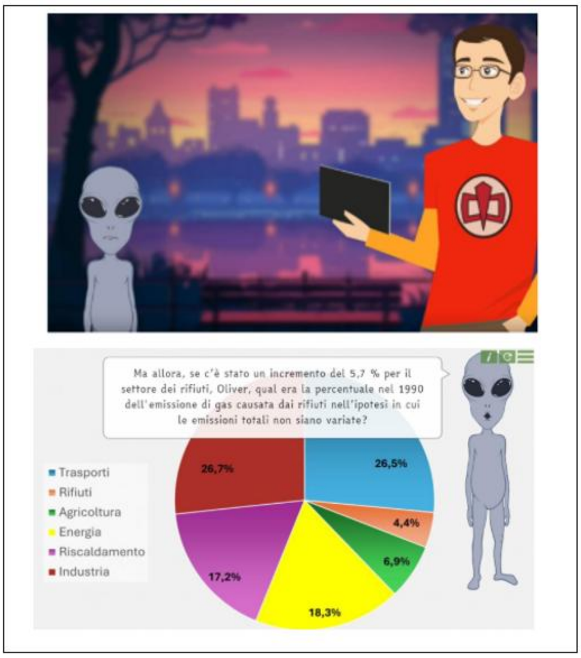

Figure 6 - Frame from the serious game "Don't open that spacecraft".

Figure 7 shows the activities related to the Murderer game on the platform dedicated to each course (the same applies for the other serious game). There is a dedicated Moodle section. Before playing the serious game, an initial questionnaire is asked to be filled. The instructions and rules of the serious games are provided on a dedicated Moodle page, but the same instructions also appear when the game is opened. After that the game can be played, followed by a final questionnaire.

Figure 7 - Activities related to the Murderer game on the Moodle platform.

Students can see their score at the end of the serious game, which ranges from 0% to 100%. This percentage reflects the number of points obtained relative to the total number of points available in the game. The winning range is set between 75% and 100%, a relatively high threshold designed to make the game more challenging. Some answers are only partially correct, meaning that the user can receive a lower score for them than for completely correct answers. The attached files contain useful hints for progressing in the game, so watching the videos is mandatory. Each step and its attached files can be viewed as many times as is required until a choice is made. Once the game is complete, it is possible to view the final score, along with all the gameplay details. A histogram is also provided (Figure 8), with the bars corresponding to the answers given at each step, from the first to the last. A bar at 100 represents a completely correct answer, while a bar at 0 represents a completely incorrect answer. Partially correct answers have an intermediate value between 0 and 100 depending on the accuracy of the response.

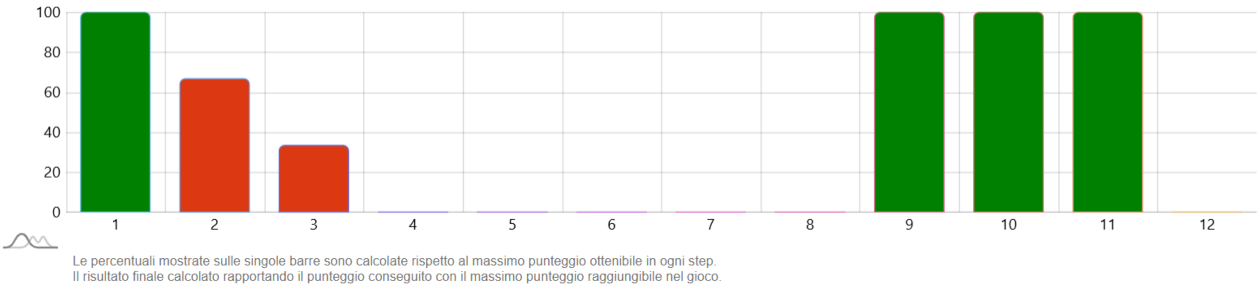

Figure 8 - Gameplay details at the end of game session.

After completing the game, clicking on "*Watch the replay of the match*" allows players to view the full video of the gameplay, including the answers provided by the user.

The design of the serious games followed the process outlined by Cruz et al. (2023), as illustrated in Figure 9. This covered the entire process for the Murderer game and up to the release of the v0.1 version of the "Alien game", since one main experiment was conducted. First, the contents and key concepts on which the serious game focuses were identified. Fundamental mathematical concepts were selected to support the development of both disciplinary and transversal competencies, ensuring alignment with the learning objectives and learning outcomes of the intended courses for the experimentation phase. Once the key concepts had been established, the game design phase began. It also includes adapting to the technology used - specifically, the E-Core Editor - to ensure effective implementation. A playable version (v.0.0) was then released and reviewed by members of the DELTA Research Group at the University of Turin. The reviewers are mathematicians who have experience of studying digital methodologies for teaching and learning Mathematics. The review addressed disciplinary and content-related aspects, as well as technical aspects aimed at identifying bugs or errors during gameplay. Based on the feedback received, a revised version (v.0.1) was created and used in the pilot study. An additional step was carried out for the Murderer game: based on the data obtained from the pilot study, a further revised version (v.1.0) was developed and used during the second experimentation.

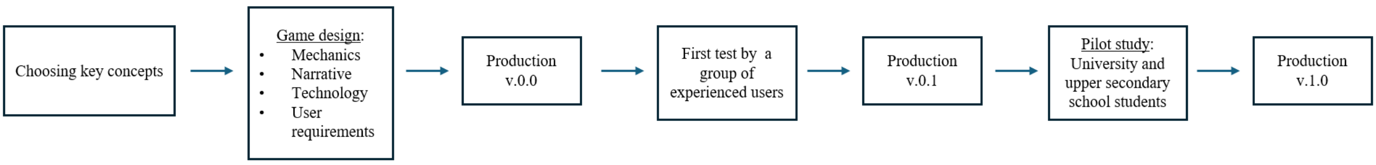

Figure 9 - The process of developing serious games.

The design of the two serious games began with a guided template, through which the general aspects of the games were defined, including their titles, target, descriptions and objectives. Both educational and gameplay aspects had to be defined. The didactic components included learning outcomes, they were formulated based on the Dublin descriptors, with the expected competencies and skills identified. Learning outcomes include transversal (not strictly related to a specific discipline) and discipline specific skills. Table 3 shows learning outcomes for the two games.

Dublin Descriptors	Murderer game	Alien game
Knowledge and understanding	Identify the main properties of elementary functions. Identify a suitable representation to illustrate the problematic situation.	Describe the main types of data and statistical sources. Know the basic concepts of descriptive statistics (mean, variance, mode, etc.) and how to apply them. Identify correct and misleading graphical representations.
Applying knowledge and understanding	Analyse the problematic situation, represent the data, interpret it and translate it into a mathematical model. Develop the problem-solving process by solving the problematic situation in a coherent, complete and correct manner, applying the necessary knowledge.	Use graphical tools to accurately represent a data set. Apply critical reading criteria to graphs and infographics.
Making judgements	Obtain information and derive solutions from the identified mathematical model. Evaluate information to make decisions.	Evaluate the reliability and completeness of statistical representations. Assess the accuracy and transparency of the use of data in the media.
Communication Skills	Interpret, analyse, and discuss information.	Interpret and discuss charts and tables.
Learning skills	Break down complex problems into smaller, more manageable parts to make them easier to understand and solve. Match digital tools and resources to specific objectives.	Recognise situations in which data is misused. Compare different representations of the same dataset. Match digital tools and resources to specific objectives.

Table 3 - Design of serious games in terms of learning outcomes.

Once the content had been defined from a didactic perspective, the game design phase began. During this phase, the previously identified educational contents were translated into concrete game elements. The aim was to create an interactive experience that it is both educational and engaging. This phase was developed considering the theoretical frameworks identified in the literature and presented in Section 2. The design was based on the three steps identified by Huang et al. (2013) for DGBL, game structure, game involvement and game appeal. These three phases summarised the main characteristics required for the design process. However, it became clear that it was necessary to explicitly define the behaviours adopted by players in relation to the learning outcomes. For example, one of the learning objectives was related to the ability to choose an appropriate representation for modelling the problematic situation. Therefore, behaviours that would enable this outcome to be achieved had to be introduced into the game, such as the option to make more general choices that could encompass multiple cases rather than being limited to the specific case presented. The following steps were identified for the design process:

1. Game structure: it is the game's basic structure, the core elements around which everything else is built. It includes several components such as rules, game genre, storyline, goals and explanations of game tasks (Huang et al., 2013).
2. Game dynamics: these include the actions performed by the player, the strategies adopted, the expected behaviour, how they respond to challenges and their interaction with the system (Werbach & Hunter, 2012). In the design of the games, for each step, losing or winning (or partially winning) behaviours were identified. It is what Plass et al. (2016) called a “game mechanic”. This refers to the core actions or behaviours associated with learning or assessment tasks within a game. An example is shown in Table 4.
3. Game appeal: it includes the game’s sound, graphics, dialogues, images, videos and animations, which can foster both cognitive and emotional engagement among learners.

Step n.	Winning profile behaviour	Losing profile behaviour
1	A methodical and precise approach: the player either asks for additional details or realises that only an estimate of the time can be made at best.	Impulsive behaviour: the player makes a rough estimate of the time available, even when sufficient information is not yet known.

2	Partially winning profile: the player provides an approximate time estimate, but it is not the most accurate. Winning profile: the player estimates the time of the murder as precisely as possible.	The player's estimation of the time of the murder is incorrect.
3	Partially winning profile: the player selects a reference system that is suitable for representing the character's path in this stage only. Winning profile: the reference system is also suitable for representing the characters' paths in subsequent steps.	Impulsive behaviour: the player traces the character's path using the available information without identifying an appropriate reference system.
4	The player recognises the graph representing Emily's path among the proposed ones.	The player does not recognize the graph representing Emily's path among the proposed ones.
5	The player recognises that Emily's path does not represent a function.	The player does not rule out Emily as a potential killer because she/he fails to recognise that the path does not represent a function.
6-7	Methodical and precise approach: before blaming Olivia, the player wants to understand the reference system considered to draw consistent conclusions. Once the reference system is specified, the player realizes that Olivia could not have passed through the living room.	The player fails to recognize that Olivia's passage through the living room depends on the reference system considered and therefore draws hasty conclusions.
8-9	The player uses all the available clues (even/odd function and quadratic function) to correctly translate the requirements that the function must satisfy for Jack to have passed through the hall and to be guilty.	The player fails to consider all the available clues or is unable to translate them into the requirements that the function must satisfy to implicate Jack.
10-11	The player uses all the available clues, such as the even/odd quadratic function and the derivative, to correctly translate them into the requirements that the function must satisfy for Mark not to be considered a potential suspect.	The player fails to consider all the available clues or is unable to translate them into the requirements and considers Mark as a potential suspect.
12	The player identifies Jack as the murderer.	The player identifies Emily, Olivia, or Mark as the murderer.

Table 4 - Expected behaviour for winning/losing profile for the Murderer game.

The guided design template is aligned with the editor's settings and requirements, helping to facilitate the implementation phase. It also includes information about the text, images, videos and link files to be inserted into the game. The editor also allows differentiated steps to be created based on the user's response. In the Murderer game, however, this feature was not included, meaning that both winning and losing profiles at each step led to the same next step through a linear structure (Figure 10). In contrast, in the "Alien game", this feature was implemented using a branched structure (Figure 11). Specifically, there were steps where the losing profile led to an intermediate guided step to prompt the player to reflect on the incorrect choice, and the winning profile led to an intermediate step for further exploration, such as introducing a variation of the problem scenario. These two structures were employed to investigate whether students' responses differed depending on the structure used. The editor does not allow users to go back once a choice has been made, to give weight and significance to the responses provided. This reflects real-life situations, in which actions cannot be retraced. The difficulty of the two serious games increases gradually, allowing players to build their skills and confidence step by step. This gradual increase is designed to maintain engagement by providing a new challenge and reducing frustration.

Figure 10 - Linear structure of the Murderer game.

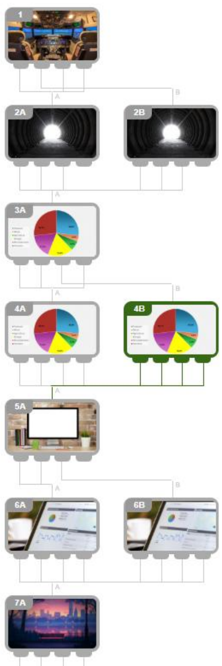

Figure 11 - Branched structure of the Alien game.

GenAI tools were used to design the Alien game. Open online tools (available at https://www.futuretools.io/) were used, for image generation and for the creation of the alien avatar Allan-meta. In addition, these tools were also employed in the development of the storyline of the serious game. It was also decided that students should be allowed to use GenAI tools to help them solve the serious games. To this end, prompts in the form of the attached images (Figure 12) were suggested for use with GenAI tools, encouraging students to use these tools to help them recall certain concepts they had forgotten (e.g. the properties of elementary functions).

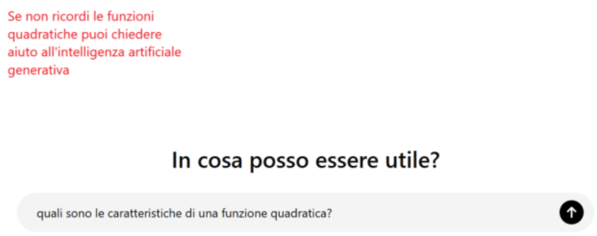

Figure 12 – Example of prompt provided to students in the Murderer game.

5. Methodology

This chapter provides a detailed overview of the didactic experimentation carried out in higher education and upper secondary school settings, particularly for data collection and analysis. It describes the methods and tools used during the experimental phase. These methods form the basis of the results presented in the next chapter. The chapter also provides an in-depth discussion of the design considerations underlying the development of the pre- and post-questionnaires used in the experiment, emphasising their structure, purpose and function in the data collection process.

This research adopts a mixed-methods approach to investigate the design, development, implementation, and evaluation of serious games in a DLE for Mathematics education. The following research questions guided the methodological choices.

(RQ1) How can the effectiveness and impact of the developed serious games be evaluated, both in terms of technical and game-related aspects (such as usability, game appeal, engagement, and perceived usefulness) and discipline-related dimensions (such as motivation, attitude, and math anxiety)?

To this end, two questionnaires were designed and administered to the students: an initial one to be completed before playing the serious game and a final one to be completed afterwards. The final questionnaire investigates the use of serious games in a DLE for mathematics learning, including technical and game-related aspects. Both questionnaires aimed to detect possible changes in discipline-related dimensions and attitudes in Mathematics before and after the gaming experience. The initial questionnaire was developed based on constructs already established in the literature concerning the perception of Mathematics. The final questionnaire was built drawing on existing scales from the literature assessing usability, technology acceptance and engagement. However, in most cases, these items were not originally developed specifically for the game-based learning context. Additional ad hoc items were therefore created to address the impact of DGBL on Mathematics learning. To ensure the quality of the data collected, for both questionnaires, reliability analyses were conducted to assess internal consistency using Cronbach’s alpha, both for the entire questionnaire and for item groups within each construct. An Exploratory Factor Analysis (EFA) was conducted on the final questionnaire using responses collected during the 2023-24 academic year. This analysis aimed to identify empirically emerging groups of underlying factors and to examine whether they were consistent with the dimensions originally intended to be measured. Based on the results of the pilot study in 2023-24 academic year, the final questionnaire was refined for subsequent didactical experimentation in the 2024-25 academic year, whose responses were used to perform a Confirmatory Factor Analysis (CFA). Finally, based on the identified factors and the original scales, also to allow comparisons with existing literature, descriptive and inferential statistical analyses were conducted.

(RQ2) Do serious games designed improve students' approach to, and learning outcomes in, Mathematics?

The impact of the developed serious games on learning outcomes was analysed by examining the performance of university students on exam questions related to the topics covered by the games. More specifically, questions on single-variable functions were considered for “The murderer at the High-Tech Institute” (the Murderer game), while questions on descriptive and inferential statistics were considered for “Don't open that spacecraft” (the Alien game). As it was not possible to administer a pre-test, the exam results of students who played the serious games during the 2023-24 and 2024-25 academic years were compared with those of a control group: students from the same academic year who did not play the game, and students from the previous academic year when such tools were not implemented. However, high-school mathematics grades reported in the initial questionnaire were also considered to assess whether there were any differences in the incoming level of mathematical background between students enrolled in the 2023-2024 and 2024-2025 academic years. It should be noted that this information was not available for the control group composed of students from previous academic years. The main experimental sample comprised undergraduate and postgraduate students from the University of Turin. However, the game-based activities were also proposed to secondary school students, their teachers and PhD students.

(RQ3) How can game-based activities be designed to be effectively employed by teachers in their instructional practices and by students in their learning process?

Drawing on the literature review, the design of the serious games and the results obtained from the experimental studies, a new digital game-based model was developed for integration within a DLE. The model was designed to facilitate the implementation of game-based activities by both students and teachers. The aim is to adapt the model for use in as many other disciplines as possible, beyond mathematics.

(RQ4) What advantages can new GenAI tools offer in terms of developing and using serious games?

To investigate the potential advantages offered by new GenAI tools, GenAI-based resources were incorporated into the design of the Alien game to generate game components such as images and the storyline. In addition, students were provided with specific tools, prompts and guidance on how to use GenAI as a support for completing the required tasks within the game. By integrating these AI-assisted resources, the study aimed to explore not only how students engage with the games but also how GenAI might enhance their learning experience. To assess this, targeted questions were included in the final questionnaire in the second study in the 2024-25 academic year, allowing an analysis of both the extent and manner of GenAI usage, as well as its potential impact on students' performance and learning outcomes.

5.1 Design and Implementation of Initial and Final Questionnaires

The initial and final questionnaires were developed based on the studies presented in Chapter 2, with a particular focus on the validated questionnaires. The items were chosen from the selected scales based on the specific constructs that the research aimed to evaluate, thus ensuring alignment with the study's objectives. It is specified that not all items from the scales were included to avoid making the questionnaires too long and due to the presence of items not aligned with the intended constructs. Regarding the initial questionnaire (QI), the items primarily focused on students' attitudes towards Mathematics and were drawn from the Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991) and the PISA 2012 Technical Report (OECD, 2014). The same items were included in the final questionnaire to allow a direct comparison of students' perceptions of Mathematics before and after playing the serious game. To facilitate pre- and post-comparison, the items were grouped into eight dimensions related to the attitudes towards Mathematics, based on their content and the construct they were intended to measure in the questionnaire. These dimensions are Perception of Mathematics, Interest and motivation in Mathematics, Math anxiety, Persistence in Mathematics, Persistence in problem-solving, Development of critical thinking, Perception of the course, and Extrinsic motivation. In addition to these questions, participants were asked to provide personal information, such as their age and gender, the type of school they previously attended, and their Mathematics grade in upper secondary school, to gain an understanding of their mathematical background. They were also asked to provide information on the number of hours they spent playing video games.
Regarding the final questionnaire (QF), to measure the impact of the serious games on students, it was needed a tool that would allow to measure 5 dimensions:

1) Impact of digital game-based learning on Mathematics learning.
2) Students' perceived effectiveness of the serious game for learning Mathematics.
3) Need for support, training and complexity of the game.
4) Gameplay experience and appeal of the serious game.
5) Engagement, in terms of students' involvement, as well as their experience of flow and immersion while playing the game.

A review of the literature revealed that the Technology Acceptance Model (TAM) (Davis,1989) and the System Usability Scale (SUS) (Brooke, 1996; Lewis, 2018) are commonly used tools for measuring digital game-based learning. Items for engagement were primarily added from Game Experience Questionnaire (GexpQ) (Poels et al., 2007), and then from the Game Engagement Questionnaire (GengQ) (Cruz et al., 2023). The theoretical constructs from the TAM and SUS were adapted to the serious game context to measure the perceived effectiveness of the serious game for learning Mathematics. The "Perceived Usefulness" section of the TAM was used to design the questionnaire, as this refers to the extent to which a person believes that using a particular technology will improve the performance at work or the overall experience. Since the GexpQ and GengQ already measured constructs related to engagement and were well suited to the context DGBL, the items derived from them were not extensively adapted. In addition, 14 new items were designed ex novo to assess the impact of DGBL on students' perceptions of their Mathematics learning, in line with the objectives of the research project. Given the questionnaire's somewhat self-developed nature, more studies than the initial questionnaire were conducted on final questionnaire to enhance its robustness. Therefore, Cronbach's alpha was not the only statistic evaluated; an EFA was conducted in the pilot study, followed by a CFA in the second study. This approach was chosen also because one of the study's main goals is to develop a final questionnaire specifically designed to measure the impact of the serious game. The items included in the initial questionnaire were therefore not directly linked to this objective but were primarily used for comparing initial and final perceptions.

Most of the items in both QI and QF - namely, those related to the attitudes towards Mathematics, and those derived from the scales (TAM, SUS and GExpQ and GEngQ), as well as the newly developed items assessing the impact of DGBL on Mathematics learning - were measured on a six-point Likert scale ranging from 1 to 6. Students were asked to indicate the extent to which they agreed with each statement, with 1 meaning "strongly disagree" and 6 meaning "strongly agree". "N/A" (Not Applicable) responses were allowed for the DGBL and TAM items for participants who felt they could not express an opinion on the use of DGBL in Mathematics. Questions that use a Likert scale are labelled with their respective codes: TAM for the items based on the Technology Acceptance Model, SUS for the items based on the System Usability Scale, ENG for the items derived from the GExpQ and GEngQ and DGBL for the new items created by the research group. For the items regarding attitudes towards Mathematics in QI and QF, the following label are used Perception of Mathematics (PM), Interest and motivation in Mathematics (IMM), Math anxiety (MA), Persistence in Mathematics (PersM), Persistence in Problem Solving (PersPS), Development of Critical Thinking (CT), Perception of the Course (PC), Extrinsic Motivation (EM). In addition, two open-ended questions were added to gather comments, observations, suggestions and feedback on any difficulties experienced. Tables 5 and 6 show the structures of the initial and final questionnaires, respectively, which were given to students taking part in the pilot study of the Murderer game in the 2023-24 academic year. Tables 7 and 8 show the items on the initial and final questionnaires that used a Likert scale with their label. The QF includes the same items as the QI to allow for a pre-post comparison. The only difference is that the questions in QF are set in the context of the serious game rather than Mathematics in general. It should be noted that the QI initially incorporated DGBL and TAM items, which were included in the first experiment for comparison with the QF. However, these questions were later deemed irrelevant as students had never used this approach before and would therefore have responded randomly. Consequently, they were not considered and were removed from the second experiment.

Item Group (Dimension)	ID	Number of items
Personal information (i.e. gender, secondary school background) and number of hours spent playing video games	/	8
Perception of Mathematics	PM	3
Interest and motivation in Mathematics	IMM	4
Math anxiety	MA	4
Persistence in Mathematics	PersM	6
Persistence in problem solving	PersPS	4
Development of critical thinking	CT	8
Perception of the course	PC	9
Extrinsic motivation	EM	4

TOTAL QUESTIONS	50

Table 5- Structure of the initial questionnaire from the pilot study.

Item Group (Dimension)	ID	Number of items
Perception of Mathematics	PM	3
Interest and motivation in Mathematics	IMM	4
Math anxiety	MA	4
Persistence in Mathematics	PersM	6
Persistence in problem solving	PersPS	4
Development of critical thinking	CT	8
Perception of the course	PC	9
Extrinsic motivation	EM	4
Items designed ex novo to assess the impact of DGBL on Mathematics learning	DGBL	14
Technology Acceptance Model	TAM	6
System Usability Scale	SUS	10
Engagement	ENG	32
Comments, observations, suggestions and difficulties.	/	2
TOTAL QUESTIONS		106

Table 6- Structure of the final questionnaire from the pilot study.

Considering your perception of Mathematics, how much do you agree with the following statements	**ID**
I am just not good at Mathematics.	PM_A
I learn Mathematics quickly.	PM_B
I keep studying until I understand Mathematics material.	PM_C
I do Mathematics because I enjoy it.	IMM_A
I am interested in the things I learn in Mathematics.	IMM_B
Learning Mathematics is worthwhile for me because it will improve my career (prospects, chances).	IMM_C
I will learn many things in Mathematics that will help me get a job.	IMM_D
I get very tense when I have to do Mathematics homework.	MA_A
I feel helpless when doing a Mathematics problem.	MA_B
When I take a test, I think about how poorly I am doing compared with other students.	MA_C
When I take tests, I think of the consequences of failing.	MA_D
Thinking about Mathematics lessons (you may refer to your past experience), to what extent do you agree with the following statements	

I prefer course material that really challenges me so I can learn new things.	PersM_A
If I study in appropriate ways, then I will be able to learn the material in this course.	PersM_B
During class time I often miss important points because I'm thinking of other things.	PersM_C
I often feel so lazy or bored when I study for this class that I quit before I finish what I planned to do.	PersM_D
I work hard to do well in this class even if I don't like what we are doing.	PersM_E
When course work is difficult, I either give up or only study the easy parts.	PersM_F
When confronted with a problem, I give up easily.	PersPS_A
I put off difficult problems.	PersPS _B
When confronted with a problem, I do more than what is expected of me.	PersPS _C
I like to solve complex problems.	PersPS _D
Thinking about your studies in Mathematics and your past experience, how much do you agree with the following statements	
When reading for this course, I make up questions to help focus my reading.	CT_A
When a theory, interpretation, or conclusion is presented in class or in the readings, I try to decide if there is good supporting evidence.	CT_B
When I become confused about something I'm reading for this class, I go back and try to figure it out.	CT_C
If course readings are difficult to understand, I change the way I read the material.	CT_D
I treat the course material as a starting point and try to develop my own ideas about it.	CT_E
I ask myself questions to make sure I understand the material I have been studying in this class.	CT_F
I try to change the way I study in order to fit the course requirements and the instructor's teaching style.	CT_G
I often find that I have been reading for this class but don't know what it was all about	CT_H
Thinking about this course, to what extent do you agree with the following statements	
I am very interested in the content area of this course.	PC_A
I think the course material in this class is useful for me to learn.	PC_B
I'm confident I can learn the basic concepts taught in this course.	PC_C
I'm confident I can understand the most complex material presented by the instructor in this course.	PC_D
It is my own fault if I don't learn the material in this course.	PC_E
I expect to do well in this class.	PC_F
I think I will be able to use what I learn in this course in other courses.	PC_G
I believe I will receive an excellent grade in this class.	PC_H
I'm confident I can do an excellent job on the assignments and tests in this course.	PC_I
The most important thing for me right now is improving my overall grade point average, so my main concern in this class is getting a good grade.	EM_A
Getting a good grade in this class is the most satisfying thing for me right now.	EM_B
If I can, I want to get better grades in this class than most of the other students.	EM_C
It is important for me to learn the course material in this class.	EM_D

Table 7- Items expressed in the Likert scale from the initial questionnaire of the pilot study.

To what extent do you think that the use of digital game-based learning (i.e. the use of games for learning, such as the Murderer game in Mathematics can	*ID*
Enhance learning.	DGBL_A
Increase your engagement.	DGBL_B
Stimulate your participation.	DGBL_C
Promote an inclusive learning approach.	DGBL_D
Promote a cooperative learning approach.	DGBL_E
Promote a personalized learning approach.	DGBL_F
Stimulate your interest.	DGBL_G
Stimulate interaction.	DGBL_H
Enable skills development.	DGBL_I
Meet your needs as a student.	DGBL_L
Reduce of the gender gap in scientific disciplines.	DGBL_M
Increase your participation in the learning process.	DGBL_N
Improve your perception of Mathematics.	DGBL_O
Reduce math anxiety.	DGBL_P
Allow you to achieve my learning goals faster.	TAM_A
Improve your performance.	TAM_B
Increase your productivity.	TAM_C
Increase your effectiveness.	TAM_D
Make Mathematics easier to learn.	TAM_E
Speed up learning Maths.	TAM_F
Based on your experience as a player of the Murderer game, to what extent do you agree with the following statements?	ID
I think I would like to use this game a lot.	SUS_A
The content and topics of the game are complex.	SUS_B
I had difficulty playing because the tool is complex to use.	SUS_C
I think I would need the support of a technician to use this game.	SUS_D
I found that the different features of this game were well integrated.	SUS_E
I found the game very difficult to use.	SUS_F
I felt very confident in using the game.	SUS_G
This game has a lot of things that interest me.	SUS_H
This game needs more introductory explanations.	SUS_I
The scenes in this game are very attractive.	SUS_L
Thinking about your experience as a player of the Murderer game, indicate how you felt during the game	
I was interested in the game's story.	ENG_A

I thought it was fun.	ENG_B
I was fully occupied with the game.	ENG_C
I thought about other things.	ENG_D
I felt competent.	ENG_E
I thought it was hard.	ENG_F
It was aesthetically pleasing.	ENG_G
I forgot everything around me.	ENG_H
I lost track of time.	ENG_I
I felt bored.	ENG_J
I felt successful.	ENG_K
I felt that I could explore things.	ENG_L
I enjoyed it.	ENG_M
I was fast at reaching the game's targets.	ENG_N
I felt pressured.	ENG_O
I felt challenged.	ENG_P
I found it impressive.	ENG_Q
I was deeply concentrated in the game.	ENG_R
I felt frustrated.	ENG_S
It felt like a rich experience.	ENG_T
I lost connection with the outside world.	ENG_U
I found it hard to get back to reality.	ENG_V
I felt time pressure.	ENG_W
I had to put a lot of effort into it.	ENG_X
I felt satisfied.	ENG_Y
I felt disoriented.	ENG_Z
I felt exhausted.	ENG_AA
I felt powerful.	ENG_AB
I felt weary.	ENG_AC
I felt proud.	ENG_AD
I feel like I just can't stop playing.	ENG_AE
The game feels real.	ENG_AF

Table 8- Items expressed in the Likert scale from the final questionnaire of the pilot study.

Considering the results obtained from the first pilot study, revisions were made to the initial and final questionnaires, as well as to the Murderer game. These changes are illustrated in the Results section.

5.2 Data Collection

Two didactical experimentations of the Murderer game were conducted: a pilot study during the 2023-24 academic year, followed by a second experimentation during the 2024-25 academic year. The Alien game was tested only once, during the 2024-25 academic year. This was because it was designed concurrently with, and following, the first experimentation, to draw on the findings from the initial study and considering them into the game design. The first pilot study was conducted with students from the University of Turin between October 2023 and November 2024, including the final examination period of the academic year. The second study was conducted with university students from October 2024 to November 2025. For university students, 1-2 bonus points were awarded for completing the activities related to the serious game, including filling out an initial questionnaire, completing at least one attempt of the serious game with a score of at least 50%, and completing the final questionnaire. In parallel, the activities were also carried out with students from Italian upper secondary schools that belong to the national “Problem Posing & Solving” (PP&S) network of schools (Fissore et al., 2024a). The pilot study of the Murderer game involved also upper Italian secondary school students from May to September 2024. The second study, which tested both the serious games, took place from January to September 2025. Some teachers asked their students to complete activities related to the serious game during the summer holidays.
The research methodology of the experimentations, involving both undergraduate and postgraduate students, as well as upper secondary school students, provided the collection of three types of data:

- initial questionnaire
- serious game sessions (i.e. scores, number of attempts)
- final questionnaire

The sample of students consists of those who completed all three activities. For some degree programs, it was also possible to consider exam results (administered using an automatic assessment system) relating to the topics covered by the serious games. As it was not possible to administer a pre-test, the exam results of students who played the serious games during the 2023-24 and 2024-25 academic years were compared with those of a control group: students from the same academic year who did not play the game, and students from the previous academic year when such tools were not implemented. However, high-school mathematics grades reported in the initial questionnaire were also considered to assess whether there were any differences in the incoming level of mathematical background between students enrolled in the 2023-2024 and 2024-2025 academic years. It should be noted that this information was not available for the control group composed of students from previous academic years. The exams used to compare the various student samples are of a comparable difficulty level. Moreover, the comparison was carried out by considering the different exam sessions scheduled each year. Thus, the test considered for each academic year is not just one. For example, this involved up to seven sessions and seven tests for the Biotechnology degree programme. In the case of multiple exam attempts, only the best attempt was considered. The analysis focused on two questions related to functions of real variables to assess the impact of the Murderer game on learning outcomes, and on two questions related to descriptive and inferential statistics in the context of the Alien game. Questions related to functions in a real variable were scored on a scale from 0 to 6, while questions related to descriptive and inferential statistics were scored on a scale from 0 to 3. The exam results were then compared with the data from the students’ game sessions, given that they had the opportunity to make multiple attempts.
In addition, during the 2024-25 academic year, teachers who had experimented with one or both serious games in their secondary school classes were asked, on a voluntary basis, to complete a questionnaire about the games after the didactical experimentation. Teachers were asked to complete the sections of the questionnaire related to DGBL, TAM and SUS variables, as these are the ones most closely associated with the use of DGBL in teaching. Furthermore, the Murderer game was presented to PhD students on the Digital Humanities programme during a doctoral course on the topic titled "Gamification and Serious Games". Post-game interviews were conducted with seven PhD students who did not have a strong mathematical background. These interviews aimed to gather feedback on the quality of the game and the experience of playing it, including aspects such as engagement, usability, and overall appeal, rather than evaluating their performance in the game.

5.3 Data Analysis

5.3.1 Data Analysis: University Sample

Reliability analyses were carried out on the initial and final questionnaire responses to assess internal consistency, using Cronbach's alpha for both the entire questionnaire and the item groups within each construct identified in the literature. The aim was to assess the internal consistency of each variable by examining how well the items within each question set measured the intended construct, that is, the extent to which participants' responses were consistent and reliable for each construct. A second reliability analysis was then carried out, considering the entire questionnaires, to check the consistency and overall reliability of the questionnaire and Cronbach's alpha again was calculated. To explore any significant relationships between the variables involved, particularly those in the final questionnaire, a correlation matrix was calculated based on the mean scores obtained from the SUS, TAM, DGBL and ENG items. The normality of the four variables mentioned above, calculated based on their mean scores, was examined to determine whether they followed a normal distribution. To assess this, both the Kolmogorov–Smirnov test and the Shapiro–Wilk test was conducted. The analyses were conducted by excluding the "N/A" (Not Applicable) responses, which were allowed for the DGBL and TAM items for those who felt they could not express an opinion on the use of DGBL in Mathematics. If there were only a few "N/A" responses, the average was calculated considering only the questions where "N/A" was not selected. However, if there were many "N/A" responses, the participant was excluded from the sample.

An Exploratory Factor Analysis (EFA) was conducted on the final questionnaire using responses collected during the 2023-24 academic year. This analysis aimed to identify empirically emerging groups of underlying factors and to examine whether they were consistent with the dimensions originally intended to be measured:

1) Impact of digital game-based learning on Mathematics learning.
2) Students' perceived effectiveness of the serious game for learning Mathematics.
3) Need for support, training and complexity of the game.
4) Gameplay experience and appeal of the serious game.
5) Engagement, in terms of students' involvement, as well as their experience of flow and immersion while playing the game.

EFA was performed on each variable individually and on all variables together, using Principal Component Analysis (PCA) with Varimax rotation and Kaiser normalization with convergence for rotation performed in 3 or more iterations. Based on the results of the EFA and reliability analysis, the final questionnaire was refined for experimentation in the 2024-25 academic year. The responses from the second experiment of the Murderer game and from the first experiment of the Alien game were used to perform a Confirmatory Factor Analysis (CFA). It was decided that the CFA should be conducted on the responses to the final questionnaire for both serious games, to assess whether they both confirmed to the identified factors. In the CFA, the following fit indices were calculated: the Comparative Fit Index (CFI), t Tucker–Lewis Index (TLI), and the Root Mean Squared Error of Approximation (RMSEA). Reliability analyses were performed once more for the second experimentations. Based on the identified factors and the original scales, also to allow comparisons with existing literature, descriptive and inferential statistical analyses were conducted. Depending on data normality, either parametric tests (t-test, ANOVA) or non-parametric tests (Mann-Whitney U or Kruskal-Wallis) were conducted to compare the groups and assess whether statistically significant differences were present. For the sake of completeness, non-parametric tests were performed even when the distributions met the normality assumptions, to provide further confirmation of the findings. Questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores. To compare the results with existing scales in the literature, the questionnaire was transformed from a 1-6 Likert scale to a 0-100 scale. Qualitative analyses were conducted on the open-ended questions in the questionnaire relating to suggestions, potential and difficulties encountered. Similar comments were identified in the analysis of the open-ended questions, categories were defined, and frequencies were counted. Unique comments were counted only once.

Changes in students' attitudes towards Mathematics were evaluated by comparing data from the questionnaires completed before and after playing the Murderer game. The Wilcoxon signed-rank test for paired samples was

used to analyse the difference between responses to the QI and QF collected from the same students. Null hypothesis of the test is: "The change in the mean is not significant"; while alternative hypothesis is: "The change in the mean is significant". The Wilcoxon signed-rank test for paired samples is sufficiently accurate for a sample size greater than 20. Due to their similar educational area (healthcare), students from the Medical Radiology, Imaging and Radiotherapy Techniques and Prevention Techniques in the Environment and Workplaces programmes were combined, given the small number of students in each programme. It was not possible to perform the same test for students enrolled on the Strategic and Security Sciences programme, as the sample size was smaller than 20. It was also deemed inappropriate to merge this group with the Biotechnology or Healthcare students, primarily due to their status as master's students rather than undergraduates, and secondarily due to their belonging to a different field. This test was applied to the quantitative responses (Likert scale from 1= strongly disagree to 6= strongly agree) provided by each student for both the QI and QF to establish whether the changes in mean response between the QI and QF are statistically significant. The obtained results were classified as moderately significant for $0.05 \leq$ p-value ≤ 0.10 and highly significant for p-value <0.05. This analysis involved comparing the mean values obtained from the QI and QF item scores for each of the eight dimensions: Perception of Mathematics, Interest and motivation in Mathematics, Math anxiety, Persistence in Mathematics, Persistence in problem solving, Development of critical thinking, Perception of the course and Extrinsic motivation. Subsequently, a comparison between the initial and final questionnaires was carried out at the single-item level to further explore which items within each dimension contributed most positively or negatively to the observed changes. A reliability analysis was conducted prior to running the Wilcoxon test. This involved calculating Cronbach's alpha for the responses to both the QI and QF, considering all eight dimensions together as well as each group of items (dimension) separately, to verify the internal consistency of the various scales. To further examine the internal consistency of each subscale, Cronbach's Alpha was calculated for the eight theoretical dimensions identified a priori.
For the analyses conducted on the exam test to compare the experimental and control groups, descriptive statistics of the exam scores were first computed. In some cases, relative frequency distributions across score ranges were calculated to examine how students' scores varied across ranges in the experimental and control groups. The cumulative distribution of exam scores was then computed and displayed, showing the proportion of students who achieved at least a given score for each group. These analyses were performed to investigate whether the experimental group obtained higher exam scores than the control group. Since no pre-test was possible to administer and no prior information on students' mathematical background was available for the control group, the only comparison possible of the mathematical background of students was the consideration of the secondary school mathematics grades reported in the QI by students participating in the 2023-24 and 2024-25 experimental groups. Descriptive statistics were calculated using the data collected during the game sessions. The exam results were then compared with the data from the students' game sessions, given that they had the opportunity to make multiple attempts. The analysis examined whether there was a correlation between the exam score and the number of attempts, and between the exam score and the mean score obtained in the serious game attempts. Spearman's correlation coefficient was used for this analysis. Correlations were also computed between the mean scores obtained in the various attempts at the serious game and the students' responses to the final questionnaire, as well as between the number of attempts in the serious game and the students' responses to the final questionnaire. Spearman's correlation coefficient was used for this analysis.
The study in the 2024-25 academic year surveyed students, asking if they used GenAI tools to solve and how they used them for the required tasks within the game. Qualitative responses from students who used GenAI tools were analysed by GenAI tool and use type. Mean scores across game attempts were calculated for both GenAI users and non-users, and independent-samples t-tests were conducted to assess whether GenAI usage had a significant impact on students' performance.

5.3.2 Data Analysis: Secondary School Sample

Data analyses were conducted separately: one for the didactical experimentation carried out during the 2023-24 school year, and another for the second experimentation conducted during the 2024-25 school year. Descriptive statistics (mean, median, and standard deviation) were computed for each variable (TAM, SUS, DGBL, and ENG), as well as for the factors identified in the final questionnaire for the second experimentation.

Questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores. To compare the results with existing scales in the literature, the questionnaire was transformed from a 1-6 Likert scale to a 0-100 scale. Descriptive statistics were also calculated using data collected during the game sessions, and correlations were examined between questionnaire responses and the average score obtained across multiple attempts at the serious game. Normal variable parametric tests (such as t-tests and ANOVAs) were conducted to compare the groups and assess whether there were any statistically significant differences.

6. Results

This chapter presents the results of the analysis of data collected during experiments involving two serious games, highlighting qualitative and quantitative findings. Specifically, two experiments were conducted with the Murderer Game and one with the Alien Game. This section provides a detailed overview of how participants interacted with and appreciated the games, and the learning outcomes they achieved. The chapter illustrates also the changes made to the serious game and to the final questionnaire used in the subsequent experimentation, based on the results of the pilot study. The analysis of these results aims to define a game-based learning model and to identify effective tools for assessing the impact of the serious games. These findings were then used to refine the model and address the research questions of the study, ultimately leading to the conclusions.

To address the research questions guiding this thesis, two experiments of "The murderer at the High-Tech Institute" (the Murderer game) were conducted during the 2023-24 (pilot study) and 2024-25 academic years. Additionally, one experiment involving the serious game "Don't open that spacecraft" (the Alien game) was conducted during 2024-25 academic year. Both experiments involved university and upper-secondary school students and their teachers. The Murderer game was also presented to in-service STEM secondary school teachers during a workshop and PhD students on the Digital Humanities programme, which is offered jointly by the Universities of Turin and Genoa. After the first pilot study, the results were analysed and changes were made to the Murderer game. These changes were also considered when developing the Alien game. Additionally, modifications were made to the final questionnaire given to students after they had completed at least one attempt of the game. The results are presented below and organised as follows:

- The first experiment of the Murderer game with university students during the 2023-24 academic year, along with the subsequent revisions made to the game and the final questionnaire.
- The second experimentation of the Murderer game in a university context during the 2024-25 academic year.
- The first experimentation of the Alien game with university students during the 2024-25 academic year.
- A presentation of the experiments carried in secondary school setting and with PhD students.

In line with the research questions - namely, how the effectiveness and impact of the developed serious games can be evaluated, the extent to which they improve students' approach to and learning outcomes in Mathematics, and the advantages that new GenAI tools can offer in the development and use of serious games - the main results for the sample of university students concern:

- The findings from the exploratory factor analysis of the final questionnaire conducted in the pilot study, followed by the confirmatory factor analysis carried out in the second study.
- Students' perceptions of the serious game across multiple dimensions, including engagement, usability, the impact of DGBL on mathematical learning, and technology acceptance. Comparisons were also made by degree programme.
- Further results examine changes in students' perceptions of Mathematics, as well as the impact of serious games on learning outcomes.

6.1 Results from the Pilot Study of the Murderer Game in Higher Education

The study, conducted from October 2023 to November 2024 involved 156 university students. The sample comprises students who completed the initial questionnaire before playing the serious game, then played the game at least once and filled out the final questionnaire afterwards. This sample includes 145 first-year undergraduate students enrolled in a scientific (non-mathematical) degree programme and 11 first-year master students enrolled in a strategic sciences degree programme. In particular, 20 students from undergraduate degree course in Medical Radiology, Imaging and Radiotherapy Techniques, 106 students from undergraduate degree course in Biotechnology, 19 students from undergraduate degree course in Prevention Techniques in the Environment and Workplaces, and 11 students from postgraduate degree course in Strategic and Security Sciences. The sample consists of 106 females, 49 males and one participant who preferred not to indicate the gender.

6.1.1 Results of the Reliability and Exploratory Factor Analysis from the Final Questionnaire

Table 9 shows the results of Cronbach's alpha calculation for the single variables (DGBL, TAM, SUS and ENG) of the first version of the final questionnaire used for the pilot study. Very high values were obtained for DGBL, TAM and ENG, above 0.9. In contrast, a slightly lower value was obtained for SUS, indicating a lower inner consistency of the items of the scale used to measure usability. However, the value is still considered acceptable. Cronbach's alpha was calculated for 30 items when the three variables DGBL, TAM and SUS were combined, and the questionnaire was found to be highly reliable (Cronbach's alpha = 0.951). Including also the ENG variable (62 items) increased Cronbach's alpha to 0.958, indicating excellent internal consistency across the entire set of items. Such high values indicate strong internal coherence among the items. However, it should be noted that Cronbach's alpha is sensitive to the number of items in a scale. Therefore, the observed increase is likely due not only to conceptual coherence, but also to the larger item pool. Nevertheless, the results confirm the robustness of the instrument for subsequent analyses. The Cronbach's alpha value remains stable for each item included, with values very similar to the Cronbach's alpha identified. This indicates that none of the items have a negative effect on the scale's internal consistency, and that removing individual items would not substantially improve the scale's overall reliability.

Variable	Number of items	Cronbach's alpha value
DGBL	14	0.947
TAM	6	0.959
SUS	10	0.708
ENG	32	0.899
DGBL, TAM, SUS combined	30	0.951
DGBL, TAM, SUS and ENG combined	62	0.958

Table 9 - Results of Cronbach's alpha calculation.

Table 10 shows the results of the Kolmogorov-Smirnov test and the Shapiro-Wilk test for normality of the four variables, performed on the average of the items of TAM, SUS DGBL and ENG. Based on the results, it can be said that the four variables are approximately normal. DGBL can be considered normal as both tests suggest normality. SUS and ENG may be close to normal, although the Kolmogorov-Smirnov test does not confirm normality. TAM on the other hand, show some evidence of normality according to the Kolmogorov-Smirnov test, but not according to the Shapiro-Wilk test.

	Kolmogorov-Smirnov test			**Shapiro-Wilk test**		
	p-value	Statistic	gl	p-value	Statistic	gl
TAM	0.065	0.071	147	0.012	0.976	147
SUS	0.036	0.076	147	0.058	0.982	147
DGBL	0.200	0.054	147	0.143	0.986	147
ENG	0.04	0.093	147	0.138	0.986	147

Table 10 – Normality of the variables.

The internal consistency of the questionnaire is also supported by the results of the correlation matrix calculated on the mean scores of the DGBL, TAM, SUS and ENG items. Table 11 shows the correlation matrix, which was calculated based on the means of the four variables. It shows the Pearson correlation coefficient (r) for each pair of variables, along with the corresponding p-value from the two-tailed significance test. The four variables exhibit significant correlations with one another, with p-values below 0.01, suggesting a high level of confidence in the validity of these results. In particular, the correlation between TAM and DGBL is very strong (0.879), indicating a robust and highly consistent relationship between these two constructs. This is consistent with the fact that students who perceived the serious game as an effective way of enhancing Mathematics learning are also more likely to attribute greater usefulness to this approach. The correlations between the other pairs of variables are also positive and moderate, suggesting meaningful associations, although not as strong as that between TAM and DGBL. The positive correlation between engagement and all three variables suggests that positive perceptions of the game's usability and usefulness enhance student engagement.

		TAM	**SUS**	**DGBL**	**ENG**
TAM	r	1	0.428	0.879	0.658
	p-value		<0.001	<0.001	<0.001
SUS	r	0.428	1	0.514	0.645
	p-value	<0.001		<0.001	<0.001
DGBL	r	0.879	0.514	1	0,668
	p-value	<0.001	<0.001	<0.001	<0.001
ENG	r	0.658	0.645	0.668	1
	p-value	<0.001	<0.001	<0.001	

Table 11- Correlation matrix on the mean scores.

To understand if there are underlying factors that can be identified, first it was conducted an exploratory factor analysis (EFA) on each of the three variables individually (SUS, DGBL, TAM and ENG), to determine whether there were a single factor or multiple factors within each variable. It identifies only one factor for the TAM items, with very high factor loadings of around 0.9. Two components are identified for usability (Table 12): one group consists of the items SUS_A, SUS_E, SUS_H, SUS_G, SUS_L, and the other group consists of the items SUS_B, SUS_C, SUS_D, SUS_F, SUS_I. The first dimension appears to be related to a positive perception of the game, encompassing interest, visual attractiveness, ease of use, and a sense of confidence when using it. The second one seems to be related to difficulties or issues in using the game. These two dimensions reflect two different aspects of the game experience: one related to ease of use and enjoyment, and the other related to difficulties with the interface and understanding of the game. Factor analysis for DGBL identifies two factors: except DGBL_M "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce of the gender gap in scientific disciplines*" and DGBL_P "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce math anxiety*", all other items belong to the same factor, with factor loadings ranging from 0.669 to 0.920. This aspect was of particular interest, so it was worth understanding how the two items performed in the PCA for the four combined variables.

Item	**ID**	**Factor 1 loadings**	**Factor 2 loadings**
I think I would like to use this game a lot.	SUS_A	0.869	
This game has a lot of things that interest me.	SUS_H	0.846	
I found that the different features of this	SUS_E	0.782	

game were well integrated.			
The scenes in this game are very attractive.	SUS_L	0.657	
I felt very confident in using the game.	SUS_G	0.580	
I had difficulty playing because the tool is complex to use.	SUS_C		0.858
I think I would need the support of a technician to use this game.	SUS_D		0.854
This game needs more introductory explanations.	SUS_I		0.739
I found the game very difficult to use.	SUS_F		0.686
The content and topics of the game are complex.	SUS_B		0.515

Table 12 - Rotated component matrix for SUS variable factorial analysis.

The results of the initial exploratory factor analysis conducted on the entire final questionnaire indicated that the TAM, USA, and DGBL items - particularly those from TAM and DGBL - tended to load on overlapping factors. In contrast, the ENG items formed a largely separate and distinct factor. Based on this pattern, a second set of exploratory factor analyses was carried out: one EFA focusing on the TAM, USA, and DGBL constructs, and a separate EFA dedicated exclusively to the ENG items. This distinction was theoretically predictable, as the ENG construct was expected to represent a dimension conceptually different from the impact of digital game-based learning, usability, and perceived usefulness. EFA was then repeated on the SUS, DGBL, TAM variables together, to identify the factors which emerge globally from the item of the final questionnaire related to these variables. Table 13 shows the rotated component matrix for the three variables together obtained from the PCA with convergence for rotation performed in 10 iterations. Factor analysis suggests that four distinct groups can represent the data. As shown in Table 13, Factor 1 primarily consists of the DGBL items and includes three items from the TAM. As expected, DGBL_M and DGBL_P are not included in Factor 1. These have high factor loadings from 0.6 to 0.893, suggesting a homogeneous group with strong internal correlation. Factor 2 comprises three TAM items and one DGBL item, all of which have relatively high factor loadings (>0.6). Factor 3 includes five SUS items and one DGBL item, the latter with a relatively low factor loading. Factor 4 groups the remaining usability items with a factor loading around 0.6. As expected from the correlation results, the factor analysis revealed that the DGBL and TAM variables share some common characteristics, as the TAM variables are distributed across two factors strongly associated with the DGBL items. In the rotated component matrix (Table 13), it is possible to observe that the usability-related items are mainly distributed across Factors 3 and 4. This suggests that the usability items do not represent a single construct but can be divided into two distinct dimensions. The usability variables do not have high factor loadings on Factors 1 and 2, indicating that usability is a separate construct with no overlap with the DGBL and TAM variables.

The items of the three initially considered variables were examined to understand what each factor from the factorial analysis represents and what can be detected using it. It is possible to observe that Factor 1 contains almost all the items of DGBL (Table 13), which are related to the impact of digital game-based learning on Mathematics learning. This factor addresses various aspects, including the perception that serious games can enhance learning, increase engagement, and stimulate participation and interaction. It was decided to label this factor as "The impact of digital game-based learning on the Mathematics learning experience", which was the first aspect to be measured. Factor 2 contains items related to the perceived effectiveness of digital game-based learning for Mathematics learning and its potential to enhance the learning experience. It examines the benefits of serious games for learning Mathematics, for example, whether they help simplify and accelerate Mathematics learning. In this set of items there is also the DGBL_P question, which is related to the reduction of math anxiety through to the use of the serious game, which could be interpreted as a benefit of the serious game for learning Mathematics. This factor as named "Perceived effectiveness for learning Mathematics". Factor 3 focuses on the user's interaction with the game, particularly the perceived complexity, any difficulties encountered, and the potential need for technical support. It also examines the level of interest and satisfaction users feel when interacting with the tool. Furthermore, this factor includes DGBL_M item "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*". Upon further consideration, this item could identify a potential difficulty arising from the game in terms of gender equality. Thus, factor 3 is called "Complexity of the game, possible difficulties and need for support". The last one, Factor 4, focuses on the overall value of the game, emphasising the user's perception of the game experience and the appeal of the serious game. This factor examines how enjoyable the game is for the user and the extent to which it is perceived as a fun and rewarding experience. Factor 4 has been entitled "Game experience and serious game appeal".

Item	**ID**	**Factor 1** "The impact of digital game-based learning on the Mathematics learning experience" loadings	**Factor 2** "Perceived effectiveness for learning Mathematics" loadings	**Factor 3** "The complexity of the game, possible difficulties and need for support" loadings	**Factor 4** "Game experience and serious game appeal" loadings
Stimulate your participation.	DGBL_C	0.893			
Increase your engagement.	DGBL_B	0.884			
Stimulate your interest.	DGBL_G	0.871			
Stimulate interaction.	DGBL_H	0.857			
Enhance learning.	DGBL_A	0.856			
Promote an inclusive learning approach.	DGBL_D	0.840			
Enable skills development.	DGBL_I	0.776			
Promote a cooperative learning approach.	DGBL_E	0.774			
Increase your participation in the learning process.	DGBL_N	0.753			
Promote a personalized learning approach.	DGBL_F	0.707			
Meet your needs as a student.	DGBL_L	0.676			
Increase your productivity.	TAM_C	0.676			
Increase your effectiveness.	TAM_D	0.658			
Improve your performance.	TAM_B	0.637			
Improve your perception of Mathematics.	DGBL_O	0.601			
Allow you to achieve my learning goals faster.	TAM_A		0.678		
Speed up learning Maths.	TAM_F		0.674		
Reduce math anxiety.	DGBL_P		0.653		
Make Mathematics easier to learn.	TAM_E		0.622		
I had difficulty playing because the tool is complex to use.	SUS_C			0.849	

I think I would need the support of a technician to use this game.	SUS_D			0.842	
This game needs more introductory explanations.	SUS_I			0.718	
I found the game very difficult to use.	SUS_F			0.652	
The content and topics of the game are complex.	SUS_B			0.575	
Reduce of the gender gap in scientific disciplines.	DGBL_M			0.398	
I felt very confident in using the game.	SUS_G				0.695
This game has a lot of things that interest me.	SUS_H				0.658
The scenes in this game are very attractive.	SUS_L				0.597
I think I would like to use this game a lot.	SUS_A				0.589
I found that the different features of this game were well integrated.	SUS_E				0.587

Table 13 - Rotated component matrix for the three variables (DBL, TAM, SUS).

Subsequently, an exploratory factor analysis was conducted on engagement items using the using PCA with convergence for rotation performed in 9 iterations. This construct was assumed to represent a distinct dimension compared to the impact of DGBL, TAM and SUS variables examined in the previous analyses. The analysis identified six main factors (Table 14):

- Factor 1$_{_ENG}$ - Enjoyment and involvement: It represents positive engagement with the game and the experience. It incorporates elements of interest, aesthetic satisfaction, enjoyment and personal enrichment. Almost all items show high positive factor loadings (≥ 0.70). This factor represents the core dimension of the construct, suggesting a coherent cluster of items related to participants' engagement or involvement. Except for item ENG_C, this factor aligns closely with the "Positive affect" and "Sensory and Imaginative Immersion" components of the GexpQ Core Module by IJsselsteijn, et al. (2007; 2013), both of which capture enjoyment, interest, and aesthetic engagement

- Factor 2$_{_ENG}$ - Negative feelings and pressure: it groups together the negative aspects of the gaming experience, such as stress, fatigue, confusion, and time pressure. This factor, as expected, groups together negatively worded items ("I felt pressured", "I felt frustrated" etc.). The factor loadings are high in this case as well. In this case the factor does not coincide with a single component identified in the GexpQ, but with multiple components, linked to "Tension/Annoyance", "Challenge", "Returning to Reality" and "Tiredness" scales.

- Factor 3$_{_ENG}$ - Competence and achievement: it reflects a sense of mastery and efficacy, with loadings approximately 0.7, except for the ENG_Y item. This factor clearly corresponds to the "Competence" and "Positive Experience" components in the GexpQ Core Module, which measure perceived skill, success and satisfaction during gameplay.

- Factor 4$_{_ENG}$ - Flow: it describes the state of flow or deep cognitive absorption, characterized by intense focus and loss of temporal and spatial awareness. It shows consistent loading factors approximately 0.7, except for the ENG_R item. This factor corresponds directly to the "Flow" dimension in the GexpQ Core Module, which assesses immersion, absorption, and time distortion during play.

- Factor 5$_{_ENG}$ - Loss of control and game realism: contains items related to game immersion, as well as an unexpected item related to game realism. The factor loadings are around 0.66–0.70. The grouping of seemingly unrelated items can be explained by the fact that two of the three items came from different questionnaires (ENG_V from the GexpQ while ENG_AE and ENG_AF from the GengQ). It was expected, due to their similarity, that Factors 5 and 6 would load on a single factor.

- Factor 6$_{_ENG}$ - Disengagement and boredom: it represents the opposite of engagement, reflecting distraction and boredom. This factor contains only two items, with factor loadings that are not so high. In the GexpQ Core Module they belong to the "Negative affect" component.

Some items - specifically ENG_F, ENG_K, ENG_P, ENG_Q, and ENG_X - did not load clearly on any single factor. These items may be ambiguously worded or may not measure any specific latent dimension effectively.

Item	**ID**	**Factor$_{ENG}$ 1** "Enjoyment and involvement" loadings	**Factor$_{ENG}$ 2** "Negative feelings and pressure" loadings	**Factor$_{ENG}$ 3** "Competence and achievement" loadings	**Factor$_{ENG}$ 4** "Flow" loadings	**Factor$_{ENG}$ 5** "Loss of control and game realism" loadings	**Factor$_{ENG}$ 6** "Disengagement and boredom" loadings
I was interested in the game's story.	ENG_A	0.855					
I thought it was fun.	ENG_B	0.836					
I was fully occupied with the game.	ENG_C	0.781					
It was aesthetically pleasing.	ENG_G	0.789					
I felt that I could explore things.	ENG_L	0.679					
I enjoyed it.	ENG_M	0.808					
It felt like a rich experience.	ENG_T	0.742					
I felt pressured.	ENG_O		0.725				
I felt frustrated.	ENG_S		0.732				
I felt time pressure.	ENG_W		0.641				
I felt disoriented.	ENG_Z		0.743				
I felt exhausted.	ENG_AA		0.843				
I felt weary.	ENG_AC		0.795				
I felt competent.	ENG_E			0.719			
I was fast at reaching the game's targets.	ENG_N			0.713			
I felt satisfied.	ENG_Y			0.515			
I felt powerful.	ENG_AB			0.683			
I felt proud.	ENG_AD			0.723			
I forgot everything around me.	ENG_H				0.743		
I lost track of time.	ENG_I				0.785		
I was deeply concentrated in the game.	ENG_R				0.538		

I lost connection with the outside world.	ENG_U				0.746		
I found it hard to get back to reality.	ENG_V					0.662	
I feel like I just can’t stop playing.	ENG_AE					0.699	
The game feels real.	ENG_AF					0.661	
I thought about other things.	ENG_D						0. 823
I felt bored.	ENG_J						0. 564

Table 14 - Rotated component matrix from the factor analysis of engagement items.

6.1.2 Results from the Final Questionnaire and Comparative Analysis across Groups

The results of the one-way ANOVA manage to examine if there are significant differences in the mean values of the four variables based on the students' degree program. Significant differences were found only for the variables DGBL and ENG, with a p-value of 0.028 and 0.043. For the other variables, no significant differences were found (p-value of 0.073 for TAM and 0.063 for SUS). Thus, the degree programme weakly influences DGBL and ENG, but not the other variables. Compared with the factors identified in the factorial analysis for TAM, SUS, and DGBL, the degree program has a weak influence on Factor 3, while it does not significantly affect the other factors. Of the six factors identified for the ENG variable, Factors 1 (Enjoyment and Involvement) and 4 (Flow) differed significantly across degree programmes (p-value = 0.046 and p-value = 0.024, respectively).

To understand the patterns and trends within each of the four variables, descriptive statistics were calculated for the mean, median and standard deviation of each variable on a scale of 0–100 (Table 15). The analyses were conducted by excluding the "N/A" (Not Applicable) responses, which were allowed for the DGBL and TAM items for those who felt they could not express an opinion on the use of DGBL in Mathematics. If there were only a few "N/A" responses, the average was calculated considering only the questions where "N/A" was not selected. However, if there were many "N/A" responses, the participant was excluded from the sample. In this way, the sample became 152 for TAM items, 148 for DGBL, and remained 156 for SUS and ENG.

		TAM	**SUS**	**DGBL**	**ENG**
N (sample)	Valid	152	156	148	156
	Missing	4	0	8	0
Mean		56.85	63.38	61.50	50.75
Median		55.00	62.00	61.43	48.75
Standard deviation		24.71	12.81	19.94	12.58

Table 15 – Descriptive statistics for TAM, SUS, DGBL and ENG variables.

Note that the mean, median and standard deviation were calculated based on the average of the items for each variable. The focus was on the four initial variables, particularly the SUS, TAM and GexpQ/Geng scales, to make meaningful comparisons with results from the literature. Table 15 shows that the mean values for TAM and ENG are not particularly high (56.85 and 50.75), while DGBL and SUS are around 60. The value of 56.85 suggests that, on average, respondents' scores related to TAM are slightly above median value of the scale (55.00). The standard deviation for TAM (24.71) is relatively high, meaning there is considerable variability in the responses and that individuals may have significantly different opinions or experiences regarding the acceptance of technology. With a mean engagement score of 50.75, participants' level of engagement is, on average, slightly above the midpoint of the 0-100 scale, indicating a moderate level of engagement. The median (48.75) is close to the mean, implying that the distribution of scores is relatively symmetrical. The standard

deviation (SD = 12.58) suggests that, although most participants reported similar levels of engagement, there were still some differences in how engaged individuals felt overall. The mean score of SUS (63.38) indicates that, on average, respondents rated the system's usability higher than the middle of the scale. The median value of SUS (62.00) is very close to the mean, which suggests that the distribution of responses is relatively symmetric. The standard deviation of SUS (12.81) is lower than that of TAM, indicating less variability among respondents' evaluations of system usability, with fewer extreme differences in opinions. The mean score of items of DGBL (61.50) suggests that, on average, respondents had a fairly positive view of the use of digital game-based learning in Mathematics. Although not as high as the TAM, the standard deviation (19.94) suggests that there is still a considerable spread in opinions or experiences related to the use of digital game-based learning in Mathematics. It is worth noting that the item DGBL_M "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*" was the one among the DGBL items that obtained the highest mean and median, 72.2 and 80 respectively. The least appreciated aspects were those related to personalised learning. Questions such as "*To what extent do you think the use of digital, game-based learning in Mathematics can promote a personalised learning approach?*" and "*How well does it meet your needs as a student?*" received lower scores, compared to other items in the DGBL scale, respectively 53.25 and 58.05.
Table 16 shows the comments obtained from the final questionnaire, which was answered on a voluntary basis, and their respective frequencies. The higher frequency of comments (6) indicates that the game is well-structured and interesting. However, there were also reports of difficulties understanding the rules and how the tool works. This is why the instructions were clarified further in the second version of the serious game.

Comments, suggestions, and difficulties encountered	Frequencies
A valid, well-executed and interesting activity.	6
Allow players to go back to previous questions and modify their answers.	3
Keep the data available for a longer time.	3
The game is well-structured, but some instructions need clarifying.	2
Increase the difficulty by including topics beyond just functions.	2
The answer selection is unclear.	2
Design a similar game that does not require prior knowledge of mathematics to make it more accessible.	1
Add some steps between points 11 and 12.	1
The game is too abstract.	1
Explain how to use the software more clearly.	1
It is difficult to remember the names of the suspects.	1
Some clues are unclear.	1
Difficulty understanding certain topics.	1

Table 16 – Qualitative comments and their frequencies.

Descriptive statistics were also calculated for the mean and standard deviation (SD) of each programme individually (Table 17). Descriptive statistics shows that students Strategic and Security Sciences degree programme expressed more consistent and favourable responses, since they have the highest mean scores for all four variables, suggesting a more positive perception towards technology acceptance, usability, game-based learning and engagement. The relatively low standard deviations in this group indicate consistent responses among students. In all degree programmes, TAM and ENG achieved the lowest average scores, confirming the impressions from the descriptive statistics for all students taken together. The Bachelor of Medical Radiology, Imaging and Radiotherapy Techniques group has the lowest mean score for all the four variables, for TAM items (48.17) and ENG (47.41), indicating a less level of engagement and less favourable perception of technology acceptance compared to the other groups. Prevention Techniques in the Environment and Workplaces and Biotechnology students show intermediate results, with slightly higher mean scores than the Radiology group but lower than Strategic and Security Sciences. Given that Biotechnology students represent the largest sample, their mean scores are closely aligned with the overall averages (Table 15), reflecting the general trend of the total sample.

DEGREE PROGRAMME		TAM	SUS	DGBL	ENG
Medical	Sample	20	20	19	20
Radiology,	Mean	48.17	60.60	53.88	47.41

Imaging and Radiotherapy Techniques	SD	24.31	12.19	20.12	12.58
Prevention Techniques in the Environment and Workplaces	Sample	18	19	17	19
	Mean	61.48	60.95	63.36	49.08
	SD	26.18	16.86	19.42	12.94
Biotechnology	Sample	104	106	102	106
	Mean	56.26	63.40	61.10	50.70
	SD	24.70	11.88	19.80	12.30
Strategic and Security Sciences	Sample	10	11	10	11
	Mean	72.00	72.55	76.91	60.28
	SD	15.89	12.20	14.78	11.50

Table 17 – Descriptive statistics for the variables according to the degree programme.

On average, female participants gave more positive responses for the TAM, SUS, DGBL and ENG variables (Table 18), achieving scores higher than the overall mean (Table 15). This result may be explained by the fact that, being generally more experienced with high-quality commercial video games, male participants may have had more points of comparison and therefore evaluated the educational game more critically. Further investigation was conducted to establish whether male participants played more frequently than female participants. A Kruskal-Wallis test was conducted to compare videogame playing frequency between female and male students. The results showed a significant difference between groups reporting a higher videogame playing frequency for male than female students. The sample size is one smaller than expected because one participant preferred not to indicate the gender.

Variable	Genre	N	Mean	Standard deviation
TAM	Female	104	58.27	25.38
	Male	47	53.70	23.41
	Total	151	56.85	24.71
SUS	Female	106	65.55	12.80
	Male	49	58.69	11.79
	Total	155	63.38	12.81
DGBL	Female	101	62.32	20.80
	Male	46	59.83	18.24
	Total	147	61.50	19.94
ENG	Female	106	52.38	13.37
	Male	49	47.26	10.07
	Total	155	50.75	12.58

Table 18 - Descriptive statistics according to genre of participants.

A Kruskal-Wallis test was performed on the factors outlined in the final questionnaire, categorized by gender. Male participants tended to achieve higher scores in factors Factor1 “The impact of digital game-based learning on the Mathematics learning experience" (p-value= 0.29) and $Factor_{ENG}1$ “Enjoyment and involvement” (p-value= 0.046), while female participants showed significantly higher scores in factor $Factor_{ENG}6$ “Disengagement and boredom” (p- value < 0.01). Another Kruskal-Wallis test was conducted to examine whether gameplay frequency had a significant effect on the factors outlined in the final questionnaire. The analysis revealed no significant differences.

After exploratory factor analysis, descriptive statistics were also computed for the extracted factors: four for the DGBL, TAM and SUS variables, and six for the ENG variable. Regarding the DGBL, SUS and TAM-related factors, the highest mean scores were obtained for Factor 3 “The complexity of the game, possible difficulties and need for support” (Table 19), indicating that students found the game technically accessible and easy to use (note that questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores). Participants also reported a moderately positive perception of the impact of DGBL on their Mathematics learning experience (Factor 1). Regarding Factor 2, “Perceived effectiveness for learning mathematics”, a rating slightly above the median scale value was obtained. However, there was a fair amount of dispersion, suggesting that some students found the serious game effective for learning, while others did not. Factor 4 "Game experience and serious game appeal" showed that, although students generally appreciated the overall experience, they did not find the game particularly appealing. The engagement factors revealed a mixed experience. $Factor_{ENG}$ 1 "Enjoyment and involvement" received a moderate evaluation, whereas $Factor_{ENG}$ 3 "Competence and achievement" and $Factor_{ENG}$ 4 "Flow" were rated below the midpoint of the scale, suggesting that students did not always feel competent or rewarded for their performance and did not experience a deep sense of immersion. Particularly low values were observed for the "Loss of control and game realism" factor. This may be due to a low level of immersion, or possibly because the corresponding items were misinterpreted. These items originated from Factor 5, where a grouping of seemingly unrelated items had previously been identified. Conversely, higher values were obtained for Factor 2 "negative feelings and pressure", indicating that students did not experience negative emotions or pressure during the game, as well as for Factor 6 "disengagement and boredom", suggesting that participants did not report feelings of boredom or disengagement (note that questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores). Overall, the data indicate that although students appreciated some aspects of the game-based learning experience, but the level of immersion and flow remained limited. The descriptive statistics of the extracted factors are consistent with those obtained for the original variables. Specifically, higher mean values were found for usability and for the impact of digital game-based learning on Mathematics learning, whereas lower values emerged for the engagement-related dimensions. No correlations were found between students' upper secondary school grades and the variables and factors of the questionnaire.

Items	**Factor 1** "The impact of digital game-based learning on the Mathematics learning experience"	**Factor 2** "Perceived effectiveness for learning Mathematics"	**Factor 3** "The complexity of the game, possible difficulties and need for support"	**Factor 4** "Game experience and serious game appeal"	**$Factor_{ENG}$ 1** "Enjoyment and involvement"	**$Factor_{ENG}$ 2** "Negative feelings and pressure"	**$Factor_{ENG}$ 3** "Competence and achievement"	**$Factor_{ENG}$ 4** "Flow"	**$Factor_{ENG}$ 5** "Loss of control and game realism"	**$Factor_{ENG}$ 6** "Disengagement and boredom"
Mean	**60,86**	**53,06**	**76,85**	**49,47**	**54,18**	**81,12**	**43,26**	**34,39**	**18,68**	**69,68**
Median	**60,00**	**50,00**	**80,00**	**52,00**	**51,43**	**86,67**	**44,00**	**30,00**	**13,33**	**70,00**
Standard deviation	**22,36**	**25,06**	**15,87**	**19,42**	**22,42**	**18,36**	**21,32**	**24,69**	**21,64**	**21,05**

Table 19 - Descriptive statistics for the extracted factors.

6.1.3 Changes in Attitudes towards Mathematics from the Initial to the Final Questionnaire

Changes in students' attitudes towards Mathematics were evaluated by comparing data from the questionnaires completed before and after the game. The reliability analysis conducted on all 42 items related to attitudes towards Mathematics yielded very high Cronbach's alpha values for both QI (0.870) and QF (0.888). These values indicate excellent internal consistency for both questionnaires. To further examine the internal consistency of each subscale, Cronbach's Alpha was calculated for the eight theoretical dimensions identified a priori (Table 20). A low Cronbach's alpha value was obtained for the dimension of "Perception of Mathematics" both in QI (0.484) and QF (0.548). Reliability analysis suggests that Cronbach's alpha would be higher if the item "I keep studying until I understand mathematics material" were deleted from the "Perception of Mathematics" dimension, with values of around 0.700 in QI and around 0.595 in QF. A further consideration suggests that this item is more appropriately included within the "Persistence in Mathematics" dimension.

Indeed, when this item is included in the "Persistence in Mathematics" dimension, the resulting Cronbach's alpha value is 0.719 for QI and 0.706 for QF. It means that the new item is consistent and it contributes positively to the overall measurement of the construct. This will also be taken into consideration in the second experiment. With the exception of the "Development of Critical Thinking" and "Persistence in Mathematics" dimension, high Cronbach's alpha values were obtained for the other dimensions (Table 20), which confirms strong item coherence.

Item Group (Dimension)	**Number of items**	**Cronbach's Alpha (QI)**	**Internal Consistency (QI)**	**Cronbach's Alpha (QF)**	**Internal Consistency (QF)**
Perception of Mathematics	3	0.484	Low	0.548	Low
Interest and motivation in Mathematics	4	0.841	Very good	0.800	Very good
Math anxiety	4	0.802	Very good	0.786	Good
Persistence in Mathematics	6	0.662	Discrete	0.634	Moderate
Persistence in problem solving	4	0.744	Good	0.723	Good
Development of critical thinking	8	0.601	Discrete	0.748	Good
Perception of the course	9	0.870	Very good	0.865	Very good
Extrinsic motivation	4	0.822	Very good	0.777	Good

Table 20 - Calculations of Cronbach's alpha for QI and QF regarding attitudes towards Mathematics.

The results of the Wilcoxon signed-rank test for paired samples are presented for each of the eight measured dimensions for which significant values were obtained. The QI and QF responses of 139 students were compared. This included 102 students from the Biotechnology programme, and 37 students from the Prevention Techniques in the Environment and Workplaces programme, as well as the Medical Radiology, Imaging and Radiotherapy Techniques programme. These results are organised by degree programme. For the Biotechnology programme, the results indicate that a marked improvement was observed in the reduction of math anxiety (-0.96) and extrinsic motivation (-0.30) from pre- to post-intervention. As shown in Table 21, very low p-values were obtained (except for the second question), indicating highly significant statistical results. Note that the questions are phrased in reverse (a positive outcome corresponds to low response values: 1-2). Students reported feeling less anxious about Mathematics and experiencing lower levels of extrinsic motivation (e.g. the desire to achieve a good grade) when they engaged with Mathematics in the context of the game. As with the Biotechnology degree programme, also students from the Prevention Techniques in the Environment and Workplaces and the Medical Radiology, Imaging and Radiotherapy Techniques programmes demonstrated significant improvements, with notable reductions in math anxiety (-0.58) and extrinsic motivation (-0.45). An item-by-item comparison (Table 21, Table 22) shows that the greatest decrease was found in both samples for the fourth question, "When I take a math test I think of the consequences of failing / While I was playing, I thought about the consequences of failing"(a decrease of at least one unit), and for the question "I get very tense when I have to do maths homework/ I was very tense while playing the serious game"(a decrease of about one unit). In the final questionnaire, this question was adapted to the context of the game. The results on math anxiety are coherent with the descriptive statistics results, which revealed that students' scores were higher for $Factor_{ENG}$ 2 - "Negative feelings and pressure". This indicates that students felt less pressure and did not experience negative emotions during the game.

Regarding the dimension "Persistence in Mathematics", the results indicate a decline in persistence (-0.48) from before to after playing the serious game for Prevention Techniques in the Environment and Workplaces and Medical Radiology, Imaging and Radiotherapy Techniques sample. Note that the questions are phrased directly in this case; therefore, low scale values (1-2) correspond to negative responses. Consequently, a decrease indicates a more negative evaluation in the QF than in the QI. It would be interesting to verify whether the same occurs in the second experiment. The other dimensions remained almost identical.

	Math anxiety For each statement, indicate a response from 1 to 6, where 1 = "Not at all true for me" and 6 = "Extremely true for me."							
Item	**QI-MA_A**: I get very tense when I have to do Mathematics homework	**QF-MA_A**: I was very tense while playing the serious game	**QI-MA_B**: I feel helpless when doing a Mathematics problem	**QF-MA_B**: I feel helpless when doing a Mathematics problem	**QI-MA_C**: When I take a test I think about how poorly I am doing compared with other students	**QF-MA_C:** While I was playing, I thought about how badly I was doing compared to the other students	**QI-MA_D**: When I take a math test I think of the consequences of failing	**QF-MA_D:** While I was playing, I thought about the consequences of failing
Mean	3.15	2.33	2.72	2.51	2.83	1.84	3.59	2.30
Standard deviation	1.44	1.34	1.43	1.26	1.57	1.16	1.52	1.36
Increase/ Decrease in the mean	-0.82		-0.21		-0.99		-1.29	
p-value	0.0000108		0.0411		0.0000000482		0.000000000144	

Table 21- Wilcoxon signed-rank test for paired samples: math anxiety (Biotechnology degree programme).

	Math anxiety For each statement, indicate a response from 1 to 6, where 1 = "Not at all true for me" and 6 = "Extremely true for me."							
Item	**QI-MA_A:** I get very tense when I have to do Mathematics homework	**QF-MA_A:** I was very tense while playing the serious game	**QI-MA_B:** I feel helpless when doing a Mathematics problem	**QF-MA_B**: I feel helpless when doing a Mathematics problem	**QI – MA_C:** When I take a test I think about how poorly I am doing compared with other students	**QF-MA_C:** While I was playing, I thought about how badly I was doing compared to the other students	**QI-MA_D:** When I take a math test I think of the consequences of failing	**QF-MA_D:** While I was playing, I thought about the consequences of failing.
Mean	3.59	2.73	2.92	2.95	3.000	2.51	3.40	2.40

Standard deviation	1.72	1.30	1.55	1.39	1.58	1.27	1.59	1.42
Increase/ Decrease in the mean	-0.86		0.03		-0.49		-1.000	
p-value	0.0000108		0.47 (> 0.1)		0.037		0.00012	

Table 22- Wilcoxon signed-rank test for paired samples: math anxiety (Prevention Techniques in the Environment and Workplaces and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes).

6.1.4 Exam Results: a Comparison of the Intervention and Control Groups

The exam test results (administered using an automatic assessment system) for Biotechnology degree programme from the 2021-22 academic year (when the serious game was not used) were compared with those from the 2023–24 academic year (when the serious game was played). The control group consist of 72 students who took the exam in the 2021–22 academic year. The experimental group consist of 95 students who played the serious game at least once and took the exam in the 2023-24 academic year, and who had completed the serious game at least once. The two questions considered are similar to those analysed in previous years, focusing on functions of a real variable - the same topic covered by the Murderer game. Each question was scored on a scale from 0 to 4. Table 23 and Table 24, as well as Fig. 3, show a comparison of the scores obtained in the two academic years for the question about the function study. As shown in Table 23, both the mean and median scores increased substantially, moving from 2.25 (mean and median) in 2021-22 to 2.96 (mean) and 2.97 (median) in 2023-24. The decrease in standard deviation from 1.40 to 0.89 indicates that the scores in 2023-24 were more closely grouped around the mean, confirming that improvement was widespread across the sample. There was also an increase in the percentage of students achieving the median score of 2 or higher, rising from 64% to 93%.

	2021-22 academic year	**2023-24 academic year**
Mean score	2.25	2.96
Standard Deviation	1.40	0.89
Median score	2.25	2.97
Percentage of students obtained a score above the median score (≥2)	64%	93%

Table 23- Comparison of the exam results of students on the Biotechnology degree programme between the control and experimental groups for the question related to the function study.

Table 24 shows that the percentage of students obtaining a score below 1.5 was significantly lower in 2023-24 than in 2021-22. A similar, but less pronounced, trend can be observed for the percentage of students who scored in the 1.6-2.3 range. It is notable that the percentage of students in the 3.6-4.0 range was more than double in 2023-24 than in 2021-22. However, it is worth noting that more students achieved the maximum score (4.0) in 2021-22 than in 2023-24 (25% vs. 10%).

	Range 0-1.5	**Range 1.6-2.3**	**Range 2.4-3.5**	**Range 3.6-4.0**
2021-22 academic year	36%	18%	21%	25%%
2023-24 academic year	5%	12%	45%	38%

Table 24- Number of students in each score range for the question related to the function study.

Fig. 13 represents cumulative distribution plot of exam scores, and it shows, for each score from 0 to 4, the proportion of students who achieved *at least* that score. It can be observed that in the 2023-24 academic year, the scores obtained for the question related to the function study are significantly higher than those from 2021-22, as the red line corresponding to the 2023-24 results lie above the blue one for 2021-22. The final reversal is because a higher proportion of students achieved the maximum score in 2021-22 (25%) than in 2023-24

(9%). Overall, Fig. 13 clearly shows that students performed better on this question in the 2023–24 academic year than in the 2021–22 academic year.

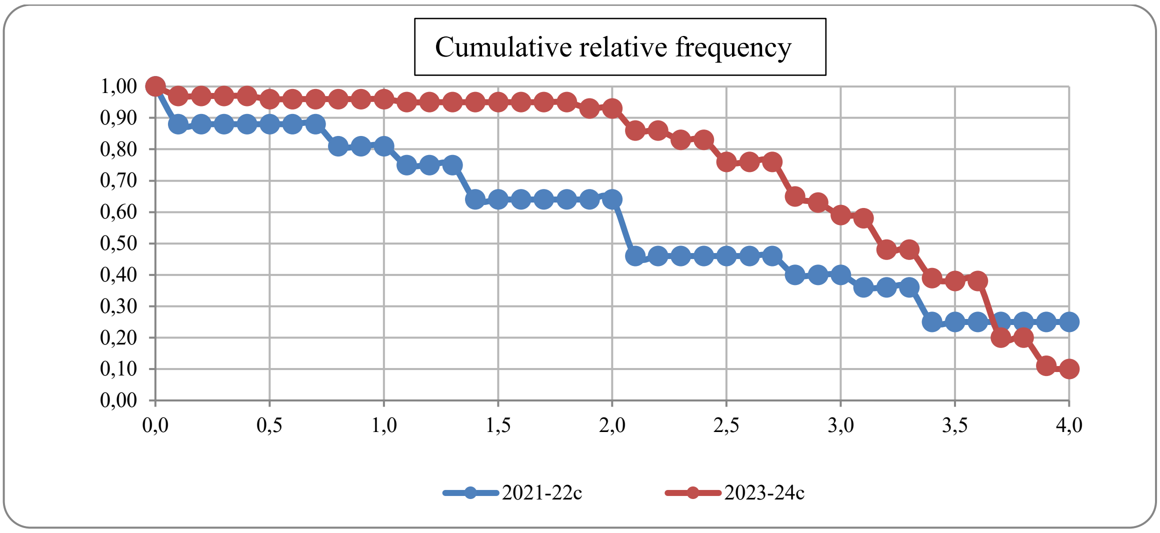

Figure 13 - Cumulative distribution plot of exam scores for the for the question related to the function study.

Regarding the second question, which was about the properties of functions based on their graphs, the average score increased by about 0.1 from the 2022-23 academic year to the 2023-24 academic year. Meanwhile, the median and standard deviation remained unchanged. There was also an increase in the percentage of students achieving the median score of 2 or higher, rising from 55% to 61%. The percentage of students who obtained a score below 1.5 was similar in both academic years, though slightly lower in 2021-22 (Table 25). A higher percentage of students obtained a score in the 1.6-2.3 range in 2021-22 (38% vs. 22%). As with the previous question, the trend reverses in the 2.4-4.0 score range, with a greater number of students obtaining a score above 2.4 in 2023-24 than in 2021-22. As with the previous question, the cumulative distribution plot of exam scores (Fig. 14) shows that, although less evident, the line for the 2023-24 results is equal to or above that for the 2021-22 results. This pattern suggests that, overall, students in 2023-24 achieved slightly higher scores, with a subtle shift in performance distribution towards better results.

	Range 0-1.5	**Range 1.6-2.3**	**Range 2.4-3.5**	**Range 3.6-4.0**
2021-22 academic year	27%	38%	28%	7%%
2023-24 academic year	29%	22%	39%	10%

Table 25- Number of students in each score range for the question concerning the properties of functions based on their graphs.

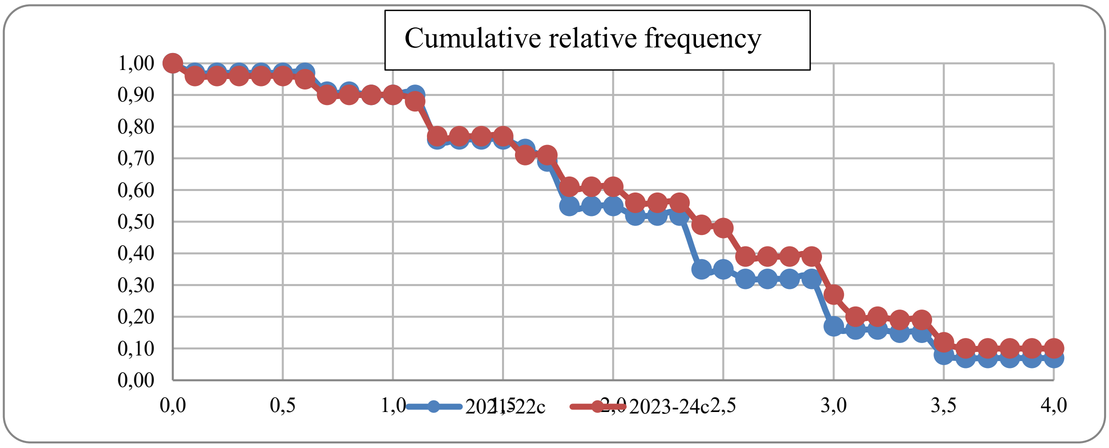

Figure 14 - Cumulative distribution plot of exam scores for the for the question concerning the properties of functions based on their graphs.

The same comparison was made for students of Medical Radiology, Imaging and Radiotherapy Techniques degree programme. The exam test results (administered using an automatic assessment system) from the 2022-23 academic year (when the serious game was not used) were compared with those from the 2023-24 academic year (when the serious game was played). The control group consists of 18 students who took the exam in the 2022-23 academic year. The experimental group consists of 18 students who played the serious game at least once and took the exam in the 2023-24 academic year, and who had completed the serious game at least once. Two questions related to functions in a real variable- a topic covered by the serious game- were taken into consideration. In particular, the first question considers elementary functions, while the second focuses more specifically on the function study. First question was scored on a scale from 0 to 4.5 and second question on a scale from 0 to 6. As shown in Table 26, there was a substantial increase in both the mean and median scores, rising from 1.48 and 0.90 in 2022-23 to 2.28 and 2.25 in 2023-24. The percentage of students achieving the median score or higher increased threefold, rising from 11% to 33%.

	2022-23 academic year	2023-24 academic year
Mean score	1.48	2.28
Median score	0.90	2.25
Standard Deviation	1.40	1.18
Percentage of students obtained a score above the median score	9%	33%

Table 26- Comparison of the exam results of students on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme between the control and experimental groups for the question related to the elementary functions.

Fig. 15 represents cumulative distribution plot of exam scores and it shows, for each score from 0 to 4.5, the proportion of students who achieved *at least* that score. Looking at the two curves, the one corresponding to 2023-24 consistently stays above the one for 2022-23. This indicates that a greater proportion of students achieved at least that result in 2023-24 compared to the previous year. This is particularly noticeable for intermediate scores (between 1 and 3), where the red curve remains significantly above the blue one. This suggests that many more students achieved medium-to-high results in 2023-24, and that the cumulative decline towards lower scores was less pronounced. Overall, Fig. 15 clearly shows that students performed better in the 2023-24 academic year than in 2022-23.

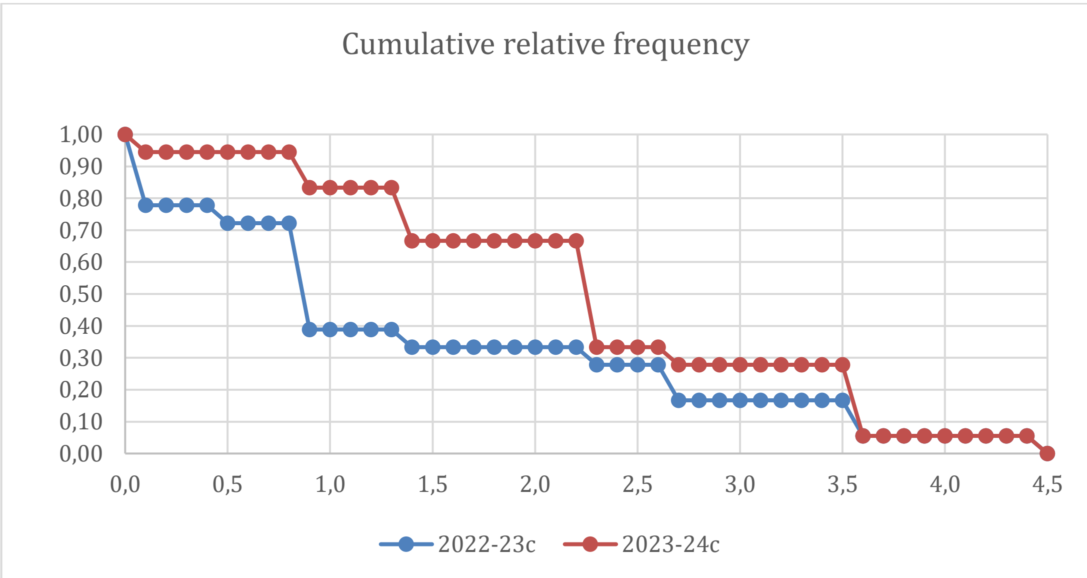

Figure 15 - Cumulative distribution plot of exam scores for the for the question related to the elementary functions.

Regarding the question on the function study, no improvement in exam scores was observed for students in the 2023-24 academic year compared with those in 2022-23. In fact, looking at the mean and median values (Table 27), there is a slight decline. A similar pattern appears in the cumulative relative frequency plot (Figure 16), where the gap between the two curves is even less pronounced, indicating that the scores for the function–study question are largely comparable. The only noticeable difference occurs between scores of 1 and 2, where the red line for 2023–24 is above the blue line for 2022–23, indicating that more students achieved a score of at least 1 in the latter year.

	2022-23 academic year	2023-24 academic year
Mean score	2.83	2.64
Median score	3.00	2.40
Standard Deviation	2.34	1.83
Percentage of students obtained a score above the median score	50%	33%

Table 27- Comparison of the exam results of students on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme between the control and experimental groups for the question related to function study.

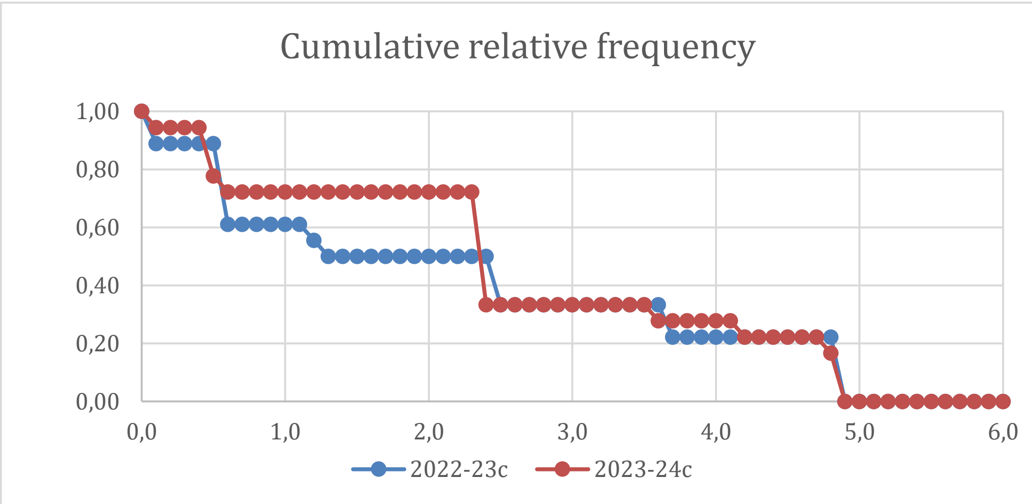

Figure 16 - Cumulative distribution plot of exam scores for the for the question related to the function study.

Spearman correlation analysis of students enrolled on the Biotechnology degree programme revealed a significant weak positive association between the average score of the two exam questions considered and Factor 2 "Perceived effectiveness for learning Mathematics" (r = 0.253, p-value = 0.015). No significant correlations were found for the remaining factors and variables (DGBL, TAM, SUS, ENG). For students enrolled on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme, Spearman's correlation analysis revealed no significant associations between exam scores and responses to the final questionnaire.

6.1.5 Results from Game Sessions of the Murderer Game

In the 2023-24 academic year, a total of 417 gameplay attempts were made by 156 university students who also completed both the initial and final questionnaires. The average number of attempts per student was 2.67, with a standard deviation of 1.59. Considering that the scale ranges from 0% to 100%, the average score obtained is 70.27%.

A statistically significant, though weak, positive correlation was found between mean scores obtained in the various attempts at the serious game and mean scores achieved in the two exam questions related to functions (the game's topic). This correlation was based on data from game sessions involving 95 Biotechnology degree programme students who took the exam. The correlation was found to be statistically significant (p-value = 0.011), though weak (r = 0.259). This suggests that students who performed better in the serious game also tended to perform slightly better in the exam. A second analysis was conducted to determine whether the average serious game score was more strongly related to one of the two exam questions. A stronger correlation coefficient (r = 0.305, p-value = 0.003) was found for the function study question, whereas a very weak, non-significant correlation (r = 0.164, p-value = 0.112) was observed for the question concerning function properties based on graphs. Therefore, as the number of attempts increases, the score on the function study question tends to increase slightly. This result is consistent with the improvements observed in the experimental group compared to the control group on the function study question (Fig. 3). No such conclusion can be drawn for the question concerning the properties of functions based on their graphs. The correlation between exam results and the number of attempts made on the serious game was also studied. However, no significant correlation was found, so it cannot be concluded that those who made more attempts performed better in the exam.

Correlation analyses were conducted between the number of attempts completed in the serious game and the responses of the 156 participants to the final questionnaire. No significant relationships were found. Specifically, the number of attempts in the serious game did not significantly correlate with mean scores on the DGBL, TAM, SUS and ENG scales. Similarly, no significant correlations were found between the number of attempts and any of the factors identified in the final questionnaire. Overall, these results suggest that participants' performance in terms of the number of attempts made in the serious game was not associated with their reported perceptions of DGBL, usability evaluations, engagement levels or technology acceptance ratings.

Regarding the correlation between the average of scores obtained in the various attempts in the serious game and the 156 students' responses to the final questionnaire, it emerges that only Factor 3 ("The complexity of the game, possible difficulties and need for support") shows a moderate, statistically significant, positive linear correlation with the average of scores obtained in the various attempts in the serious game (r= 0.249, p-value= 0.03). It is reasonable to assume that those who achieved higher average scores perceived the technical and conceptual complexity of the game as lower. All other factors were found to have no correlation, or a weak correlation that was not significant at the standard 0.05 level. Consistently, analysing the correlation with the four variables of the final questionnaire (DGBL, SUS, TAM, ENG) it was found that the only element showing a statistically significant correlation with the average scores obtained in the various attempts in the serious game was the SUS variable (r=0.185, p-value=0.021). This suggests that as the average score increases, there is a significant, but weak, tendency for usability to increase as well.

A Spearman correlation test was also performed for students on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme to investigate whether there was an association between the exam scores relating to the two questions on functions of a real variable, and the number of attempts and average scores obtained in the serious game. No significant correlations were found, most likely due to the small sample size.

6.2 Revising the Murderer Game following the Pilot Study and Designing the Alien Game

After analysing the results of the pilot study, the insights gained were applied to revising the Murderer game and designing the Alien game.

The following modifications were implemented:

- Scene adjustments: some details in the scenes were changed. For example, the sizes of the characters were adjusted so that they were no longer the same height as certain objects, making the game more realistic. Additionally, in the responses related to step 4, an image originally indicated a measurement of 20 metres in an intermediate position between the point coordinate and the vector. This has been corrected to make it clear that the measure refers specifically to the vector.
- Scoring system and win conditions: a scoring system and a threshold for winning were introduced. In the first version, players could win by collecting 75% of the points without identifying the culprit. In the revised version, however, scores were weighted so that correctly identifying the culprit is now necessary to win. Specifically, 20 out of 63 points are awarded for identifying the culprit at the last step. However, correctly guessing the culprit alone is not enough: players must also achieve a total score of at least 48 points (75% of the total). The aim is to give more weight to the knowledge required to identify the culprit than to luck, to avoid identifying the culprit by chance.
- Integration of Gen-AI tools: prompts for use with ChatGPT were suggested to help with unclear topics, enabling players to solve the serious game with the aid of an external tool. The goal of this integration

is not simply to provide the game solution, since the game context is specific, but to encourage critical and responsible use of AI in study, with outputs that should be critically evaluated. Attempts were made to integrate GenAI in view of the limited personalisation results observed in the pilot study. The least appreciated aspects were indeed those related to personalised learning. Questions such as "*To what extent do you think the use of digital, game-based learning in Mathematics can promote a personalised learning approach?*" and "*How well does it meet your needs as a student?*" were given lower scores, ranging from 53.25 to 58.05.

- Improved spoken feedback: feedback is now provided by real voices, whereas in the first version it was automatically generated, making it feel more mechanical and less realistic.
- Clearer instructions: based on comments regarding the previous instructions, they were revised to be clearer, both on the Moodle page and within the game itself. The instructions have been organised into a step-by-step format for better structure, with real screenshots illustrating each step. Key information, such as the number of points required to win, has been highlighted in bold or underlined for emphasis.
- Branched structure: for the Alien game differentiated steps were introduced based on the user's response. Losing led to an intermediate guided step to prompt reflection on the incorrect choice, while winning led to an intermediate step for further exploration, such as introducing a variation of the problem scenario. This modification was also introduced in response to the less positive feedback regarding the contribution of the serious game to personalised learning. The intention was to implement a branched structure, as this could provide a higher degree of personalisation by allowing for either further exploration or guided steps towards the solution, depending on the response provided.

6.3 Revising the Initial and Final Questionnaires for the Second Experimentation of the Serious Games

There were no substantial changes to the QI. The only modification was the removal of the questions about the impact of the DGBL on Mathematics learning and technology acceptance (DGBL and TAM items), which had been included in the QI in the first experiment for comparison with the QF. However, these questions were later deemed irrelevant, as students had never previously used this approach and would therefore have responded randomly.

Some modifications were made to the QF based on the results of the first experiment, primarily due to the recognition that the questionnaire was rather lengthy. Thus, items that were similar, redundant or unclear were removed. The changes made to the different groups of items are summarised below:

- DGBL: the DGBL_C item "*To what extent do you think that the use of DGBL (i.e., the use of games for learning, such as the game the Murderer game) in Mathematics can stimulate your participation?*" was removed, as it was very similar to the DGBL_B item "*To what extent do you think that the use of DGBL… can increase your engagement*?". Analysis showed that the mean responses to the two questions were identical and that 83% of students gave the same answer to both. Therefore, it was decided to remove the first question. Moreover, DGBL_N item "*To what extent do you think that the use of DGBL… can increase your participation in the learning process*" provides an accurate measure of the participation dimension.
- TAM: similarly, the TAM_C item "*To what extent do you think that the use of DGBL… can increase my productivity*" from the TAM model was removed. This decision was made because this item is similar to TAM_D item "*To what extent do you think that the use of DGBL… can increase my effectiveness*". Moreover, analysis showed that the mean responses to the two questions were quite identical and that 75% of students gave the same answer to both. Therefore, it was decided to remove the first question. The original TAM_C item, "*Increase my productivity*", has been replaced with two more specific items. One item is TAM_G which assesses the extent to which using DGBL in Mathematics can develop students' critical thinking skills (i.e. their ability to analyse facts, evidence and observations to make reasoned judgements). The other item (TAM_H) assesses the extent to which it can develop problem-solving skills. This change was made to provide a more precise evaluation of the educational impact of the game by focusing on the development of cognitive skills rather than a

generic notion of productivity. It also allows to explicitly ask about the acquisition of these two transversal competencies, which are not limited to Mathematics.

- SUS: the items SUS_F ("I found the game very difficult to use") and SUS_G ("I felt very confident in using the game") were removed, as their content was partially redundant with previous SUS items that already investigated the difficulties encountered in using the serious game. Two additional items were included in the SUS items: SUS_M "The contents were easy to read" and SUS_N "The contents were easy to understand". These items were specifically designed to evaluate the accessibility of the game content for students with specific learning disorders. The decision to add these items was also motivated by informal feedback collected during the pilot study, in which both instructors and students reported that students with specific learning disorders particularly appreciated the serious game. Therefore, an explicit measure of this aspect was introduced in the second experimentation. Furthermore, the SUS_C item was rephrased positively, changing from "I had difficulty playing because the tool is complex to use" to "I found the game easy to use".
- ENG: the scale was reduced because several items focused specifically on the game itself rather than its educational use. As the main objective was to measure learning engagement through the game, only the most relevant items were retained, resulting in a shorter scale. These modifications were also guided by the results of the pilot study. The EFA of the ENG scale revealed a six-factor solution, which appeared disproportionately complex compared with the factor structures observed for the DGBL, TAM, and SUS scales. $Factor\_{ENG}4$ "Flow" and $Factor\_{ENG}5$ "Loss of control and game realism" showed strong conceptual overlap, since both dimensions capture deep immersion in the game. According to student feedback and informal interviews, the previous version contained many items, which made this section too long and imposed a considerable cognitive load on participants. To improve the efficiency of the instrument and reduce respondent fatigue, items showing conceptual redundancy or limited added value were removed. This meant that the number of scales had to be reduced. Furthermore, students' answers suggested that the meaning of certain items was not clearly perceived (for example "I found it impressive"); consequently, these items were removed from the revised version of the questionnaire. The engagement scores, particularly those relating to Factors 3, 4 and 5, were not particularly high. The length of the questionnaire may have caused student fatigue, which may explain this. Consequently, the questionnaire was shortened to enhance clarity and minimise respondent burden. The item ID used in the questionnaire design description and reported in Table 8, paragraph 5.1, was retained. Table 28 shows the items maintained for the ENG variable after the pilot study. Some items - ENG_F, ENG_K, ENG_P, ENG_Q and ENG_X - did not load clearly on any single factor. These items may be ambiguously worded or may not effectively measure a specific latent dimension. All these items were removed except for ENG_F and ENG_P, which were retained because they measure aspects that it was considered essential to include.

An open-ended mandatory question was added to the final questionnaire to explore students' perceptions of the potential of the serious game for learning Mathematics. Additionally, as GenAI elements were introduced in the second version of the Murderer game and in the Alien game, students were asked to indicate whether they had used GenAI tools during the activity, which tools they had used, and for what purposes. In Table 28, the second version of the questionnaire after the pilot study is presented; it should be noted that the questions from the initial questionnaire that remained unchanged are also included.
Since several modifications were made to the final questionnaire, its psychometric properties were re-evaluated. Specifically, reliability analyses, exploratory and confirmatory factor analyses were conducted to assess internal consistency and to verify whether the original factor structure was still supported by the data of the second study conducted during 2024-2025 academic year.

To what extent do you think that the use of digital game-based learning (i.e. the use of games for learning, such as the Murderer game) in Mathematics can	***ID***
Enhance learning.	DGBL_A
Increase your engagement.	DGBL_B

Promote an inclusive learning approach.	DGBL_D
Promote a cooperative learning approach.	DGBL_E
Promote a personalized learning approach.	DGBL_F
Stimulate your interest.	DGBL_G
Stimulate interaction.	DGBL_H
Enable skills development.	DGBL_I
Meet your needs as a student.	DGBL_L
Reduce of the gender gap in scientific disciplines.	DGBL_M
Increase your participation in the learning process.	DGBL_N
Improve your perception of Mathematics.	DGBL_O
Reduce math anxiety.	DGBL_P
Allow you to achieve my learning goals faster.	TAM_A
Improve your performance.	TAM_B
Increase your effectiveness.	TAM_D
Make Mathematics easier to learn.	TAM_E
Speed up learning Maths.	TAM_F
Develop your critical thinking skills (i.e. their ability to analyse facts, evidence and observations to make reasoned judgements).	TAM_G
Develop your problem-solving skills.	TAM_H
Based on your experience as a player of the Murderer game, to what extent do you agree with the following statements?	ID
I think I would like to use this game a lot.	SUS_A
The content and topics of the game are complex.	SUS_B
I found the game easy to use.	SUS_C
I think I would need the support of a technician to use this game.	SUS_D
I found that the different features of this game were well integrated.	SUS_E
This game has a lot of things that interest me.	SUS_H
This game needs more introductory explanations.	SUS_I
The scenes in this game are very attractive.	SUS_L
The contents were easy to read.	SUS_M
The contents were easy to understand.	SUS_N
Thinking about your experience as a player of the Murderer game, indicate how you felt during the game	
I was interested in the game's story.	ENG_A
I thought it was fun.	ENG_B
I thought about other things .	ENG_D
I felt competent.	ENG_E

I thought it was hard.	ENG_F
It was aesthetically pleasing.	ENG_G
I felt bored .	ENG_J
I enjoyed it.	ENG_M
I was fast at reaching the game's targets.	ENG_N
I felt pressured.	ENG_O
I felt challenged.	ENG_P
I was deeply concentrated in the game.	ENG_R
It felt like a rich experience.	ENG_T
I felt disoriented .	ENG_Z
The game feels real.	ENG_AF

Table 28- Second version of the final questionnaire after the pilot study.

6.4 Results from the Second Experimentation of the Murderer Game in Higher Education

The study, conducted from October 2024 to November 2025 (including the final examination session of the academic year), involved 204 university students. The sample comprises students who completed the initial questionnaire before playing the serious game, then played the game at least once and filled out the final questionnaire afterwards. As three students provided too many "N/A" responses to the DGBL and TAM-related items in the final questionnaire, their responses concerning these two variables were not included in the analysis. The sample consists of 110 females and 94 males. This sample includes 163 first-year undergraduate students enrolled in a non-mathematical scientific degree programme, 35 undergraduate students in the bachelor's programme in Strategic and Security Sciences, and 6 postgraduate students in the master's programme in Strategic and Security Sciences. There are 22 students on the Medical Radiology, Imaging and Radiotherapy Techniques degree course, 117 on the Biotechnology degree course, 24 on the Prevention Techniques in the Environment and Workplaces degree course, and 35 on the Strategic and Security Sciences degree course.

6.4.1 Results of the Reliability and Factor Analyses of the Questionnaires

Table 29 shows the results of Cronbach's alpha calculation for the single variables (DGBL, TAM, SUS and ENG) of the final questionnaire for the second experimentation of the Murderer game. As in previous study, very high values were obtained for DGBL, TAM above 0.9 and a slightly lower value was obtained for SUS (0.716). However, the value is still considered acceptable. Compared to the previous experiment, where Cronbach's alpha values were around 0.95, they are now approximately 0.90. This reduction may indicate that redundant items have been removed. Indeed, compared to the final questionnaire used in the previous experiment, one TAM item was eliminated, and two SUS items were replaced. These were specifically those regarding the ease of reading and interpreting the game's content for students with specific learning disorders. Compared to the previous experiment, in which a Cronbach's alpha value of around 0.9 was obtained for the ENG items, this second study produced a lower value of approximately 0.2 (0.716). This is reasonable, given that the number of items was reduced by more than half. Cronbach's alpha was calculated for 45 items when the four variables DGBL, TAM, SUS, and ENG were combined, and the questionnaire was found to be highly reliable (Cronbach's alpha = 0.941), although slightly lower than in the previous study (0.951). Once again, this slight decrease may be explained by the reduced number of items, which naturally lowers the alpha value. The Cronbach's alpha value remains stable for each item, with values very similar to the Cronbach's alpha identified, in some cases, removing the item even results in a lower alpha value. This indicates that none of the items have a negative effect on the scale's internal consistency, and that removing individual items would not substantially improve the scale's overall reliability.

Variable	N. items	Cronbach's alpha value
DGBL	13	0.904
TAM	7	0.894
SUS	10	0.716
ENG	15	0.782
DGBL, TAM, SUS and ENG combined	45	0.941

Table 29 - Results of Cronbach's alpha calculation for the second experimentation.

Table 30 shows the results of the Kolmogorov-Smirnov test and the Shapiro-Wilk test for normality of the four variables, performed on the average of the items of TAM, SUS DGBL and ENG. Results indicate that DGBL and SUS do not significantly deviate from a normal distribution ($p > 0.05$), and can therefore be considered normally distributed. In contrast, ENG and TAM show significant departures from normality ($p < 0.05$), indicating that these variables are not normally distributed.

	Kolmogorov-Smirnov test			Shapiro-Wilk test		
	p-value	Statistic	gl	p-value	Statistic	gl
TAM	0.038	0.065	201	0.030	0.985	201
SUS	0.200	0.055	201	0.266	0.991	201
DGBL	0.057	0.062	201	0.063	0.997	201
ENG	0.042	0.064	201	0.017	0.983	201

Table 30 - Normality of the variables.

The internal consistency of the questionnaire is also supported by the results of the correlation matrix calculated on the mean scores of the DGBL, TAM, SUS and ENG items. Table 31 presents the correlation matrix, which was calculated based on the means of the four variables, along with the Spearman correlation coefficient (r) for each pair of variables and the corresponding p-values from the two-tailed significance test. All four variables show strong and statistically significant positive correlations ($p < .001$), suggesting a high level of confidence in the validity of these results. As in previous study, the strongest association is between DGBL and TAM ($r = 0.818$), suggesting a very tight relationship between these two constructs. As mentioned above, it is consistent with the fact that students who perceived the serious game as an effective way of enhancing Mathematics learning are also more likely to attribute greater usefulness to this approach. The correlations between the other pairs of variables, which range between 0.610 and 0.717, further confirm that usability and engagement are important aspects of the user experience in a game-based learning environment. In this experiment, higher correlations were observed compared to the previous study. For example, the correlation between SUS and TAM increased from 0.428 to 0.617, and the correlation between DGBL and SUS increased from 0.514 to 0.672. These higher values indicate stronger relationships among the variables and a more coherent pattern of responses.

		TAM	SUS	DGBL	ENG
TAM	r	1	0.617	0.818	0.610
	p-value		<0.001	<0.001	<0.001
SUS	r	0.617	1	0.672	0.717
	p-value	<0.001		<0.001	<0.001
DGBL	r	0.818	0.672	1	0,669

	p-value	<0.001	<0.001	<0.001	<0.001
ENG	r	0.610	0.717	0.669	1
	p-value	<0.001	<0.001	<0.001	

Table 31- Correlation matrix on the mean scores.

An exploratory factor analysis of the revised ENG scale was carried out on the answers to the final questionnaire after playing the Murderer game. This revealed a substantially more restrictive factorial structure than the first analysis. While the original 32-item scale produced a six-factor solution, the reduced form of the scale (15 item) yielded only three clearly interpretable factors. Note that questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores. These items are identified as "reverse" in Table 32. The first factor aggregates (with factor loadings ranging from 0.464 to 0.862) items related to positive engagement, such as enjoyment, interest, aesthetics and positive involvement in the game. The second factor captures (factor loadings ranging from 0.401 to 0.790) aspects associated with negative engagement, including pressure, boredom, distraction and disorientation experienced during the serious game. The third factor groups (factor loadings ranging from 0.527 to 0.653) items related to competence, challenge and effectiveness in carrying out the game and its objectives. While the shortened version of the ENG scale provides a more concise measurement tool, some of the finer details that emerged in the original, more extensive structure are inevitably lost. One notable aspect that is less distinctly represented in the shortened scale is "flow", a concept widely regarded as a key element of engagement in learning and gaming environments. Consequently, the three-factor solution that emerged offers a broader, though less detailed, view of students' engagement. Although positive engagement may still reflect certain aspects of flow, it no longer captures its full complexity.

Items	**ID**	**Factor 1** "Positive engagement" loadings	**Factor 2** "Negative engagement" loadings	**Factor 3** "Competence, challenge and effectiveness in carrying out the game and its objectives" loadings
I thought it was fun.	ENG_B	0.862		
I was interested in the game's story.	ENG_A	0.845		
I enjoyed it.	ENG_M	0.786		
It felt like a rich experience.	ENG_T	0.741		
I was deeply concentrated in the game.	ENG_R	0.718		
It was aesthetically pleasing.	ENG_G	0.532		
The game feels real.	ENG_AF	0.464		
I felt pressured (reverse).	ENG_O		0.790	
I felt disoriented (reverse).	ENG_Z		0.776	
I thought it was hard (reverse).	ENG_F		0.605	
I felt bored (reverse).	ENG_J		0.536	
I thought about other things (reverse).	ENG_D		0.401	
I felt competent.	ENG_E			0.653
I was fast at reaching the game's targets.	ENG_N			0.577
I felt challenged.	ENG_P			0.527

Table 32 - Rotated component matrix for ENG variable.

A confirmatory factor analysis (CFA) was performed to test whether the data from the new final questionnaire fit the structure hypothesised in the theoretical model derived from the exploratory factor analysis of the DGBL, TAM and SUS variables, as measured by the final questionnaire of the pilot study. The data considered comes from students' responses after playing both serious. A first CFA was conducted, but the initial model showed unsatisfactory fit indices: the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI) and the Root Mean Squared Error of Approximation (RMSEA). It was examined which items were not performing adequately within the CFA model. Items with very low factor loadings values were identified as candidates for removal, as these items do not contribute to measuring the construct and negatively affect the overall model fit. They all belonged to Factor 3 "The complexity of the game, possible difficulties and need for support":

- SUS_B: "*The content and topics of the game are complex*".
- SUS_D: "*I think I would need the support of a technician to use this game*".
- SUS_I: "*This game needs more introductory explanations*".
- DGBL_M: "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*".

The suggested removal of items was carried out one by one which led to an improvement in the model fit. However, a closer evaluation revealed that two of the flagged items: SUS_B and DGBL_M captured conceptually meaningful aspects of the tool and were therefore important to retain. For this reason, both items were reintroduced to reassess the model. However, their simultaneous inclusion resulted in unacceptable fit indices. A second attempt was made by reinserting only SUS_B. In this case, the model still achieved an acceptable level of fit. The removal of items SUS_D and SUS_I led to a revision of the name of Factor 3, changing it from "The complexity of the game, possible difficulties and need for support" to "The complexity of the game and possible difficulties". Additionally, introducing two theoretically justified correlated error terms, since the items measure related aspects - specifically between TAM_G ("*To what extent do you think that the use of DGBL... in Mathematics can develop your critical thinking skills?")* and TAM_H ("*To what extent do you think that the use of DGBL... in Mathematics can develop your problem-solving skills?"*), and between TAM_B ("*To what extent do you think that the use of DGBL... in Mathematics can improve your performance?"*) and TAM_D ("*To what extent do you think that the use of DGBL.. in Mathematics can increase your effectiveness?"*), further improved the overall model fit. These adjustments resulted in a model that achieved a satisfactory fit, with values consistent with conventional standards: the CFI reached 0.902, the TLI 0.891 (acceptable, close to 0.90), the RMSEA 0.079 and the SRMR 0.056, indicating a model that performs well and is theoretically coherent. The model is considered acceptable, even when the SUS_B item is included. DGBL_M appears to represent a distinct construct from Factor 3. While it does not meaningfully contribute to measuring this factor, it nonetheless captures a unique aspect of the questionnaire, that is related to the reduction of gender gap in scientific disciplines. This suggests that DGBL_M may be valuable in understanding a separate facet of participants' perceptions or experiences, even if it does not fit neatly into the established factor model. In fact, the factor loading for this question was already very low in the pilot study. Examining the factor loadings reveals that Factor 1, "The impact of digital game-based learning on the Mathematics learning experience", exhibits very high loadings ranging from 0.68 to 0.86, indicating a strong and stable factor. Factor 2, "Perceived effectiveness for learning Mathematics", also shows very high loadings, with most items exceeding 0.8, except in a few cases. Factor 3, "The complexity of the game, possible difficulties", demonstrates generally acceptable loadings, except for SUS_B. However, this item was retained due to its conceptual importance. Finally, Factor 4, "Game experience and serious game appeal", shows acceptable loadings overall, between 0.7 and 0.8. The weakest item, SUS_L ("The scenes in this game are very attractive"), has a loading of 0.401. However, similar to the rationale for retaining USA_B, it was considered important and therefore kept in the model.
Section 8.1.2 presents the final model of the questionnaire to evaluate a serious game developed for the mathematical field, considering all the analyses carried out up to this point.

6.4.2 Results from the Final Questionnaire and Comparative Analysis across Groups

To understand the patterns and trends within each of the four variables, descriptive statistics were calculated for the mean, median and standard deviation of each variable on a scale of 0-100 (Table 33). The answers come from the final questionnaire responses after playing the Murderer game.

		TAM	SUS	DGBL	ENG
N (sample)	Valid	201	201	201	201
	Missing	3	3	3	3
Mean		59.86	60.73	62.39	57.52
Median		60.00	62.00	61.54	57.33
Standard deviation		20.34	12.65	17.23	12.30

Table 33 - Descriptive statistics for TAM, SUS, DGBL and ENG variables.

As shown in Table 33, the mean values for the four variables range from 57.52 to 62.38. As in the first experiment, the lowest values were obtained for the ENG and TAM variables (57.52 and 59.86, respectively, on a 0-100 scale), although both are noticeably higher than before study, about +3 and +7 points, respectively. This improvement may perhaps be related to the enhancements made to the game components (e.g. improved voice-over feedback, better introductory explanations of the game and more realistic characters). Regarding the DGBL items, there was an increase of almost one point, reaching 62.39. The only variable whose value decreased compared to the previous edition is SUS (60.73 vs 63.38).

Compared to the earlier experiment, the median values increased (except for SUS, which remained unchanged) and the standard deviations decreased, meaning that participants' responses were more consistent and showed less variability. As in the previous study, the highest standard deviation was obtained for TAM (20.34), though this is lower than in the earlier edition (24.71). This was followed by DGBL (17.23), which was also lower than in the previous edition (19.94). This suggests that there is still considerable spread in opinions or experiences related to the acceptance of tools, such as serious games, and the use of DGBL in Mathematics. The standard deviation for the SUS and ENG variables is around 12.5, indicating that there is less variability among respondents in terms of system usability and engagement, and that there are fewer extreme differences in opinion. It is worth noting that, once again, the DGBL_M item "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*" obtained the highest mean and median within the DGBL scale: 69.35 (even if it is lower than in the previous edition, when it was 72.2) and 80, respectively. The Mann-Whitney U test revealed no statistically significant differences in answers between male and female students for this item (p-value = 0.933). These results suggest that male and female students show a comparable level of agreement in their responses to this item.

In the previous edition, the least appreciated aspects were those related to personalized learning. Questions such as "*To what extent do you think the use of digital, game-based learning in Mathematics can promote a personalized learning approach?*" and "*How well does it meet your needs as a student?*" received the lowest scores of the DGBL scale. In the second experimentation of the serious game, the mean score for the first question improved from 53.25 to 63.62, while the mean score for the second question remained nearly unchanged.

Looking more closely at the mean scores on a 0-100 scale, the items SUS_M, "*The contents were easy to read*" (63.68), and SUS_N, "T*he contents were easy to understand*" (59.3), obtained the highest values on the SUS scale. Regarding the questions related to the extent to which DGBL can develop critical thinking skills (TAM_G) and problem-solving skills (TAM_G), the mean scores were 65.12 and 69.64, respectively.

Table 34 shows the comments obtained from the final questionnaire, which was answered on a voluntary basis, and their respective frequencies. The higher frequency of comments (18) indicates that the game is complete and useful. Many students (11) emphasised the need for more time to read the texts. However, as stated in the instructions, the texts and dialogues can be replayed as many times as desired for as long as desired.

Comments, suggestions, and difficulties encountered	Frequencies
Complete and useful activity.	18
Allow more time to view the texts.	11
Include other mathematical topics as well.	5
Make the questions clearer.	3
Limited activity for learning.	3
Improve the graphics.	2
Make answer selection more intuitive.	2
Tailor the game to a specific course of study from a mathematical perspective.	2
No difficulties.	1
Allow viewing the correct answers.	1

Game too simple, make it more engaging.	1
Useful application of AI.	1
Allow viewing the texts during the quiz.	1
Difficulty understanding the explanation of the suspects' paths.	1
Make the game easier for those less skilled in Mathematics.	1
Difficulty understanding the evaluation method.	1

Table 34 – Qualitative comments and their frequencies.

Table 35 reports the responses provided in answer to the open-ended question about the potential of the serious game. Students were required to provide an answer. Most students (120) found the serious game useful for revising the topics covered in class and for applying mathematical concepts (51). It is worth noting the response of the six students, who identified the serious game as an effective assessment activity.

Potential of the serious game	Frequencies
It is useful and interesting for learning and revising Mathematics.	120
It helps you to see Mathematics being applied in a practical way.	51
It stimulates problem solving.	21
It is also a good assessment activity.	6
Encourages collaboration among groups.	5
It encourages the development of transversal skills.	3
Questions are phrased in a more understandable way.	1
It stimulates competitiveness among students.	1

Table 35 – Qualitative comments and their frequencies about potential of the serious game.

The Kruskal-Wallis test and one-way ANOVA results for the normally distributed variables DGBL and SUS allow to determine whether there are significant differences in the mean values of the variables based on students' degree programmes.

- DGBL: the Kruskal-Wallis test shows a statistically significant difference among groups (p-value = 0.006). A one-way ANOVA also confirmed the presence of a significant difference in mean DGBL scores across the degree programmes (p-value = 0.003), which is consistent with the nonparametric result. Post-hoc tests (Tukey HSD and Bonferroni) indicate that Biotechnology degree programme differs significantly from the others, particularly Medical Radiology, Imaging and Radiotherapy Techniques and Prevention Techniques in the Environment and Workplaces programmes, showing a higher and more positive perception of DGBL.
- TAM: the Kruskal-Wallis test indicates a statistically significant difference (p-value = 0.013).
- SUS: the Kruskal-Wallis test reveals a statistically significant difference among groups (p-value = 0.008). A one-way ANOVA also confirms a significant difference in mean usability scores across programmes (p-value = 0.010), consistent with the Kruskal-Wallis result. Post-hoc tests (Tukey HSD and Bonferroni) show that the only significant difference is between Biotechnology degree programme and Prevention Techniques in the Environment and Workplaces programme: students in the first programme rated usability significantly higher than those in the second.
- ENG: a statistically significant difference also emerges for the ENG variable (p-value = 0.031), indicating that mean ENG scores differ across degree programmes.

It was also calculated descriptive statistics for mean and standard deviation (SD) for each variable according to the programme degree (Table 36). Except for the degree programmes "Medical Radiology, Imaging and Radiotherapy Techniques" and "Prevention Techniques in the Environment and Workplaces", all other degree programmes reflect the overall trend highlighted by the statistical analyses of the full sample, namely that the mean scores for the ENG variable were the lowest. In contrast, the mean scores for the TAM variable were lower for these two degree programmes. Furthermore, it is noteworthy that these two programmes consistently scored lower than the others across all four variables, with differences of up to 10 points. This partially confirms the findings of the first experiment, in which the group studying for a bachelor of Medical Radiology, Imaging and Radiotherapy Techniques obtained the lowest mean scores for all four variables. This indicates lower favourable perception of the serious game than the other groups. Students in Biotechnology gave higher scores for all four variables compared to the previous study, followed by students on the bachelor's programme in Strategic and Security Sciences. Note that this programme of study was not part of the earlier experiment, so it is not possible to make a comparison. Except for the items in the SUS scale, Biotechnology students provided mean scores that were approximately 4 to 7 points higher than last edition. The mean scores provided by Strategic and Security Sciences students also suggest that, despite coming from a different field to Biotechnology students, they appreciated the serious game. It should also be noted that lower scores than in the previous experiment were observed among students on the master's programme in Strategic and Security Sciences. These students had been the group that had appreciated the serious game the most in the pilot study. However, this group comprised only six students, which is a very small sample size.

DEGREE PROGRAMME		TAM	SUS	DGBL	ENG
Medical Radiology, Imaging and Radiotherapy Techniques	Sample	21	21	21	21
	Mean	50.09	56.10	53.04	51.94
	SD	19.36	14.47	18.07	14.43
Prevention Techniques in the Environment and Workplaces	Sample	23	23	23	23
	Mean	51.75	54.71	53.75	53.74
	SD	18.28	9.13	17.19	12.03
Biotechnology	Sample	116	116	116	116
	Mean	63.05	63.09	65.27	59.06
	SD	21.53	12.17	16.76	12.29
Strategic and Security Sciences (Bachelor's program)	Sample	35	35	35	35
	Mean	60.91	60.51	64.62	58.59
	SD	16.34	12.27	15.87	10.54
Strategic and Security Sciences (Master's program)	Sample	6	6	6	6
	Mean	57.14	56.00	59.49	55.56
	SD	12.39	18.49	12.26	9.15

Table 36 - Descriptive statistics for the variables according to the degree programme.

Based on the results of the Mann-Whitney U test conducted on the four variables (DGBL, SUS, ENG and TAM) in relation to the gender variable, it was found that the resulting p-values for all four tests were less than the significance level of 0.05. The p-values were found to be less than 0.01 for DGBL, SUS and ENG, while the p-value for TAM was found to be 0.03. Therefore, the null hypothesis that the distribution of the variables is the same for male and female participants is rejected. As in previous experiment, female participants gave more positive responses on average for all four variables (Table 37), achieving scores higher than the overall mean (Table 33). Once again, a test was conducted on play frequency to see if there were any differences between the two groups. This was done to support the idea that male participants, being generally more experienced with high-quality commercial video games, may have had more points of comparison and therefore evaluated the educational game more critically. The independent-samples t-test showed again a significant difference in video game frequency between the two groups of students in this second sample, indicating that the male group plays significantly more than the female group.

Variable	Genre	N	Mean	Standard deviation
	Female	109	63.72	21.72
	Male	92	55.27	17.63
TAM	Total	201	59.86	20.34
SUS	Female	109	63.47	13.50
	Male	92	57.48	10.77
	Total	201	60.73	12.65
DGBL	Female	109	65.92	18.04
	Male	92	58.20	15.29
	Total	201	62.39	17.23
ENG	Female	109	60.13	12.15
	Male	92	54.42	11.79
	Total	201	57.52	12.30

Table 37 - Descriptive statistics according to genre of participants.

After conducting exploratory factor analysis for the ENG variable and confirmatory factor analysis for the DGBL, TAM and SUS variables, descriptive statistics were computed for the extracted factors (Table 38). There were four factors for the DGBL, TAM and SUS variables and three factors for the ENG variable (obtained for the reduced scale). A total of 201 responses to the final questionnaire, completed after at least one attempt at the Murderer game, were analysed. Regarding the DGBL, SUS and TAM-related factors, the highest mean scores were obtained for Factor 1, followed by Factor 3 and $Factor_{ENG}$ 2 (with similar values. Compared to the values obtained in the pilot study, Factor 1 increased by approximately one unit, while Factor 3, “The complexity of the game, possible difficulties”, decreased significantly. This may be because students perceived the game's content as more complex than in the previous edition. In contrast, Factor 2, “Perceived effectiveness for learning Mathematics”, increased by almost seven points. As in the pilot study, Factor 2 showed a considerable degree of dispersion, suggesting that some students found the serious game effective for learning, while others did not. Factor 4, “Game experience and serious game appeal”, increased by approximately two points compared to the previous experiment, although it still had the lowest overall score of the four factors. Engagement is positively associated with Factor 4 scores. Correlation analysis revealed a strong positive Pearson correlation between engagement and Factor 4, labelled “Game experience and serious game appeal” ($r = 0.71$, $p < .001$). To further examine this relationship, a linear regression analysis was conducted to assess the predictive role of engagement on Factor 4. The resulting regression model is statistically significant ($p <0\ .001$), indicating that engagement scores significantly predict Factor 4. The standardized regression coefficient shows a strong effect ($\beta = 0.70$), the standard error of the estimate is 12.81. The model explains a substantial proportion of variance in the dependent variable, accounting for 50.4% of the total variance. No direct comparison can be made between the engagement factors and the questionnaire results from the previous experiment because the number of items and factors was reduced. Nevertheless, the average scores obtained for the three current Engagement factors are higher (three out of four factors were less than or slightly more than 50.00). One possible explanation for this is that the shorter questionnaire prevented students from becoming impatient or fatigued, resulting in more positive and focused evaluations. The highest score was obtained for $Factor_{ENG}2$, “Negative engagement”. This reflects the fact that students reported the least pressure, boredom and disorientation during the game among the three factors. A slightly lower mean score was recorded for $Factor_{ENG}1$, “Positive engagement” (57.41), which refers to aspects such as enjoyment, interest, aesthetics, and positive involvement in the game. The lowest mean score (53.26) was obtained for $Factor_{ENG}3$, “Competence, challenge, and effectiveness in carrying out the game and its objectives”. This indicates that, among all engagement-related aspects, the dimension tied to learning objectives received the lowest average score. Note that the items that were worded in reverse - where a positive response corresponded to a low score - were rephrased so that positive responses aligned with higher scores on the Likert scale. No correlations were found between students' upper secondary school grades and the variables and factors of the questionnaire.

Items	**Factor 1** “The impact of digital game-based learning on the Mathematics learning experience"	**Factor 2** “Perceived effectiveness for learning Mathematics”	**Factor 3** “The complexity of the game, possible difficulties”	**Factor 4** “Game experience and serious game appeal”	**$Factor_{ENG}$ 1** “Positive engagement”	**$Factor_{ENG}$ 2** “Negative engagement”	**$Factor_{ENG}$ 3** “Competence, challenge and effectiveness in carrying out the game and its objectives”
Mean	61.71	59.74	61.39	51.32	57.41	61.05	53.26
Median	60.00	60.00	60.00	55.00	60.00	64.00	55.00
Standard deviation	19.09	20.042	15.49	18.20	18.67	13.21	14.38

Table 38 - Descriptive statistics for the extracted factors.

A Mann-Whitney U test was conducted to examine potential gender differences in the factor scores, using gender as the grouping variable. The analysis included 201 valid responses. The results revealed statistically significant differences between genders for all examined factors, with the sole exception of $Factor_{ENG}3$ (p-value = 0.075), for which no significant difference was detected. Overall, the analysis revealed differentiated profiles of perceptions and reactions to the serious game, indicating a degree of gender sensitivity. Indeed, when mean scores were examined by gender, female students consistently reported higher values, often by as much as 10 points, compared to male students. $Factor_{ENG}3$ was the only factor that does not follow this pattern. Examining the various factors relating to participants' gaming habits, as measured by gameplay frequency, reveals no significant differences in their responses.

6.4.3 Changes in Attitudes towards Mathematics from the Initial to the Final Questionnaire

As in previous experiments, changes in students' attitudes towards Mathematics were assessed by comparing questionnaire data collected before and after playing the Murderer game. A reliability analysis was conducted on the responses of 204 students to all 42 items related to attitudes towards Mathematics. Very high Cronbach's alpha values were obtained for both QI (0.886) and QF (0.888). These values indicate excellent internal consistency for both questionnaires. The Cronbach's alpha value obtained for the QI is slightly higher than that of the previous experiment, while the value for the QF remains the same. Based on the results of the pilot study, the original eight dimensions were reduced to seven by merging the first two. The resulting dimensions are: Perception of Mathematics, Interest and motivation, Math anxiety, Persistence in Mathematics, Persistence in Problem Solving, Development of critical thinking, Perception of the course and Extrinsic motivation. To further examine the internal consistency of each subscale, Cronbach's Alpha was calculated for the seven theoretical dimensions identified a priori (Table 39). Unlike in the previous experimentation, no particularly low values were obtained. Instead, the scales showed Cronbach's alpha values of around 0.650 or higher. The lowest Cronbach's alpha values were observed in the QI for Persistence in Mathematics and in the QI and QF for the Development of Critical Thinking dimension, which had already exhibited moderate values in previous experiment. Based on the results of the pilot study, the item "I keep studying until I understand Mathematics material" was moved from the Perception of Mathematics, interest, and motivation dimension to Persistence in Mathematics as this grouping seemed to be more coherent. The Cronbach's alpha calculated in the QI for this revised group of items is 0.657, which is comparable to the value obtained in the previous study (0.662), while the Cronbach's alpha value in the QF increased from 0.634 in the previous experiment to 0.754 in this second experiment. Conversely, including the item in the Perception of Mathematics, interest, and motivation dimension would decrease the internal consistency (from 0.813 to 0.801 in the QI), indicating a poorer fit.

Item Group (Dimension)	Number of items	Cronbach's Alpha (QI)	Internal Consistency (QI)	Cronbach's Alpha (QF)	Internal Consistency (QF)
Perception of Mathematics, interest, and motivation	6	0.813	Very good	0.832	Very good
Math anxiety	4	0.841	Very good	0.785	Good
Persistence in Mathematics	7	0.657	Discrete	0.754	Good
Persistence in problem solving	4	0.706	Good	0.691	Discrete
Development of critical thinking	8	0.642	Discrete	0.692	Discrete
Perception of the course	9	0.845	Very good	0.873	Very good
Extrinsic motivation	4	0.722	Good	0.763	Good

Table 39 - Cronbach's alpha for QI and QF regarding attitudes towards Mathematics.

As in the previous experiment, the Wilcoxon signed-rank test for paired samples was used to analyse the differences between the responses to the QI and QF questionnaires collected from the same students (before and after playing the game). The QI and QF responses of 204 students were compared. This included 117 students from the Biotechnology programme, and 46 students from the Prevention Techniques in the Environment and Workplaces programme, as well as the Medical Radiology, Imaging and Radiotherapy Techniques programme. Furthermore, the responses of students enrolled in the bachelor's and master's degree programmes in Strategic and Security Sciences were analysed collectively. The results of the Wilcoxon signed-rank test on the mean item scores across the seven dimensions are presented by degree programme. For the Biotechnology programme, the results confirmed those obtained in the first experiment. An improvement was observed in the reduction of math anxiety (p-value= 0.000000007) and extrinsic motivation (p-value= 0.00452) in mean values, although it was less marked this time (-0.61 and -0.20, respectively, versus -0.96 and -0.30 in the first experiment). The other dimensions remained almost unchanged.
An item-by-item comparison (Table 40) shows that the greatest decrease was found for the fourth question, "When I take a math test, I think of the consequences of failing / While I was playing, I thought about the consequences of failing" (a decrease of more than one unit), similarly to the first experimentation.

	Math anxiety For each statement, indicate a response from 1 to 6, where 1 = "Not at all true for me" and 6 = "Extremely true for me."							
Item	**QI-MA_A:** I get very tense when I have to do Mathematics homework	**QF-MA_A:** I was very tense while playing the serious game	**QI-MA_B:** I feel helpless when doing a Mathematics problem	**QF-MA_B:** I feel helpless when doing a Mathematics problem	**QI – MA_C:** When I take a test I think about how poorly I am doing compared with other students	**QF-MA_C:** While I was playing, I thought about how badly I was doing compared to the other students	**QI-MA_D:** When I take a math test I think of the consequences of failing	**QF-MA_D:** While I was playing, I thought about the consequences of failing.
Mean	3.00	2.38	2.61	2.49	2.73	2.09	3.32	2.25
Standard deviation	1.49	1.44	1.47	1.26	1.44	1.34	1.60	1.37
Increase/ Decrease in the mean	-0.62		-0.12		-0.63		-1.07	
p-value	0.00014		0.23234 (> 0.1)		0.00000118		0.00000000129	

Table 40- Wilcoxon signed-rank test for paired samples: math anxiety (Biotechnology degree programme).

As with the Biotechnology degree programme, a reduction in math anxiety was also observed among students enrolled in the Prevention Techniques in the Environment and Workplaces and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes (mean change: -0.50; p-value = 0.0028). An item-by-item comparison (Table 41) shows that the magnitude of the reduction is similar across the three items that reached statistical significance. As in the previous experiment, the item "I feel helpless when doing a Mathematics problem" did not reach statistical significance. Regarding the Persistence in Mathematics dimension, the results of the first experimentation showed a decline in persistence of -0.48 for the Prevention Techniques in the Environment and Workplaces and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes. In this second experiment, a moderately significant result was found (p-value = 0.0624) with a slight worsening in this dimension (-0.10). However, this value does not permit the conclusion that there was a real decrease in the perception of persistence in Mathematics for this sample of students from pre- to post-intervention. For the other dimensions, no significant pre-post changes were detected in the sample, except for the dimension Persistence in problem solving, for which a decline of 0.30 was observed.

	Math anxiety For each statement, indicate a response from 1 to 6, where 1 = "Not at all true for me" and 6 = "Extremely true for me."							
Item	**QI-MA_A:** I get very tense when I have to do Mathematics homework	**QF-MA_A:** I was very tense while playing the serious game	**QI-MA_B:** I feel helpless when doing a Mathematics problem	**QF-MA_B:** I feel helpless when doing a Mathematics problem	**QI – MA_C:** When I take a test I think about how poorly I am doing compared with other students	**QF-MA_C:** While I was playing, I thought about how badly I was doing compared to the other students	**QI-MA_D:** When I take a math test I think of the consequences of failing	**QF-MA_D:** While I was playing, I thought about the consequences of failing.
Mean	3.30	2.59	2.80	2.77	2.86	2.18	3.22	2.66
Standard deviation	1.68	1.28	1.76	1.28	1.61	1.18	1.57	1.39
Increase/ Decrease in the mean	-0.71		-0.03		-0.68		-0.56	
p-value	0.014		0.46 (>0.1)		0.0028		0.0127	

Table 41- Wilcoxon signed-rank test for paired samples: math anxiety (Prevention Techniques in the Environment and Workplaces and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes).

A Wilcoxon test (Table 42) conducted on undergraduate and postgraduate students on the Strategic and Security Sciences programme showed a clear reduction in math anxiety for this sample of students on mean values (-0.56). For this sample of students, the results were significant for two questions and non-significant for two others. A substantial decrease was observed for the question, "When I take a math test I think of the consequences of failing/ While I was playing, I thought about the consequences of failing" (-0.74), but an even stronger decrease was seen for the third question "When I take a test, I think about how poorly I am doing compared to other students/ While I was playing, I thought about how badly I was doing compared to the other students"(-0.98). There was also a slight improvement of around 0.2 in the Perception of Mathematics, interest and motivation, and Extrinsic motivation dimensions, although the latter showed only a moderately significant p-value of around 0.10.

Item	QI-MA_A: I get very tense when I have to do Mathematics homework	QF-MA_A: I was very tense while playing the serious game	QI-MA_B: I feel helpless when doing a Mathematics problem	QF-MA_B: I feel helpless when doing a Mathematics problem	QI – MA_C: When I take a test I think about how poorly I am doing compared with other students	QF-MA_C: While I was playing, I thought about how badly I was doing compared to the other students	QI-MA_D: When I take a math test I think of the consequences of failing	QF-MA_D: While I was playing, I thought about the consequences of failing.
Mean	3.18	2.74	2.72	2.62	2.90	1.92	3.18	2.44
Standard deviation	1.63	1.39	1.75	1.27	1.64	1.11	1.56	1.47
Increase/ Decrease	-0.44		-0.10		-0.98		-0.74	
p-value	0.20 (> 0.1)		0.43 (> 0.1)		0.0060		0.00468	

Table 42- Wilcoxon signed-rank test for paired samples: math anxiety (Strategic and Security Sciences programmes).

6.4.4 Exam Results: a Comparison of the Intervention and Control Groups

The exam results for the 2024-25 academic year of the Biotechnology and Medical Radiology, Imaging and Radiotherapy Techniques, which were administered through an automated assessment system, were compared with those from the preceding academic years: 2021-22 or 2022-23, when the serious game was not used, and 2023-24, when it was implemented.
For the Biotechnology degree programme, the second experimental group comprised 86 students who played the serious game at least once and took the exam in the 2024-25 academic year (the first group, from the 2023-24 academic year, comprised 95 students). The questions, which were scored on a scale from 0 to 4, were similar to those analysed in previous years and focused on functions in a real variable - the same topic covered in the Murderer game. In both the 2023-24 and 2024-25 academic years, the average results obtained for the question related to the function study, were clearly higher than in the academic year when the serious game was not used (Table 43). However, the scores for 2023-24 were better than those for 2024-25. The same applies to the median scores, with a substantial decrease in the standard deviation. The percentage of students achieving a median score of 2 or higher increased from 64% to 93% in the 2023-24 academic year and was 87% in the 2024-25 academic year.

	2021-22 academic year (control group)	2023-24 academic year (experimental group)	2024-25 academic year (experimental group)
Mean score	2.25	2.96	2.72
Standard Deviation	1.40	0.89	0.46
Median score	2.25	2.97	2.83
Percentage of students obtained a score above the median score (≥2)	64%	93%	87%

Table 43- Comparison of the exam results of students on the Biotechnology degree programme between the control and experimental groups for the question related to the function study.

Average results improved significantly in 2024-25 compared with the previous two academic years, particularly for questions related to the properties of functions based on their graphs (Table 44). The mean score increased by 0.7 compared to 2021-22, and by 0.6 compared to 2023-24. The percentage of students achieving at least a median score of 2 or higher increased markedly, rising from 55% in 2021-22 to 61% in 2023-24 and 91% in 2024-25.

	2021-22 academic year (control group)	2023-24 academic year (experimental group)	2024-25 academic year (experimental group)
Mean score	2.14	2.23	2.89
Standard Deviation	0.99	1.08	0.81
Median score	2.29	2.29	3.20
Percentage of students obtained a score above the median score (≥2)	55%	61%	91%

Table 44- Comparison of the exam results of students on the Biotechnology degree programme between the control and experimental groups for the question related to the properties of functions based on their graphs.

These results are also confirmed by the cumulative relative frequency plots, which shows for each score the proportion of students who achieved *at least* that score. As can be seen in Fig. 17, the scores obtained for the question related to the function study are significantly higher in the academic years when the serious game was used (red and green line), with the best performance occurring in 2023-24. Similarly, higher scores are observed for the question concerning the properties of functions based on their graphs in the academic years when the serious game was played. However, in this case, the best performance comes from the 2024-25 academic year (see Fig. 18).

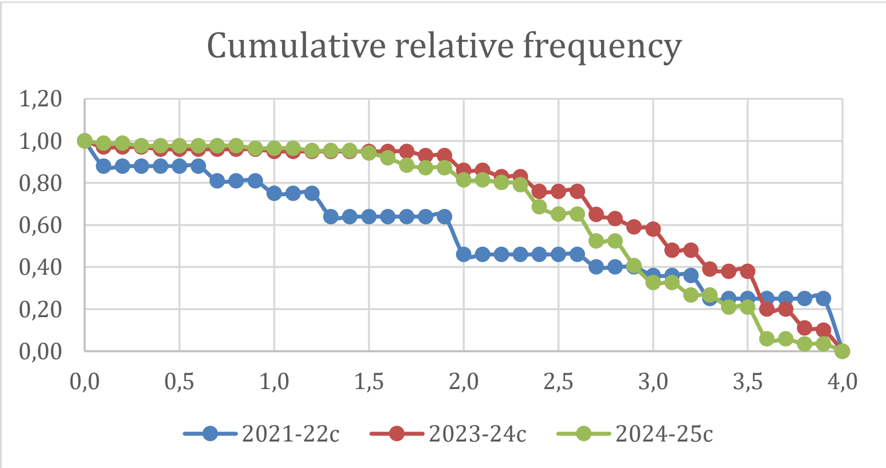

Figure 17 - Cumulative distribution plot of exam scores for the for the question related to the function study.

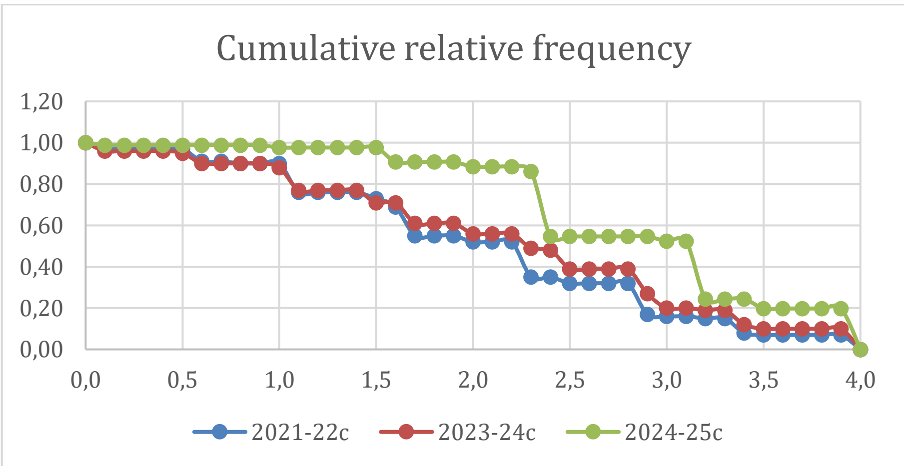

Figure 18 - Cumulative distribution plot of exam scores for the question concerning the properties of functions based on their graphs.

Table 45 compares the results of students in the Biotechnology degree programme for the 2024-25 academic year who completed the serious game before the exam (86 students) with those who did not (15 students). The comparison is based on the average score obtained in the two questions related to the topic of the serious game. The former group achieved an average score that was higher, with a difference of approximately 0.7 points compared to the latter.

	Students played the serious game (86)	Students didn't play the serious game (15)
Mean score	2.81	2.13
Standard Deviation	0.63	1.25
Median score	3.02	2.12

Table 45- experimental and control group in the 2024-25 academic year.

For the Medical Radiology, Imaging and Radiotherapy Techniques degree programme, the second experimental group comprised 21 students who played the serious game at least once and took the exam in the 2024-25 academic year (the first group, from the 2023-24 academic year, comprised 18 students). The control group consists of 18 students who took the exam in the 2022-23 academic year. The questions, which were scored on a scale from 0 to 4, were similar to those analysed in previous years and focused on functions in a real variable - the same topic covered in the Murderer game. As can be seen in Tables 46 and Table 47, which report the average scores for questions related to the serious game topic, both for the elementary functions and the function study item, there is a noticeably better performance in the 2024-25 academic year compared to previous years. It is also evident that a higher percentage of students achieved at least the median score in the 2024-25 academic year.

	2022-23 academic year (control group)	2023-24 academic year (experimental group)	2024-25 academic year (experimental group)
Mean score	1.48	2.28	4.20
Median score	0.90	2.25	4.50
Standard Deviation	1.40	1.18	0.53
Percentage of students obtained a score above the median score	9%	33%	86%

Table 46- Comparison of the exam results of students on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme between the control and experimental groups for the question related to the elementary functions.

	2022-23 academic year (control group)	2023-24 academic year (experimental group)	2024-25 academic year (experimental group)
Mean score	2.83	2.64	4.10
Median score	3.00	2.40	4.80
Standard Deviation	2.34	1.83	1.44
Percentage of students obtained a score above the median score	50%	33%	62%

Table 47- Comparison of the exam results of students on the Medical Radiology, Imaging and Radiotherapy Techniques degree programme between the control and experimental groups for the question related to function study.

Figure 19 shows the cumulative relative frequencies of exam scores for the three academic years considered. It can be observed that the curve for 2024-25 (in green) consistently remains higher compared to those of the previous years. This indicates that a higher percentage of students achieved higher scores. The 2023-24 curve (in red) lies in an intermediate position: although it shows an improvement compared to 2022-23, it remains below the 2024-25 curve. The 2022-23 curve (in blue), which refers to the year in which the serious game was not used, is consistently the lowest of the three curves, indicating poorer performance. Figure 20 also clearly shows that the 2024-25 curve (in green) is consistently higher than the others across almost the entire score range. This indicates that a larger percentage of students achieved higher scores compared to previous years. The curve also shows steeper rises towards the upper end, suggesting an overall better distribution of results. The 2023-24 curve (in red) lies in an intermediate position. Although it represents an improvement over 2022-23, it remains visibly below the 2024-25 performance, especially in the mid-to-high score range. The 2022-23 curve (in blue) is once again the lowest, indicating less positive performance. The steeper drops as the score increases show that a substantial share of students achieved low scores that year.

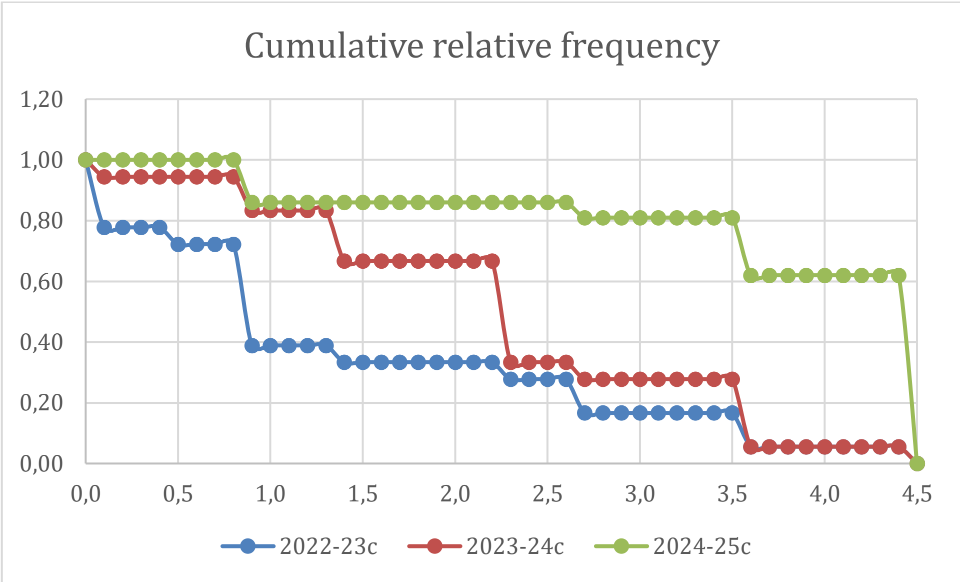

Figure 19 - Cumulative distribution plot of exam scores for the for the question related to the elementary functions.

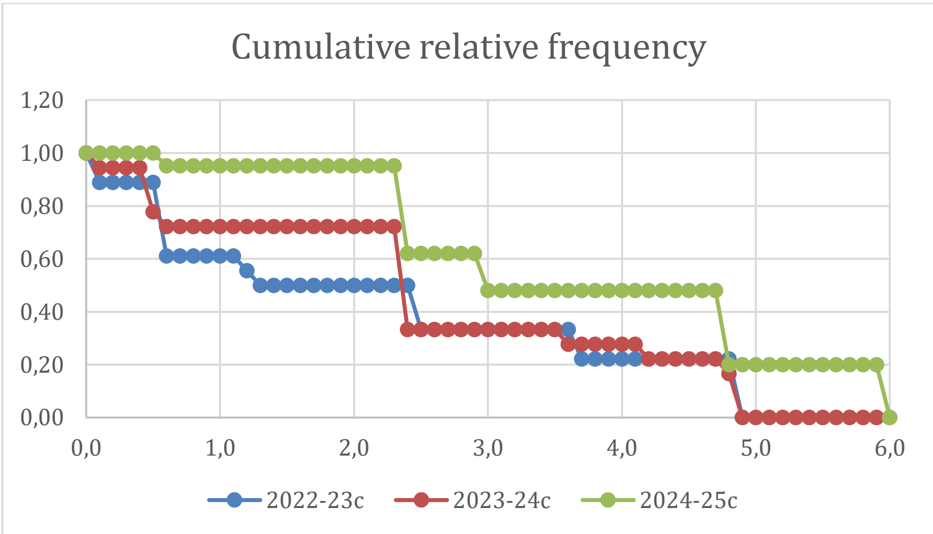

Figure 20 - Cumulative distribution plot of exam scores for the for the question related to the function study.

Unlike with the Biotechnology degree programme students, no comparison was made between those who took the exam in the same academic year and played the serious game, and those who did not. This was because the number of students who did not play was too small, only five.

A non-parametric Mann-Whitney U test was conducted on the high-school mathematics grades reported in the initial questionnaire to assess whether there were any differences in the incoming level of mathematical background between students enrolled in the 2023-24 and 2024-25 academic years. The analysis revealed no statistically significant differences between the 2023-24 and 2024-25 cohorts of students both on the Biotechnology and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes (p-value =0.351 and 0.527 respectively). The results suggest that the two cohorts have a comparable level of mathematical background on the basis of their high school mathematics grades, indicating that any differences in examination performance may not be attributed to preparation in Mathematics at high school level. It was not possible to perform the same analysis for the 2021-22 academic year, as this information was not requested for that cohort.

Spearman correlation analysis of students enrolled on the Biotechnology degree programme revealed significant, slightly negative associations between the average score of the two exam questions considered and responses to the final questionnaire. Specifically, the average exam question score was negatively correlated with TAM (r = −0.35, p-value = 0.001), DGBL (r = −0.29, p-value = 0.007), Factor 1 “The impact of digital game-based learning on the Mathematics learning experience” (r = −0.31, p-value = 0.003) and Factor 2 “Perceived effectiveness for learning Mathematics” (r = −0.33, p-value = 0.002). No significant correlations were found for the remaining factors. Spearman's correlation analysis revealed a significant, moderate negative association between the average score of the two exam questions and DGBL (r = –0.45, p-value = 0.039) for students enrolled in the Medical Radiology, Imaging and Radiotherapy Techniques degree programme. No significant correlations were found between exam performance and the other factors in the questionnaire.

6.4.5 Results from Game Sessions of the Murderer Game

During the 2024-25 academic year, 204 university students made a total of 419 gameplay attempts and completed both the initial and final questionnaires. On average, each student made 2.05 attempts, with a standard deviation of 1.31. It should be noted that the mean number of attempts per student were lower than in previous experiment. Considering that the scale ranges from 0% to 100%, the average score obtained is 72.9%. The winning range is 75% to 100%.

Based on data from game sessions involving 86 students on the Biotechnology degree programme who took the exam, there is a weak positive correlation (r = 0.198, using Spearman's correlation) between the average scores obtained in various attempts at the serious game and the average scores achieved in two exam questions related to functions (the mean of these two questions). However, this relationship is not statistically significant at the standard 0.05 level of significance (p-value = 0.069). The correlation between exam results and the number of attempts made on the serious game was also studied. However, no significant correlation was found, so it cannot be concluded that those who made more attempts performed better in the exam. A second analysis was conducted to determine whether average serious game scores or the number of attempts were more strongly related to either of the two exam questions. However, no significant correlations were found for either question.

For the sample of students of the Medical Radiology, Imaging and Radiotherapy Techniques degree programme (N=21), a moderate positive correlation (r=0.426, using Spearman's correlation) was found between the average scores obtained in various attempts at the serious game and the average scores achieved in two exam questions related to functions. This relationship marginally missed statistical significance (p=0.054).

Correlation analyses were conducted also between the number of attempts completed in the serious game and the 201 responses to the final questionnaire. Weak relationships were found, as the coefficient of correlation was under 0.2. Specifically, the number of attempts in the serious game did not significantly correlate with mean scores on the DGBL, TAM, SUS and ENG scales. Similarly, no significant correlations were found between the number of attempts and any of the factors identified in the final questionnaire. Overall, these results suggest that participants' performance in terms of the number of attempts made in the serious game was not associated with their reported perceptions of DGBL, usability evaluations, engagement levels or technology acceptance ratings.

Regarding the correlation between the average scores obtained in the various attempts at the serious game and the students' responses to the final questionnaire, only $Factor_{ENG}$ 2 (“Negative engagement”) and $Factor_{ENG}$ 3 (“Competence, challenge, and effectiveness in carrying out the game and its objectives”) show a moderate, statistically significant positive Spearman correlation with the average scores obtained in the game attempts (r= 0.221, p-value = 0.02 and r = 0.222, p-value = 0.01, respectively). This suggests that students who achieved higher average scores tended to give slightly higher responses to two out of three engagement-related factors. All other factors were found to have no correlation, or a weak correlation that was not significant at the standard 0.05 level. No significant or weak correlations were found with the four variables of the final questionnaire (DGBL, SUS, TAM, ENG).

6.4.6 The Use of GenAI in the Murderer Game

In the final questionnaire completed after the serious game, 77 out of 204 students (38%) stated that they had used GenAI tools as a support during the activity. Of the 77 students, 48 were from the Biotechnology degree programme (total sample size: 117), 11 were from the Medical Radiology, Imaging and Radiotherapy Techniques programme (total sample size: 22), 8 were from the Prevention Techniques in the Environment and Workplaces programme (total sample size: 24), 8 were from the Strategic and Security Sciences bachelor's programme (total sample size: 35) and 2 were from the Strategic and Security Sciences master's programme (total sample size: 6). Of these, 70 used AI chatbots and assistants such as ChatGPT, Copilot or Gemini. The remaining seven reported using tools that were not actually AI-based. An analysis of the qualitative responses of students who used GenAI tools revealed that the majority asked the chatbot about the characteristics of specific functions and requested visualisations of their graphs. This aligns with the objective of introducing GenAI tools, whereby students are expected to apply hints suggested by these tools in the serious game. Calculating the average scores obtained in the various attempts at the serious game shows that students who used GenAI achieved mean scores approximately 3 points higher than those who did not (Table 48).

Serious game attempts	Students used Gen-AI (77)	Students didn't use Gen-AI (127)
Mean score	71.14	68.48
Standard Deviation	14.24	12.22
Median score	70.64	67.20

Table 48- Mean scores achieved by students in various attempts of the serious game, comparing those who used GenAI with those who did not.

An independent-samples t-test was conducted to determine whether there was a significant difference in mean scores of the various attempts of the serious game between students who used GenAI and who did not. The results indicated that the difference was not statistically significant (p-value = 0.61), suggesting that the use of GenAI does not have a measurable impact on mean scores in this sample. A Mann–Whitney U test was conducted to examine whether using GenAI influenced the number of attempts at the serious game. However, the results showed that the number of attempts was distributed in the same way among students who used GenAI as among those who did not. Therefore, it cannot be claimed that the use of GenAI discourages students from attempting the game again to improve their performance.

6.4.7 Findings from the Experimentation with PhD Students

The Murderer game was presented to seven PhD students enrolled on the Doctoral Programme in Digital Humanities at the University of Genoa, in collaboration with the University of Turin. The students attended the "Gamification and Serious Games" course during the 2024-25 academic year. Post-game interviews were conducted to collect feedback on aspects such as engagement, usability, and overall appeal, rather than students' performance, since the participants did not have a strong background in Mathematics. The interviews revealed that the game was well received overall: the storytelling was described as highly motivating, with an engaging and compelling plot. They appreciated the gradual increase in difficulty, perceiving it as an effective didactical strategy. Most students considered the game highly engaging, and the availability of subtitles was valued from an inclusivity perspective as it allowed players both to read and to listen. Participants also appreciated the fact that the mathematical content was embedded within the gameplay, enabling them to learn Mathematics implicitly through the mission rather than through explicit instruction. One student noted that the game requires the use of pen and paper, a fact that is explicitly stated at the beginning. While the student found the overall game concept effective, he suggested that, as the game progresses and becomes more difficult, the educational component becomes more explicit, with Kathy's dialogue becoming increasingly mathematical.

6.5 Results from the First Experimentation of the Alien Game in Higher Education

The study, conducted from October 2024 to November 2025 involved 88 university students. The sample comprises students who completed the initial questionnaire before playing the serious game, then played the game at least once and filled out the final questionnaire afterwards. This sample includes 88 first-year undergraduate students enrolled on the Biotechnology degree programme. This is the only programme that covers the topics addressed by the serious game. The sample consist of 46 females and 42 males. The descriptive statistics for each of the four scales (TAM, SUS, DGBL and ENG) are reported in Table 49. These include the mean, standard deviation and median for each item.

	TAM	SUS	DGBL	ENG
Mean	53.47	58.25	57.75	52.77
Median	54.29	58.00	56.92	50.67
SD	22.04	13.22	18.55	12.40

Table 49 - Descriptive statistics for the variables TAM, SUS, DGBL, ENG for the Alien game.

Considering that the possible score range spans from 0 to 100, the descriptive statistics for the four scales reveal a generally moderate level of evaluation across all dimensions. The mean values are between 52.77 and 58.25, indicating that students' perceptions are slightly above the midpoint of the scale, but not approaching the upper range associated with highly positive evaluations. Of the four measures, SUS and DGBL have the highest mean scores (58.25 and 57.75). In contrast, ENG, with a mean of 52.77. TAM variable occupies an intermediate position, indicating moderate levels of technological acceptance. The medians closely mirror the means for all variables, suggesting that the distributions are relatively symmetric and not driven by extreme outliers. In terms of standard deviation, TAM has the highest standard deviation (22.04), indicating substantial differences in how students perceived technological acceptance. By contrast, SUS and ENG demonstrate lower variability (13.22 and 12.40, respectively), suggesting more consistent evaluations.
It is worth noting that, once again, also for the Alien game the DGBL_M item "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*" obtained the highest mean and median within DGBL scale: 72.72 and 80, respectively.
Regarding questions about the extent to which DGBL can develop critical thinking (TAM_G) and problem-solving (TAM_H) skills, mean scores of 74.15 and 84.01 were obtained respectively, which are the highest values on the TAM scale. These values are also higher than those obtained when only Biotechnology students were considered in the second experimentation of the Murderer game (74.15 vs. 69.48 and 84.01 vs. 73.27). Additionally, the mean values for the items "*The contents were easy to read*" (SUS_M) and "*The contents were easy to understand*" (SUS_N) were higher than those obtained in the second experiment with the Murderer game (78.69 vs. 63.68 and 71.59 vs. 59.3).
Table 50 shows the comments obtained from the final questionnaire, which was answered on a voluntary basis, and their respective frequencies. The higher frequency of comments (4) indicates that the game is interesting and valid. Three students also reported difficulty understanding the content of the game.

Comments, suggestions, and difficulties encountered	Frequencies
An interesting and valid activity.	4
The content of the questions is unclear.	3
Develop this activity for every Mathematics topic.	2
Improve the animations.	2
Make the dialogues available for longer.	1
Show the solutions to the questions.	1
This is not useful for learning statistical concepts.	1
Allow the game audio to be lowered without affecting the computer's audio.	1
Allow retrying of incorrect answers.	1
Include additional exercises beyond multiple-choice questions.	1
It is necessary to have already studied the topics covered in order to complete the activity.	1

Table 50 – Qualitative comments and their frequencies.

Table 51 reports the responses provided in answer to the open-ended question about the potential of the serious game. Students were required to provide an answer. As for the Murderer game, most students (40) found the serious game useful for revising the topics covered in class and for applying mathematical concepts (31). The serious game was identified once again as an opportunity to assess the level of understanding and put into practice the content learned in class.

Potential of the serious game	Frequencies
An interesting and useful activity for learning and revision.	40
It makes the application of Mathematics practical.	31

It stimulates problem-solving.	13
It is useful for checking the level of understanding of the topics.	5
The activity is very intuitive.	1
Encourages group work.	1

Table 51 – Qualitative comments and their frequencies about potential of the serious game.

Table 52 shows descriptive statistics computed for the extracted factors from the final questionnaire.

Items	Factor 1 "The impact of digital game-based learning on the Mathematics learning experience"	Factor 2 "Perceived effectiveness for learning Mathematics"	Factor 3 "The complexity of the game, possible difficulties"	Factor 4 "Game experience and serious game appeal"	$Factor_{ENG}$ 1 "Positive engagement"	$Factor_{ENG}$ 2 "Negative engagement"	$Factor_{ENG}$ 3 "Competence, challenge and effectiveness in carrying out the game and its objectives"
Mean	56.12	53.71	59.38	46.42	50.34	58.73	48.98
Median	55.38	53.33	53.33	45.00	46.67	60.00	50.00
Standard deviation	20.25	22.44	22.44	19.16	19.52	11.88	14.96

Table 52 - Descriptive statistics for the extracted factors.

Factors 1,2,3 have mean values ranging from 53.71 to 59.38. Factor 3, concerning the perceived complexity of the game and potential difficulties, has the highest mean value (59.38). Factors 1 and 2 address the impact of game-based learning on the Mathematics learning experience and its perceived effectiveness. These factors remain moderately positive, and their medians align closely with the means, indicating relatively symmetrical distributions. Factor 4, relating to the game's overall appeal and quality of experience, shows the lowest mean value (46.42) among the mathematics-related dimensions. This suggests that, while students considered the game to be moderately effective for learning purposes, they perceived its general appeal to be less positive. The corresponding standard deviation (19.16) reveals substantial variability in individual responses. Considering that the reverse-worded items were recoded so that higher scores consistently correspond to more positive perceptions, $Factor_{ENG}2$ ("Negative engagement") indicates that, with a mean of 58.73, students generally did not experience significant pressure, boredom or disorientation during the game. This positive interpretation is reinforced by the factor's low standard deviation (11.88), which shows that this perception was relatively uniform across respondents. The remaining two engagement-related factors show more moderate outcomes. Positive engagement ($Factor_{ENG}1$) has a mean of 50.34, suggesting that students experienced only a modest level of emotional involvement or enthusiasm while interacting with the game. $Factor_{ENG}3$, which relates to competence, challenge and effectiveness in carrying out the goals of the game, has an average of 48.98, indicating that students felt only moderately confident and effective in navigating the game and achieving their objectives. Compared to the other factors, $Factor_{ENG}1$ and $Factor_{ENG}3$ show greater variability, suggesting that students' perceptions differed notably from person to person. Engagement is positively associated with Factor 4 scores. Correlation analysis revealed a strong positive Pearson correlation between engagement and Factor 4, labelled "Game experience and serious game appeal" ($r = 0.81$, $p < .001$). To further examine this relationship, a linear regression analysis was conducted to assess the predictive role of engagement on Factor 4. The resulting regression model is statistically significant ($p < .001$), indicating that engagement scores significantly predict Factor 4. The standardized regression coefficient shows a strong effect ($\beta = 0.81$), the standard error of the estimate is 11.33. The model explains a substantial proportion of variance in the dependent variable, accounting for 65.8% of the total variance.

6.5.1 Results from the Final Questionnaire and Comparative Analysis across Serious Games

Before examining differences in students' perceptions across the two serious games, the four variables (DGBL, SUS, TAM and ENG) were evaluated to determine whether they met the assumption of normality. A total of 172 responses to the final questionnaires (one for the Murderer game and one for the Alien game) from Biotechnology degree students who played the games were considered (86 answers for each final questionnaire). Table 53 shows the results of the Kolmogorov-Smirnov test and the Shapiro-Wilk test for normality of the four variables, performed on the average of the items of TAM, SUS DGBL and ENG. Based on the results, it can be said that the three variables SUS DGBL and ENG are normal, as both the Kolmogorov-Smirnov and the Shapiro-Wilk tests yielded non-significant p-values ($p > .05$), indicating that the null hypothesis of normality could not be rejected. The tests produced mixed results for TAM, as they show some evidence of normality according to the Kolmogorov-Smirnov test but not the Shapiro-Wilk test. Therefore, it can be concluded that the TAM variable is approximately normal.

	Kolmogorov-Smirnov test			**Shapiro-Wilk test**		
	p-value	Statistic	gl	p-value	Statistic	gl
TAM	0.061	0.066	172	0.023	0.982	172
SUS	0.200	0.044	172	0.627	0.993	172
DGBL	0.098	0.062	172	0.231	0.989	172
ENG	0.200	0.056	172	0.073	0.986	172

Table 53 - Normality of the variables.

To examine whether there were differences in students' perceptions of the serious games across the four variables (DGBL, SUS, ENG and TAM), independent-samples t-tests were conducted for the three variables that met the normality assumption (DGBL, SUS and ENG). As TAM showed a deviation from normality, both a t-test and a Mann-Whitney U test were performed. Participants consistently reported higher scores in the Murderer game than in the Alien game across the means of all four variables. Significantly higher DGBL scores were obtained after playing in the Murderer game (mean = 64.34) than after playing the Alien game (mean = 57.82). This difference is statistically significant (p-value= 0.019), with a small-to-medium effect size (Cohen's D= 0.36). A similar pattern emerges for the SUS and ENG scores. The Murderer game produced higher values in terms of usability (SUS mean = 62.88) than the Alien game (mean = 58.33). This difference is statistically significant (p-value = 0.022), with a small-to-medium effect size (Cohen's D = 0.35). The ENG score was higher for the Murderer game (mean = 58.60,) than for the Alien game (mean = 52.62). This difference is statistically significant (p-value = 0.002), corresponding to a medium effect size (Cohen's D = 0.48). Participants achieved higher TAM scores (mean = 60.96) in the Murderer game than in the Alien game (mean= 53.52). This difference is statistically significant for both the t-test (p-value = 0.029) and the Mann-Whitney U test (p-value = 0.033), with a small effect size (Cohen's d = 0.33). As shown in Table 54, the mean scores for the four variables (TAM, SUS, DGBL, and ENG) are approximately 60 or higher for the Murderer game. In contrast, the corresponding values for the Alien game fall below 60, with differences exceeding five points. Note that there are slight differences in the mean scores for the Alien game compared to those in Table 49. This is because only students who played both serious games were selected for the sample, and two students were eliminated for this reason.

		TAM	**SUS**	**DGBL**	**ENG**
Murderer game	Sample	86	86	86	86
	Mean	60.96	62.88	64.34	58.60
	SD	22.15	12.56	17.28	12.58
Alien game	Sample	86	86	86	86
	Mean	53.52	58.32	57.82	52.62
	SD	22.22	13.30	18.22	12.35

Table 54 - Differences in means for the two serious games across the four variables (DGBL, USA, ENG and TAM).

In the pilot study of the Murderer game, the aspects relating to personalised learning were the least appreciated. In response, the design of the Alien game was given a branched structure to explore whether it could provide a higher level of personalisation. However, mean scores of 57.70 and 49.20 for questions such as "*To what extent do you think the use of digital, game-based learning in Mathematics can promote a personalised learning approach?*" and "*How well does it meet your needs as a student?*" indicate that personalisation did not improve.

The mean scores of the factors identified through EFA and confirmed through CFA were examined using responses from the final questionnaire of students who played both serious games. As Factor 2 ("Perceived effectiveness for learning Mathematics"), Factor 4 ("Game experience and serious game appeal"), $Factor_{ENG}2$ ("Negative engagement") and $Factor_{ENG}3$ ("Competence, challenge and effectiveness in carrying out the game and its objectives") did not show evidence of normality, the Mann-Whitney U test was performed on these factors to investigate differences in their mean scores. Further factor-level analysis confirmed the observed trend for the four variables (DGBL, TAM, SUS and ENG), which favoured the outcomes associated with the Murderer game. Six out of seven of the investigated factors exhibited statistically significant differences in mean scores between the answers obtained after playing the two serious games. The t-test results revealed significant differences for Factor 1 (p-value = 0.017), Factor 3 (p-value = 0.037), and $Factor_{ENG}1$(p-value = 0.005). In each case, the mean score for the final questionnaire completed after playing the Murderer game was significantly higher than for the Alien game. Mann-Whitney U-test results also reveal significant differences for Factor 4 (p-value = 0.027), $Factor_{ENG}2$ (p-value = 0.007), and $Factor_{ENG}3$ (p-value = 0.009). However, there was no statistically significant difference for Factor 2 between the two serious game interventions (p-value = 0.064), suggesting comparable outcomes for this dimension regardless of the game played. Table 55 shows the mean scores for the seven factors obtained for both serious games. Note that there are slight differences in the mean scores for the Alien game compared to those in Table 52. This is because only students who played both serious games were selected for the sample, and two students were eliminated.

		Factor 1 "The impact of digital game-based learning on the Mathematics learning experience"	**Factor 2** "Perceived effectiveness for learning Mathematics"	**Factor 3** "The complexity of the game, possible difficulties"	**Factor 4** "Game experience and serious game appeal"	**$Factor_{ENG}$ 1** "Positive engagement"	**$Factor_{ENG}$ 2** "Negative engagement"	**$Factor_{ENG}$ 3** "Competence, challenge and effectiveness in carrying out the game and its objectives"
Murderer game	Sample	86	86	86	86	86	86	86
	Mean	63.42	60.27	64.42	52.38	58.49	62.49	54.01
	SD	18.96	22.11	14.27	19.06	18.99	13.56	14.28
Alien game	Sample	86	86	86	86	86	86	86
	Mean	56.19	53.76	59.65	46.28	50.19	58.60	48.78
	SD	20.39	22.63	15.37	19.38	19.56	11.93	14.95

Table 55 - Differences in means across the seven factors of the final questionnaire.

6.5.2 Exam Results: a Comparison of the Intervention and Control Groups

The exam results for the Biotechnology degree programmes in the 2024-25 academic year were compared with those from the previous two years (2021-22 and 2023-24). During this time, no serious game had been used to teach the probability and statistics portion of the course. Specifically, two questions were considered: one on descriptive statistics involving indices or regression, and the other on inferential statistics involving confidence intervals or hypothesis tests. Both topics were covered in the serious game.

The experimental group comprised 73 students who played the serious game at least once and took the exam in the 2024-25 academic year. The questions, which were scored on a scale from 0 to 3, were similar to those analysed in previous years. They focused on descriptive and inferential statistics - the same topic covered in the Alien game. The control group comprises students who took the exam during the 2021-22 (88) and 2023-24 (128) academic years, before the serious game was introduced.

As can be seen in Table 56, the scores for question related to descriptive statistics in the exam are higher in the 2024-25 academic year, when the serious game was introduced, than in previous years when it was not used. The percentage of students who obtained a score above the median (≥1.5) also increased from 58/59% to 67% in the 2024-25 academic year.

	2021-22 academic year (control group)	2023-24 academic year (control group)	2024-25 academic year (experimental group)
Mean score	1.80	1,72	1.97
Standard Deviation	1.16	0.95	1.00
Median score	2.40	1.5	2.00
Percentage of students obtained a score above the median score (≥1.5)	59%	58%	67%

Table 56- Comparison of the exam results of students on the Biotechnology degree programme between the control and experimental groups for the question related to descriptive statistics.

This trend is also evident in question related to inferential statistics, where the mean score in the 2024-25 academic year was approximately 0.3-0.5 higher than those in the previous two academic years (Table 57). There was also a positive effect on the percentage of students who obtained a score above the median (≥1.5), which increased from 52% in 2021-22 and 39% in 2023-24 to 62% in 2024-25.

	2021-22 academic year (control group)	2023-24 academic year (experimental group)	2024-25 academic year (experimental group)
Mean score	1.39	1.14	1.67
Standard Deviation	1.12	1.05	0.91
Median score	1.5	0.75	1.8
Percentage of students obtained a score above the median score (≥1.5)	52%	39%	62%

Table 57- Comparison of the exam results of students on the Biotechnology degree programme between the control and experimental groups for the question related to inferential statistics.

Fig. 21 represents cumulative distribution plot of exam scores, and it shows, for each score from 0 to 3, the proportion of students who achieved *at least* that score in the question related to inferential statistics. As can be seen in the plot, the line for the 2024-25 academic year (green) is almost always above, or takes on similar values to, the two lines referring to the control groups for the 2021-22 (blue) and 2023-24 (red) academic years. The most noticeable difference concerns mid-range scores, showing that a higher proportion of people reached at least a mid-range score in 2024-25. For higher scores, the curve for 2024-25 closely resembles that for 2021-22.

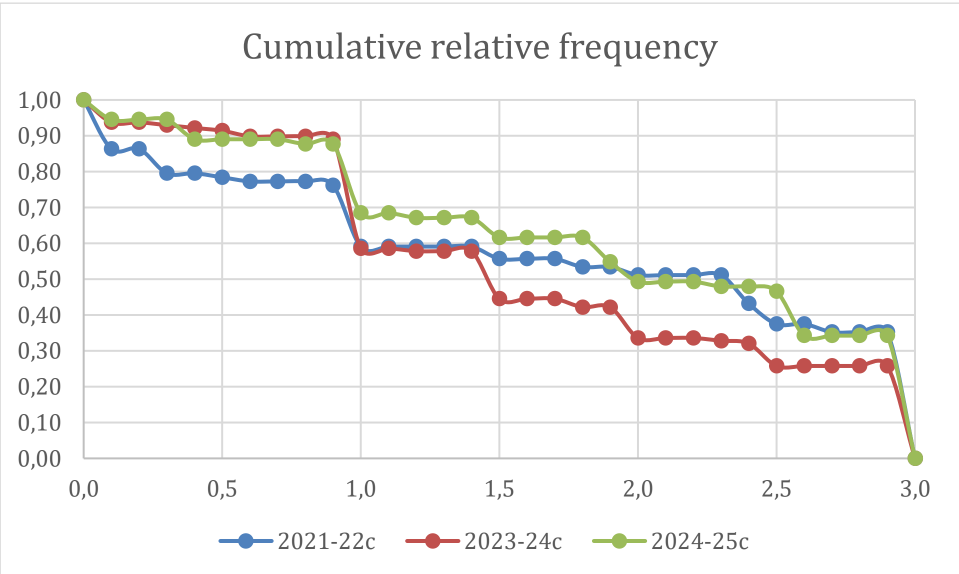

Figure 21 - Cumulative distribution plot of exam scores for the for the question related to descriptive statistics.

A more pronounced difference can be seen in the results for the experimental group relating to the question on inferential statistics (Figure 22). For lower to mid-range scores, the green graph clearly stands out, indicating a substantially larger proportion of students achieving at least a score of 2 (out of 3) than in the control group. For higher scores, however, the 2021-22 graph dominates. Therefore, while an improvement in reaching at least a mid-range score can be noted within the interval, no improvement in attaining a high score can be observed.

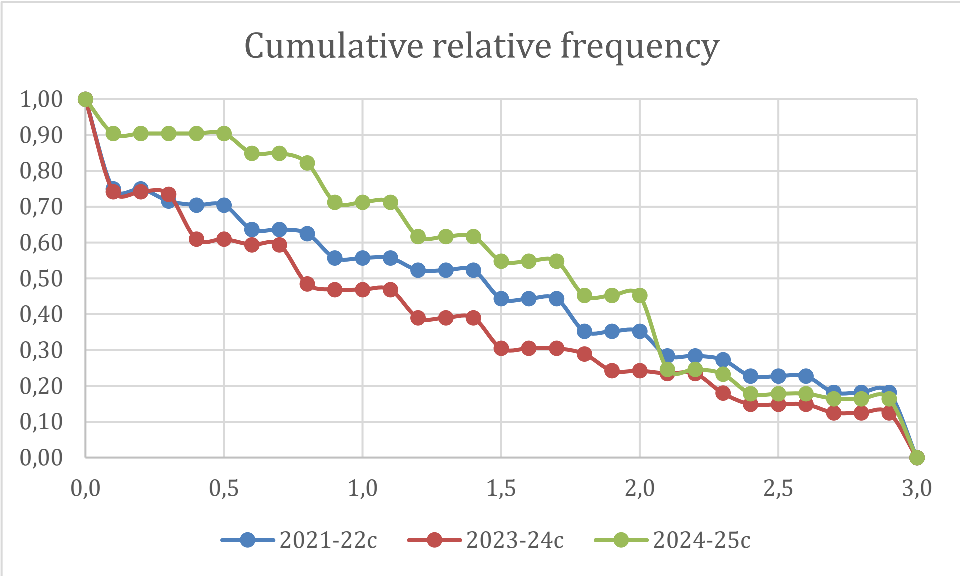

Figure 22 - Cumulative distribution plot of exam scores for the for the question concerning inferential statistics.

Table 58 compares the results of students taking the Biotechnology degree programme in the 2024-25 academic year who completed the serious game before the exam (73 students) with those who did not (29 students). The comparison is based on the average score obtained in the two questions related to the topic of the serious game. The former group achieved an average score that was higher, with a difference of approximately 0.2 points compared to the latter. The percentage of students who obtained a score above the median (≥1.5) is notably higher for the experimental group (71%) than for the control group (22%).

	Students played the serious game (73)	**Students didn't play the serious game (29)**
Mean score	1.82	1.60
Standard Deviation	0.76	0.88
Median score	1.84	1.65
Percentage of students obtained a score above the median score (≥1.5)	71%	22%

Table 58- experimental and control group in the 2024-25 academic year

6.5.3 Results from Game Sessions of the Alien Game

During the 2024-25 academic year, 88 students on the Biotechnology degree programme at university made a total of 169 gameplay attempts and completed both the initial and final questionnaires. The average number of attempts per student was 1.91, with a standard deviation of 1.01 (fewer attempts per person than in the Murderer game). Considering that the scale ranges from 0% to 100%, the average score obtained was 72.6% (similar to the mean score obtained in the Murderer game). The winning range is 75% to 100%.
Based on data from game sessions involving 73 Biotechnology students who took the exam, no correlation was found between average scores obtained in the various attempts at the serious game and average scores achieved in the two exam questions related to descriptive and inferential statistics (the game's topic). A second analysis was conducted to determine whether the average serious game score was more closely related to one

of the two exam questions. Again, no correlation was found. The correlation between the number of attempts completed in the serious game and the exam results was also examined, but no correlation was found in this case either.
Correlation analyses were conducted between the number of attempts completed in the serious game and the 88 participants' responses to the final questionnaire. No significant relationships were found. Specifically, the number of attempts in the serious game did not significantly correlate with mean scores on the DGBL, TAM, SUS and ENG scales. Similarly, no significant correlations were found between the number of attempts and any of the factors identified in the final questionnaire. Regarding the correlation between the average score obtained in the various attempts at the serious game and the students' responses to the final questionnaire, no statistically significant correlation was found between the average score obtained in the various attempts at the serious game and the seven factors identified in the final questionnaire. No significant or weak correlations were found also with the four variables of the final questionnaire (DGBL, SUS, TAM, ENG). These results confirm that, in terms of the number of attempts made, participants' performance in the second serious game was not associated with their reported perceptions of DGBL, usability evaluations, engagement levels or technology acceptance ratings.

6.5.4 The Use of GenAI in the Alien Game

In the final questionnaire completed after the serious game, 15 out of 88 students (17%) stated that they had used GenAI tools for support during the activity. This is a lower percentage than the number of students who used GenAI for support when solving the Murderer game. All the students come from the Biotechnology degree programme, which is the only one where the Alien game was tested. All students declared that they had used a chatbot, and in every case it was ChatGPT. Based on the qualitative responses to the final questionnaire, GenAI was mainly used (74% of students) to explain concepts that students had not yet studied or could not recall, or to clarify statistics concepts that they did not fully understand. Students then applied these concepts in the serious game. Some students also used the chatbot to obtain explanations of questions in the serious game. As only 15 students used GenAI to complete the serious game, compared to 73 who did not (almost five times as many), it was not considered worthwhile to calculate differences in mean scores between users and non-users, as was done in the analysis of GenAI use for playing the Murderer game.

6.6 Results from Experimental Activities in Upper Secondary Schools

The activities were also carried out with students from Italian upper secondary schools that belong to the national "Problem Posing & Solving" network of schools. The pilot study of the Murderer game involved upper Italian secondary school students from May to September 2024. The second study, which tested both the serious games, took place from January to September 2025. The sample consists of students who completed the initial questionnaire before playing the serious game, played the Murderer game at least once and then completed the final questionnaire. A total of 46 Italian upper secondary school students, from four different classes in four different schools, participated in the pilot study of 2023-24. The sample includes 25 female students and 21 male students. In terms of the type of school attended, four students were enrolled in a technical economic institute, while the majority (42 students) attended a scientific lyceum. In terms of grade level, 11 students were in the penultimate year of upper secondary school (Grade 12, as the Italian scholastic system has an extra year), while 35 students were in their final year. Overall, a total of 57 unique gameplay sessions were recorded. This corresponds to an average of 1.24 sessions per participant, indicating that most students played the serious game once, while a smaller proportion engaged in multiple game sessions. The mean score obtained in the game was 52.7%. This average performance was notably lower than that observed in the sample of university students, which was approximately 75% - the winning threshold. However, this difference is considered reasonable given the younger age of the participants in the present sample. Furthermore, while the topics addressed in the game corresponded to content that university students had recently covered in their courses, this was not the case for upper secondary school students. For these students, the game content represented a synthesis of approximately five years' curriculum, meaning they may not have recalled all the topics.

The level of appreciation for the game was assessed by analysing responses to the final questionnaire (the initial version used in the pilot study). Descriptive statistics were calculated for the mean, median and standard deviation of each variable (TAM, SUS, DGBL and ENG) on a scale of 0-100 (Table 59). The analyses were conducted by excluding the "N/A" (Not Applicable) responses, which were allowed for the DGBL and TAM items for those who felt they could not express an opinion on the use of DGBL in Mathematics. Since there were only a few "N/A" responses, the average was calculated considering only the questions where "N/A" was not selected.

	TAM	SUS	DGBL	ENG
N (sample)	46	46	46	46
Mean	41.46	51.65	47.02	46.21
Median	43.33	52	47.14	46
Standard deviation	25.73	10.48	21.11	9.27

Table 59 – Descriptive statistics for TAM, SUS, DGBL and ENG variables.

Overall, the mean scores for all variables were below or slightly above the midpoint of the scale, indicating moderate levels of perceived usefulness (TAM), perceived usability (SUS), perception of digital game-based learning for mathematics education (DGBL), and engagement (ENG). Specifically, the SUS score had the highest mean value (mean = 51.65, SD = 10.48), followed by the DGBL variable which had a mean score of 47.02 (SD = 21.11). The TAM (mean = 41.46, SD = 25.73) and ENG (mean = 46.21, SD = 9.27) variables had lower mean scores. Overall, these results suggest that students' evaluations of the serious game were generally moderate, with notable differences, particularly regarding perceived usefulness and the impact of DGBL on Mathematics learning. The responses obtained in the present study were generally low, especially compared to those collected from university student samples in previous studies. According to informal interviews and feedback from the teachers involved in the study, students often did not complete the activity in an attentive manner. Teachers reported that students often seemed impatient when filling in the questionnaires and, consequently, tended to give similar or repetitive answers, paying little attention to the precise wording of the questions. This interpretation is supported by an examination of the qualitative responses and open-ended comments included in the questionnaires, which also reflect a lack of depth in students' answers. This approach may account for the low scores in the questionnaire measures and the relatively low average performance in the serious game. Taking this feedback into consideration, a second experimental study was conducted to verify whether these impressions would also emerge in a new sample of students. Furthermore, analysis of the qualitative responses revealed that many students found the serious game complex in terms of the difficulty of its content. Examining the relationship between the average score obtained across multiple attempts at the serious game and questionnaire responses related to usability showed a positive Spearman correlation (r = 0.305, p-value = 0.039). This suggests that students who performed better in the serious game tended to give more positive evaluations of its usability. A similar pattern emerges when analysing the Spearman correlation between mathematics grades and usability scores (r = 0.325, p-value = 0.035).

A total of 74 Italian upper secondary school students, from five different classes in four different schools, participated in the second study conducted from January to September 2025. The sample included 29 female students and 43 male students, and 2 participants preferred not to indicate the gender. In terms of the type of school attended, 16 students were enrolled in a technical technological institute, while the majority (58 students) attended a scientific lyceum. In terms of grade level, 16 students enrolled in a technical institute were in the penultimate year of upper secondary school (Grade 12), as the Italian school system has an extra year. Meanwhile, 58 students were in their final year. The same 16 students played both serious games, and to avoid increasing respondent burden and potentially causing fatigue, they were asked to complete a single questionnaire. They were asked to refer to DGBL as the use of the two serious games for learning Mathematics. Regarding the Murderer game, a total of 111 gameplay sessions were recorded. This corresponds to an average of 1.50 sessions per participant. The mean score achieved in the game is 75.73%. This average score is notably higher than the scores obtained by secondary school students in the pilot study and higher than those obtained by university students. It is also slightly above 75%, the winning threshold.

Regarding the Alien game, a total of 16 gameplay sessions were recorded by 16 students. This corresponds to exactly one session per participant. The mean score achieved in the game is 52.77%. When examining the performance of the same 16 students in the Murderer game an average score of 47.73% was obtained. Therefore, the performance levels observed across the two serious games can be considered comparable.
The level of appreciation for the game was assessed by analysing responses to the final questionnaire (the revised version used in the second experimentation). Table 60 shows descriptive statistics calculated for the mean, median and standard deviation of each variable on a scale of 0-100. The analyses were conducted by excluding the "N/A" (Not Applicable) responses, which were allowed for the DGBL and TAM items for those who felt they could not express an opinion on the use of DGBL in Mathematics. Since there were only a few "N/A" responses, the average was calculated considering only the questions where "N/A" was not selected. In this second experiment, the average scores obtained for the DGBL, TAM, SUS and ENG variables were, in most cases, approximately 10 points higher. Furthermore, unlike in the pilot study, these mean values did not differ substantially from one another.

	TAM	SUS	DGBL	ENG
N (sample)	74	74	74	74
Mean	55.02	56.86	58.52	56.83
Median	57.14	56.00	58.46	56.00
Standard deviation	21.63	15.75	18.94	14.43

Table 60 – Descriptive statistics for TAM, SUS, DGBL and ENG variables.

The scores obtained for the various factors derived from the final questionnaire of the second experiment conducted in the university context were also calculated (Table 61). The results obtained are quite consistent with those already found for the total sample of 201 university students in the second pilot study (Table 38). The data show that Factor 2, "Perceived effectiveness for learning Mathematics", has the highest mean (62.84), similar to $\text{Factor}_{ENG}2$ "Negative engagement". Questions worded inversely, meaning that a positive response was given for low scores, were rephrased so that a positive response was given for high Likert scale scores. To compare the results with existing scales in the literature, the questionnaire was transformed from a 1-6 Likert scale to a 0-100 scale. By contrast, similarly to the university sample, Factor 4, "Game experience and serious game appeal", had the lowest mean score (50.00), suggesting that participants considered the overall game experience and the appeal of serious games to be less impactful. The median values largely confirm the trends observed in the means, indicating consistent central tendencies across factors.

Items	**Factor 1** "The impact of digital game-based learning on the Mathematics learning experience"	**Factor 2** "Perceived effectiveness for learning Mathematics"	**Factor 3** "The complexity of the game, possible difficulties"	**Factor 4** "Game experience and serious game appeal"	**Factor_{ENG} 1** "Positive engagement"	**Factor_{ENG} 2** "Negative engagement"	**Factor_{ENG} 3** "Competence, challenge and effectiveness in carrying out the game and its objectives"
Mean	57.69	62.84	56.49	50.00	55.90	62.43	54.05
Median	58.46	60.00	55.00	50.00	56.67	64.00	55.00
Standard deviation	20.70	22.03	18.15	21.52	23.06	14.62	13.96

Table 61 - Descriptive statistics for the extracted factors.

Examining the relationship between the average score obtained across multiple attempts at the Murderer game, and the responses to the questionnaire relating to engagement with (ENG), and perceived usefulness of (TAM),

the tool showed a positive Spearman correlation (r = 0.343, p-value = 0.03 and r = 0.248, p-value = 0.033, respectively). Thus, students who achieved higher scores in the serious game also tend to perceive it as more engaging and useful, and this relationship is statistically significant. Regarding the correlation between the average score obtained across multiple attempts at the Murderer game and the factors identified in the final questionnaire, the only significant relationship is with $Factor_{ENG}2$ (r = 0.270, p-value = 0.02). It is worth noting that, once again, also for upper secondary school sample, the DGBL_M item "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines*" obtained the highest mean and median within DGBL scale: 63.78 and 60, respectively.
To study the distribution of the variables, Kolmogorov–Smirnov and Shapiro–Wilk tests were conducted on the TAM, SUS, DGBL and ENG variables. Based on the results of these tests, all four variables are normally distributed (the p-values for the Shapiro–Wilk test is 0.15, 0.062, 0.15 and 0.68, and the p-values for the Kolmogorov–Smirnov test are 0.2, 0.2, 0.2 and 0.076 respectively). The results of the t-test for independent samples of the mean scores for the four variables revealed no statistically significant differences between the male and female groups. A one-way ANOVA was used to study differences in students' answers according to how frequently they play video games. The analysis confirmed the existence of at least one statistically significant difference in the means of the analysed groups for SUS (p-value = 0.023), TAM (p-value = 0.046) and DGBL (p-value = 0.049). Post-hoc comparisons using the Games-Howell test revealed a significant difference for the SUS variable only. Students who played the game less overall had a significantly higher mean (difference = 17.37) than those who played more. Thus, regarding the SUS variable, students who tend to play video games less gave higher usability ratings.
Teachers who had tested at least one of the two serious games in the classroom provided three responses to the final questionnaire. Particularly high values were obtained for the DGBL variable (mean item score of 4.6) and SUS variable (mean item score 4.5). Scores are rated on a scale from 1= strongly disagree to 6= strongly agree. Slightly lower results were obtained for the TAM variable (mean item score of 3.7). The lowest values in the TAM scale, with an average score of 3.0, were obtained for the items related to using serious games to speed up and simplify Mathematics learning. The teachers highlighted the following strengths: the serious game includes a challenge and uses mathematical language and objects to solve it, thereby supporting the learning of complex concepts; the theme is engaging; learners can be autonomous when searching for solutions to problems. Additionally, one teacher reported finding the proposal highly stimulating and suggested developing more games.

7. The Model

This chapter presents a new, adaptable model for designing digital, game-based activities that can be integrated into a DLE. This model builds upon the frameworks discussed in Chapter 2, which guided the design of the two serious games. It also incorporates insights gained from experimentations conducted with these games, as described in the previous two chapters. Indeed, the model was iteratively refined based on this practical experience, enabling continuous improvement and validation of its design principles. This chapter presents a new, adaptable model for designing digital, game-based activities that can be integrated into a DLE. This model builds upon the frameworks discussed in Chapter 2, which guided the design of the two serious games. It also incorporates insights gained from experimentations conducted with these games, as described in the previous two chapters. Indeed, the model was iteratively refined based on this practical experience, enabling continuous improvement and validation of its design principles. The model includes five phases that reflect the stages of this research project's development.

7.1 The model for DGBL in a DLE

Despite the increasing interest in game-based learning, the design of serious games remained largely unstructured. As Gentile et al. (2014) noted, clear guidelines based on a validated methodological framework were lacking. Later, some models have emerged in the literature and have been examined within the theoretical framework of this study. These include models developed by Plass et al. (2016), Garris et al. (2022) and Lee et al. (2023). Drawing on available models and on the design of the two serious games, as well as the results obtained from experimental studies, a new adaptable model was developed for designing digital game-based activities that can be integrated into a Digital Learning Environment (DLE). In a similar manner to Lee et al. (2023), who designed a gamified Mathematics course, the model is based on the ADDIE instructional design framework (Peterson, 2003). The model is intended to support teachers who may lack advanced digital competencies and design expertise, and it emphasises ease of use. It incorporates the principles of gamification and digital game-based learning which can be integrated into a DLE. It considers learning goals, game design elements and how to implement them in the classroom. All materials are developed using a design template to ensure consistency between learning objectives, gamification strategies, and technological implementation. The model consists of five phases that can be grouped into two main areas (Figure 23). The first focuses primarily on the design, while the second is centred on implementation in the classroom and evaluation of the learning experience, with a view to identifying any necessary refinements. In line with Vidakis et al. (2015), the model comprises a variety of interconnected tasks that could be integrated synergistically. The teacher is responsible for the game content and the educational theories underlying the game, either independently or in collaboration with educational experts. Meanwhile, game designers and programmers are responsible for the ludic and technical aspects of the game. Teachers implement the game in the classroom, where students use it as part of their learning process. Since the model does not require advanced technological skills or specialised expertise in game design, teachers with a basic level of training in game-based learning can actively engage in the game design process, potentially with the support of contemporary technologies such as Gen AI.

1. The Analysis phase focuses on understanding the profile of the learners involved and on defining clear learning goals, paying particular attention to the content to be included into the digital game-based activities, competencies and skills to be developed. To stimulate students' attention and engagement, it is recommended that contemporary topics that are relevant and relatable to students, as well as real-world applications, are covered. Curricular objectives are translated into measurable learning outcomes at this stage, which can then be addressed through interactive and game-based activities in the following steps. This ensures coherence with formal educational settings while remaining flexible enough to support other disciplinary contexts. An example of the learning content analysis in the design of the Murderer game is shown in Chapter 4, Table 3. This phase involves also reviewing the literature to identify various theoretical frameworks. These may include key concepts and topics related to gamification, as well as frameworks for designing questionnaires. These questionnaires are developed and administered in subsequent phases.

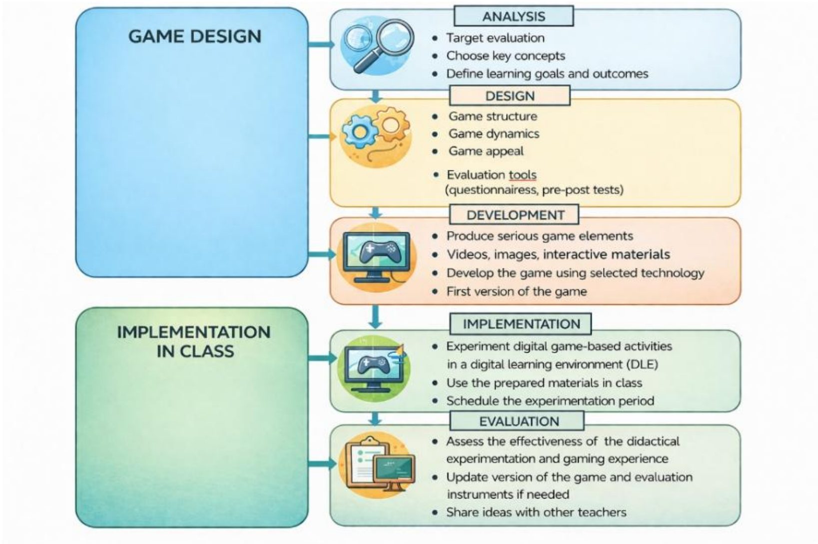

Figure 23 – The model for designing digital game-based activities integrated into a DLE.

2. The Design phase is the core of the proposed model, where the curriculum is transformed into game-based components. In this phase, designers decide whether to focus on structural or content gamification. In other words, they decide whether to transform the learning content into a game-based experience or not, or whether to combine the two. The design phase encompasses the first two stages of "game structure" and "game dynamics", as identified in the gamification process introduced in Chapter 4, in relation to the design of the two serious games. The game structure (Table 62) defines the framework of the gaming experience and includes rules, the game genre, the narrative context, the goal of the game and the progression. Game dynamics describe player actions, expected behaviours, strategies and responses to challenges, explicitly identifying conditions for winning, losing or partial success at each step (Table 63). In this sense, game dynamics correspond to the "game mechanics" defined by Plass et al. (2016) intended as the core actions or behaviours associated with learning or assessment tasks within a game. Moreover, Wiboonsin and Kasemsukpipat (2022) established a clear link between the game mechanics and the disciplinary concepts involved in the serious game designed to improve financial literacy. This scheme was adopted to align expected learner behaviours and actions with the game's dynamics and the underlying theoretical constructs (Table 63). The use of structured and design templates can be particularly effective in supporting the design of these three components, as they allow learning and gamification elements to be collected systematically and subsequently implemented within the technological environment on which the game is built. Another suggestion to reduce design complexity is to adopt recognisable narrative structures or existing game schemas, enabling learners to focus on learning objectives rather than unfamiliar mechanics (Nautiyal et al., 2024). Furthermore, teachers can opt for linear or branched structures if supported by the technological tool employed. The latter enables adaptive learning pathways based on learners' responses. This phase makes it clear how gamification and digital game-based learning interact effectively. The various elements of gamification, such as storylines, points, badges and challenges, are thoughtfully integrated into the design of the serious game. These elements are designed to promote engagement and motivation, while simultaneously guiding learners in developing skills and achieving learning outcomes, as well as providing immediate feedback.

Main characteristics of a game	
Title	The murderer at the High-Tech Institute.
Goal	Help Bill find the assassin who murdered Teresa, a mathematician at the High-Tech Institute.
Context	The story opens with Teresa's murder, immediately setting out the central challenge of determining who committed the crime. The first clue suggests that the route taken by the murderer corresponds to the graph of a real-valued function. At this point, Kathy, a mathematician at the High-Tech Institute, enters the scene. Although she was unable to recognise the killer, who was wearing a disguise, she observed several details of the route taken closely. These observations are translated into a sequence of mathematical clues, which are formulated as properties that the function must satisfy for its graph to coincide with the murderer's trajectory. Once all the clues have been analysed, the player must evaluate the possible suspects and select the one whose movements are consistent with the reconstructed path.
Rules	The game consists of 12 steps, with a maximum of four answer options for each step. Players can earn a maximum of 20 points per step by answering correctly. With a total of 63 points, the game is won by scoring

	at least 48 points. The game does not allow users to go back once a choice has been made. It is possible to see the score at the end of the serious game, which ranges from 0% to 100%. This percentage reflects the number of points obtained relative to the total number of points available in the game. The winning range is set between 75% and 100%. Some answers are only partially correct, meaning that the user can receive a lower score for them than for completely correct answers. The attached files contain useful hints for progressing in the game, so watching the videos is mandatory. Each step and its attached files can be viewed as many times as is required until a choice is made. Once the game is complete, it is possible to view the final score, along with all the gameplay details. A histogram is also provided with the bars corresponding to the answers given at each step, from the first to the last. A bar at 100 represents a completely correct answer, while a bar at 0 represents a completely incorrect answer. Partially correct answers have an intermediate value between 0 and 100 depending on the accuracy of the response. The game can be played as many times as you want. This will help you improve your performance.
Single or multiplayer game	Single player, but collaboration among students is allowed.
Structure	Both winning and losing profiles at each step led to the same next step through a linear structure.
Steps	12.
Characters	Narrator voice, Teresa, Tom, Bill and Kathy.
Technological components	The E-core editor was used, which can be integrated with any LMS.
Other features	As it is a STEM-related game, it would be preferable to have a text editor that allows formulas to be inserted.

Table 62 – Example of game structure from the Murderer game

Step n.	Concept	Game action	Winning profile behaviour	Losing profile behaviour
3	Identify a suitable reference system to illustrate the problematic situation.	Several reference systems are presented, and players must select the one they consider to be the most appropriate.	Partially winning profile: the player selects a reference system that is suitable for representing the character's path in this stage only. Winning profile: the reference system is also suitable for representing the characters' paths in subsequent steps.	Impulsive behaviour: the player traces the character's path using the available information without identifying an appropriate reference system.
4	Within an appropriate and generalisable reference system, players are asked to identify the path taken by the suspect.	Several paths are represented in different reference systems are presented, and players must select the path they consider to be the most appropriate.	The player recognises the graph representing Emily's path among the proposed ones.	The player does not recognize the graph representing Emily's path among the proposed ones.
5	Determine whether a graph represents a function.	Emily's path is described, and the first clue provided by Kathy is that the murderer's trajectory represents the graph of a function. The player is therefore asked whether Emily could be the culprit.	The player recognises that Emily's path does not represent a function.	The player does not recognize Emily as a potential killer because she/he fails to recognise that her path does not represent a function.
8-9	Recognise and apply when a function has a symmetry (even or odd) and when it is quadratic.	Additional clues are provided to the player, who must identify, among the various paths, those that satisfy the conditions of being quadratic functions and exhibiting even or odd symmetry.	The player uses all the available clues (even/odd function and quadratic function) to correctly translate the requirements that the function must satisfy for Jack to have passed through the hall and to be guilty.	The player fails to consider all the available clues or is unable to translate them into the requirements that the function must satisfy to implicate Jack.
10-11	The properties of the exponential function and the derivative of a function must be recognised and applied.	Additional clues are provided to help the player identify the paths that satisfy the conditions of being quadratic functions, exhibiting even or odd symmetry, and having the required derivative.	The player uses all the available clues, such as the even/odd quadratic function and the derivative, to correctly translate them into the requirements that the function must satisfy for Mark not to be considered a potential suspect.	The player fails to consider all the available clues or is unable to translate them into the requirements and considers Mark as a potential suspect.

Table 63 – Example of the expected actions and behavior for the winning and losing profiles, and the concepts associated with them (Game dynamics).

Step	1		
Characters	Narrator voice:  Tom:  Bill:  Kathy: 		
Main sentence	Narrator voice: Something very serious has happened at the High-Tech Institute. You absolutely must help Bill! Listen carefully to this audio message (Phone ringing).		
Attachments	Video of the phone call between Tom and Bill. Dialogues: Tom: Good morning! This is Inspector Tom. Who am I speaking with? Bill: This is Professor Bill. I'm calling from the High-Tech Institute. I found my dear colleague Teresa dead! I had come to invite her to join me for lunch, and I found her waiting for me on the small sofa in the middle of the ground-floor lounge. When my smartwatch showed exactly 12:00, I tapped her on the shoulder, then I went to wash my hands in the bathroom next to the lounge. When I came back, I found her dead on the sofa. Tom: Can you tell me what time she was killed? Tom: How am I supposed to know?		
Answers option	1.1 To determine the exact time of the murder, Bill should provide a few more details.	1.2 The murder must have taken place between 12:00 and 12:05. The bathroom is next to the lounge, so it would have taken him a maximum of five minutes to wash his hands.	1.3 While the exact time is unknown, an estimate can be made.
Score	10	0	10
Next step	Step 2	Step 2	Step 2
Feedback		Perhaps Bill could provide us with a few more details.	

Table 64 – Game appeal: the design of audiovisual and multimedia elements, such as images, videos, animations, dialogue and sounds, in the opening scene of the Murderer game.

3. The Development phase involves translating the design elements into concrete materials to bring the various components of the game to life. It covers the third and last "game appeal" phase of the gamification process. Game appeal addresses the audiovisual and multimedia elements, such as images, videos, animations, dialogue and sounds, that support cognitive processing and emotional engagement (Table 64). It includes the production of videos, images, interactive resources, quizzes, and serious game elements using appropriate digital tools and platforms. This also includes developing the game using plug-ins or editors such as E-Core, which allow integration with Moodle, as well as using GenAI tools to produce contents such as images, videos and avatars. To guarantee that the system functions correctly and efficiently, this phase requires the involvement of a programmer/ IT professional. Careful attention is given to usability, accessibility and engagement. This ensures that the final product effectively conveys educational content and provides an immersive, motivating experience that is easy to use, and it is accessible. Using an accessible and easy-to-use tool facilitates the adoption of DGBL in schools and universities, while also reducing the barriers posed by difficulties in using digital tools (Huang et al., 2013). During this phase, the first version of the game is developed and released for the didactical experimentation. This version is identified as v0.0 during the design phase of the two experimental serious games. The development phase also involves producing ad hoc materials for collecting feedback on the proposed digital, game-based activity. These materials include initial and final questionnaires, as well as pre- and post-tests.
4. The Implementation phase consists of experimenting digital game-based activities with students in educational settings. Teachers use the prepared materials in a DLE, incorporating them into regular classroom activities, blended or online learning scenarios. During this phase, it is crucial to ensure that teachers are properly trained with the necessary skills and knowledge to deliver instructional materials effectively. At this stage, the ethical issues associated with research involving students should also be given careful consideration. Where feasible, this phase also includes the revision of the first version of the game by a team of experts prior to its use with students. In the implementation phase, teachers adopt an observational role, monitoring both the learning process and the gameplay experience. This includes aspects such as engagement and usability. Every stage of the implementation phase is carefully planned.

Clear timelines, roles, and procedures are defined in advance to support both teachers and students throughout the experimentation. For instance, Table 65 provides a detailed overview of how the experimentation with the two serious games proposed was structured, including the sequence of activities, the duration of each phase, and the tools used to support the learning experience. At this stage, both teachers and students have access to tools that monitor the learning process. Teachers can, for example, review gameplay data to identify the most frequent errors and difficulties encountered by students and potentially link these to specific content that has not yet been mastered. Furthermore, students can immediately view their scores, receive feedback on their performance and identify where mistakes were made.

Order	Activity	Duration
1	Access to the course on Moodle.	5 min
2	Completion of the initial questionnaire.	15 min
3	Reading and consulting the rules of the serious game from the Moodle ad hoc page.	15 min
4	Access the serious game via the Moodle platform and play it once or more.	40 min
5	Completion of the final questionnaire.	15 min

Table 65 – Example of timetable for the didactical experimentation of the Murderer and Alien game.

5. The Evaluation phase assesses the effectiveness of the didactical experimentation and of the overall gaming experience. This phase involves various types of assessment, such as summative and formative assessments, or a combination of the two. It also involves taking direct or indirect measures, such as assessing learning outcomes through tests or gauging students' perceptions of the impact of DGBL and their level of engagement and determining how easy they find it to use. Based on these results, teachers will reflect on the design process and on the integration of the game into didactic practice. They evaluate whether revisions are needed and whether a new version of the game should be produced to address elements that did not worked as intended. The evaluation outcomes then inform an iterative refinement process that aims to align the activity as closely as possible with students’ needs as identified through the collected feedback, thereby reinforcing the adaptability of the model across different contexts. Drawing inspiration from the work of Nautiyal et al. (2024), who identified a five-step framework for designing and implementing educational board games and GBL sessions, this phase also offers opportunities to adapt the developed game for different subjects, collaborate with other educators to design new games, and share ideas and best practices. In addition, the need to update and revise evaluation instruments, such as questionnaires and pre- and post-tests, is considered.

Overall, this ADDIE-based model for DGBL offers a structured yet flexible framework that bridges gamification and digital game-based learning within a DLE. By prioritising didactical alignment, design simplicity, and practical classroom implementation, the model supports sustainable adoption by teachers, and it enables the effective use of DGBL in Mathematics and other disciplines. The model promotes the reuse of digital resources and supports continuous evaluation and improvement. This makes the approach adaptable and long-lasting in different educational contexts.

8. Discussion and Future Developments

This section examines the results obtained in this study, contextualising them within existing research and literature, and establishing links between them. It also outlines the innovative nature of the research and its potential implications. It also highlights how the findings align with, extend or challenge previous studies, offering a critical interpretation of the outcomes. Finally, it details possible future developments of the research project.

8.1 The Innovativeness of the Research and Its Possible Implications

This research involved studying and developing methodologies for designing digital game-based activities. The developed model is intended to encourage the creation of game-based open educational resources for integration into a Digital Learning Environment (DLE). The research followed four main directions:

1. The design and development of innovative game-based learning activities to improve students' approaches to Mathematics at upper secondary and university level, involving students from scientific and non-scientific degree programmes.
2. The didactical experimentation of the two serious games with upper secondary school and university students and their teachers to evaluate the impact of the tools and the proposed teaching methodologies on learners.
3. The evaluation of the impact of the proposed teaching methodologies, with a particular focus on the development of disciplinary and transversal skills, as well as students' engagement with and perceptions of the discipline.
4. The development of an adaptable, digital game-based learning model integrable into a DLE.

Before the development of the various materials, a systematic literature review of game-based learning in Mathematics education was needed to provide an overview of the current state of the art and to identify well-established and emerging trends. This represents a key innovative aspect of the project, as relatively few literature reviews address this topic, particularly focusing on higher school grades. This systematic literature review identified 36 relevant studies. These studies were analysed, and the key findings of the selected papers were summarised, paying particular attention to those related to the objectives of this thesis. Specifically, the review examines the effect of GBL and serious games on student motivation, engagement and math anxiety in Mathematics education, as well as its impact on learning outcomes. The review also explores tools for evaluating serious games and other game-based learning resources in Mathematics. The findings show that GBL is applied more frequently in primary education than in secondary and higher education due to curricular constraints, limited teacher training and attitudes towards games. Motivation and engagement are two of the most widely researched variables. While most studies indicate beneficial effects, these may vary according to educational level, prior skills and content complexity. The results regarding engagement and math anxiety are somewhat mixed. Evidence suggests that GBL can enhance problem-solving skills and higher-order thinking; however, its effectiveness depends heavily on game design, classroom integration, and teacher expertise. The review also highlights the absence of shared frameworks for measuring engagement and math anxiety, and emphasises the need for more rigorous, long-term evaluation studies, particularly at higher educational levels. This review has two limitations: only two databases were used (Google Scholar and Scopus) and the selection strategy included only studies addressing the objectives of this thesis, which may have resulted in the exclusion of relevant papers.

Following a thorough review of the literature, the design of two serious games intended to facilitate and improve students' approach to Mathematics was undertaken: one focusing on functions of a single variable, and the other centred on probability and statistics. The design phase did not focus exclusively on the two serious games. Another objective of the research was to identify a valuable tool with which to measure the impact of the proposed teaching methodologies in terms of technical and game-related aspects, such as usability, game appeal, engagement and perceived usefulness, as well as in relation to discipline-related dimensions, including motivation, attitudes towards Mathematics and math anxiety. The study also aimed to identify the strengths and weaknesses of both the serious games and the evaluation tool. This enabled the tools to be refined and improved, enhancing the students' experience. Two studies were conducted to evaluate the serious games and their associated tools: the pilot study during the 2023-24 academic year and the main study during the 2024-25 academic year. The studies involved upper secondary school students and their teachers, as well as non-mathematics university students. The latter group comprised students from four different degree programmes in the 2023-24 academic year and students from five in the 2024-25 academic year. These students were enrolled on the undergraduate courses in Medical Radiology, Imaging and Radiotherapy Techniques, in Biotechnology, in Prevention Techniques in the Environment and Workplaces, and on the postgraduate course in Strategic and Security Sciences. The experimentation during the 2024-25 academic year involved also undergraduate students in Strategic and Security Sciences and PhD students on the Doctoral Programme in Digital Humanities at the University of Genoa (in collaboration with the University of Turin), who attended the "Gamification and Serious Games" course.

8.1.1 The Design and Development of Innovative Game-Based Learning Activities

These game-based digital resources were designed using a variety of tools and implemented within a DLE to support Mathematics teaching and learning. The novelty of the project lies in:

- The research of innovative solutions that leverage technology, such as the use of an editor to develop and integrate a serious game into a DLE, to design game-based activities for Mathematics learning. The selected technologies are easily and accessible to both students and teachers, enabling the design and implementation of new learning activities without advanced digital skills, thereby promoting inclusivity and the broader adoption of game-based learning approaches.
- Considering the current state of the art, this project aims to develop digital game-based resources that stimulate not only disciplinary skills, but also problem-solving and critical thinking. The World Health Organization (WHO, 2003) recognises these as "life skills", which are essential today.
- The project emphasises player engagement and interaction, fostering active participation and meaningful learning experiences. At the same time, it aims to provide teachers with valuable information about the learning process to support them in monitoring students' progress, assessing learning outcomes and refining teaching practices.

Designing the game-based activities proved highly challenging. The most complex aspect was achieving a balance between mathematical content and gamification elements. Given the limited background and experience in this specific area, an in-depth review of the relevant literature was essential to address this challenge and explore how existing serious game design projects have been developed. Based on the findings of the literature review conducted within this research project, relatively few studies on the design and development of serious games for Mathematics at higher school levels were identified. In particular, the statistics presented in Chapter 2 reveal that studies on game-based learning in the context of primary education are more prevalent than those at upper secondary and higher education levels. Another key strength of this study is the large sample size and the fact that the experiments were conducted twice in two academic years, which enhances the robustness of the analyses and increases the reliability and generalisability of the results.
In parallel, templates and structured frameworks were key in guiding the definition of instructional content, learning outcomes and the development of various game components. Several key characteristics were given particular attention in the design of the activities. Specifically, the serious games were designed to support personalised learning paths, enabling learners to progress according to their individual needs and abilities. The activities also aimed to increase student engagement and encourage active participation in the learning process, promoting a learner-centred approach. Furthermore, the activities were designed to motivate students to engage with Mathematics by presenting mathematical concepts in meaningful and interactive contexts. Inclusivity and accessibility were core design principles, ensuring the games could be used effectively by learners with diverse learning styles and levels of prior knowledge. Selecting the content to be included in the serious games proved particularly demanding. For "The Murderer at the High-Tech Institute" (the Murderer game), the focus was placed on fundamental mathematical topics to ensure the content was accessible to a wide range of learners, including upper secondary school and university students. This approach was intended to maximise the game's

applicability and reusability across different educational levels. The second serious game "Don't open that spacecraft" (the Alien game) focused on more specific topics related to probability and statistics. These topics were chosen because they provide an essential link, enabling students - and, more broadly, all citizens - to interpret and understand data in today's world. Not all these topics are always included in curricula, however. Another key design principle was to include not only disciplinary content, but also activities aimed at fostering transversal skills such as problem solving and critical thinking. This research was carried out during a period of rapid growth and widespread adoption of artificial intelligence (AI). Consequently, it became necessary to acknowledge this shift and explore ways to coherently incorporate AI into the research project, despite this not having been originally planned, since, at the time the project was defined, AI had not yet reached this level of adoption. Efforts were made to understand how Generative AI (GenAI) tools could be integrated into game-based activities. GenAI tools were used during the design phase of the Alien game and were made available to students to use while playing the game during the 2024-25 study.
Two didactical experimentations were planned to determine whether modifications to the developed products were needed based on the results of the first experimentation. Analysing the results was highly insightful, encouraging critical reflection on the final products and highlighting potential improvements to enhance clarity, content, accessibility, and student understanding. Based on the pilot study, the Murderer game was revised, with these considerations taken into account in the design of the Alien game. Changes were also made to the final questionnaire, which was used to collect students' observations after they had played the serious game.

8.1.2 (RQ1) How Can the Effectiveness and Impact of the Developed Serious Games be Evaluated, both in Terms of Technical and Game-Related Aspects and Discipline-Related?

One stage of applying the model involved developing a final questionnaire to measure the impact of serious games on students across five dimensions:

1) Impact of digital game-based learning on Mathematics learning.
2) Students' perceived effectiveness of the serious game for learning Mathematics.
3) Need for support, training and complexity of the game.
4) Gameplay experience and appeal of the serious game.
5) Engagement, in terms of students' involvement, as well as their experience of flow and immersion while playing the game.

A final questionnaire tailored specifically to the context of Mathematics was developed (Table 66) to evaluate the serious games designed. The labels adopted for the items are consistent with those used in the previous chapters: SUS (usability), TAM (perceived usefulness), DGBL (the impact of digital game-based learning on Mathematics learning) and ENG (engagement).
The final version of the questionnaire considers all the analyses conducted, including item refinement, reliability assessment and validation of the factor structure. Table 66 contains the abbreviated set of ENG items. Nevertheless, Table 67 presents the extended version of the ENG scale for a more detailed and in-depth measure of engagement. In fact, based on the results of the pilot study, it was decided that the items on the ENG scale should be shortened, as they were considered too long. While the shortened ENG scale provides a more concise measurement tool, some of the finer details present in the original, more extensive structure are inevitably lost. One notable aspect that is not captured as clearly in the shortened scale is "flow", a concept widely regarded as a key element of engagement in learning and gaming environments (Chauhan et al., 2021; IJsselsteijn, et al., 2013). For this reason, the reduced scale is suitable for minimising administration time, or when a general measure of engagement is sufficient. It is also useful in contexts such as the present study, where engagement is being measured alongside multiple dimensions. Using a shorter scale in such cases helps to limit participant burden. However, if engagement itself is the primary focus, the full, extended version of the questionnaire is preferable for a comprehensive, multidimensional assessment.

To what extent do you think that the use of digital game-based learning (i.e. the use of games for learning, such as the Murderer game) in Mathematics can	*ID*
Enhance learning.	DGBL_A

Increase your engagement.	DGBL_B
Promote an inclusive learning approach.	DGBL_D
Promote a cooperative learning approach.	DGBL_E
Promote a personalized learning approach.	DGBL_F
Stimulate your interest.	DGBL_G
Stimulate interaction.	DGBL_H
Enable skills development.	DGBL_I
Meet your needs as a student.	DGBL_L
Reduce of the gender gap in scientific disciplines.	DGBL_M
Increase your participation in the learning process.	DGBL_N
Improve your perception of Mathematics.	DGBL_O
Reduce math anxiety.	DGBL_P
Allow you to achieve my learning goals faster.	TAM_A
Improve your performance.	TAM_B
Increase your effectiveness.	TAM_D
Make Mathematics easier to learn.	TAM_E
Can be useful for your studies.	TAM_F
Develop your critical thinking skills (i.e. their ability to analyse facts, evidence and observations to make reasoned judgements).	TAM_G
Develop your problem-solving skills.	TAM_H
Based on your experience as a player of the Murderer game, to what extent do you agree with the following statements?	ID
I think I would like to use this game a lot.	SUS_A
The content and topics of the game are complex.	SUS_B
I found the game easy to use.	SUS_C
I found that the different features of this game were well integrated.	SUS_E
This game has a lot of things that interest me.	SUS_H
The scenes in this game are very attractive.	SUS_L
The contents were easy to read.	SUS_M
The contents were easy to understand.	SUS_N
Thinking about your experience as a player of the Murderer game, indicate how you felt during the game	
I was interested in the game's story.	ENG_A
I thought it was fun.	ENG_B
I thought about other things .	ENG_D
I felt competent.	ENG_E
I thought it was hard.	ENG_F

It was aesthetically pleasing.	ENG_G
I felt bored .	ENG_J
I enjoyed it.	ENG_M
I was fast at reaching the game's targets.	ENG_N
I felt pressured.	ENG_O
I felt challenged.	ENG_P
I was deeply concentrated in the game.	ENG_R
It felt like a rich experience.	ENG_T
I felt disoriented .	ENG_Z
The game feels real.	ENG_AF

Table 66- Definitive version of the final questionnaire to evaluate a serious game for the field of Mathematics (version with the abbreviated ENG items).

Thinking about your experience as a player of the Murderer game, indicate how you felt during the game	
I was interested in the game's story.	ENG_A
I thought it was fun.	ENG_B
I was fully occupied with the game.	ENG_C
I thought about other things.	ENG_D
I felt competent.	ENG_E
I thought it was hard.	ENG_F
It was aesthetically pleasing.	ENG_G
I forgot everything around me.	ENG_H
I lost track of time.	ENG_I
I felt bored.	ENG_J
I felt that I could explore things.	ENG_L
I enjoyed it.	ENG_M
I was fast at reaching the game's targets.	ENG_N
I felt pressured.	ENG_O
I felt challenged.	ENG_P
I was deeply concentrated in the game.	ENG_R
I felt frustrated.	ENG_S
It felt like a rich experience.	ENG_T
I lost connection with the outside world.	ENG_U
I found it hard to get back to reality.	ENG_V
I felt time pressure.	ENG_W

I felt satisfied.	ENG_Y
I felt disoriented.	ENG_Z
I felt exhausted.	ENG_AA
I felt powerful.	ENG_AB
I felt weary.	ENG_AC
I felt proud.	ENG_AD
I feel like I just can't stop playing.	ENG_AE
The game feels real.	ENG_AF

Table 67- Extended version of the ENG items.

Seven factors were identified in the EFA within the final questionnaire for the version including the abbreviated ENG scale, and ten factors were identified for the extended ENG scale version (Table 68). This factorial structure requires the addition of the item "*To what extent do you think that the use of DGBL in Mathematics can reduce the gender gap in scientific disciplines*", since it captures a unique aspect related to the perceptions of DGBL and its potential to reduce the gender gap in scientific disciplines. While it does not fit well within the established factor model, having shown very low factor loadings also in the pilot study, it may still be valuable in understanding a separate dimension of participants' views.
The original TAM item "*I would find this product useful in my job*" was not included in the TAM scale as the participants are students who are still far from entering the world of work, meaning the item should not be appropriate in this context. It was therefore replaced by the TAM_F item, "*To what extent do you think that the use of DGBL… in Mathematics can speed up learning Maths*". However, this modification may have been somewhat extreme, and it could be preferable to reinstate the original TAM item and adapt it to the current context of using the serious game, for example by changing it to "*To what extent do you think that the use of DGBL… in Mathematics can be useful for your studies*". This modification was implemented in Table 66. However, the model was tested using the revised item "*To what extent do you think the use of DGBL can speed up learning Mathematics?*".

Extened version	**Factor 1** "The impact of digital game-based learning on the Mathematics learning experience"	**Factor 2** "Perceived effectiveness for learning Mathematics"	**Factor 3** "The complexity of the game and possible difficulties"	**Factor 4** "Game experience and serious game appeal"	**Factor$_{ENG}$ 1** "Enjoyment and involvement"	**Factor$_{ENG}$ 2** "Negative feelings and pressure"	**Factor$_{ENG}$ 3** "Competence and achievement"	**Factor$_{ENG}$ 4** "Flow"	**Factor$_{ENG}$ 5** "Loss of control and game realism"	**Factor$_{ENG}$ 6** "Disengagement and boredom"
Items	DGBL_A DGBL_B DGBL_C DGBL_D DGBL_E DGBL_F DGBL_G DGBL_H DGBL_I DGBL_L DGBL_N DGBL_O TAM_B TAM_C TAM_D	DGBL_P TAM_A TAM_E TAM_F	SUS_B SUS_C SUS_F	SUS_A SUS_E SUS_G SUS_H SUS_L	ENG_A ENG_B ENG_C ENG_G ENG_L ENG_M ENG_T	ENG_O ENG_S ENG_W ENG_Z ENG_AA ENG_AC	ENG_E ENG_N ENG_Y ENG_AB ENG_AD	ENG_H ENG_I ENG_R ENG_U	ENG_V ENG_AE ENG_AF	ENG_D ENG_J

Abbreviated version	Factor 1 “The impact of digital game-based learning on the Mathematics learning experience”	Factor 2 “Perceived effectiveness for learning Mathematics”	Factor 3 “The complexity of the game and possible difficulties”	Factor 4 “Game experience and serious game appeal”	Factor$_{ENG}$ 1 “Positive engagement"	Factor$_{ENG}$ 2 “Negative engagement”	Factor$_{ENG}$ 3 “Competence, challenge and effectiveness in carrying out the game and its objectives”	
Items	The same as row 2				ENG_A ENG_B ENG_G ENG_M ENG_R ENG_T	ENG_F ENG_J ENG_O ENG_Z ENG_AF	ENG_E ENG_D ENG_N ENG_P	

Table 68- Factors extracted by the factorial analyses.

It would be interesting to extend the use of this tool to other disciplines to evaluate serious games within a broader educational context. Applying it across different subject areas could reveal how effectively serious games support learning outcomes beyond Mathematics and STEM disciplines. Extending the research in this way would help to generalise the findings, refine the evaluation tool and provide practical guidance for teachers and game developers working in a variety of fields.
Similar analyses were conducted in other studies to identify emerging factors in various GBL contexts. The factor “Game experience and serious game appeal” aligns with the results of the exploratory factor analysis conducted by Huang et al., which identified “game appeal” as one of three converging factors of DGBL. Many studies in the DGBL field have adopted the TAM model. For example, Cardona Valencia (2023) et al. proposed an adaptation of the TAM model to evaluate consumer behaviour; their five-factor model comprises the following: Ease of use, Perceived usefulness, Attitude, Confidence and Intention to use. Moreover, Chauhan et al. (2021) identified eight factors in their adaptation of the TAM model: Social interaction, Use context, Perceived ease of use, Perceived enjoyment, Perceived usefulness, Attitude, Flow and Behavioural intention. However, it is worth noting that some of these factors previously identified in the literature, such as perceived usefulness and ease of use, align with those found in this exploratory factor analysis. Although the shortened ENG scale offers a more concise and manageable measurement tool, it inevitably loses some of the finer details captured by the original, more detailed structure. The multi-dimensional evaluation framework presented above shows that serious games can be thoroughly assessed in terms of their technical and game-related aspects, as well as their impact on Mathematics learning. Combining the TAM, SUS, DGBL and ENG scales, and supporting them with Cronbach's alpha calculations, EFA and CFA analyses provides a reliable tool for gaining in-depth insights into students’ experiences and perceptions of the serious games.

8.1.2.1 Patterns and trends in students' answers to the final questionnaire

To understand the patterns and trends within the DGBL, TAM, SUS and ENG variables, descriptive statistics were calculated on a scale of 0-100, based on the average of the items composing each variable, to make meaningful comparisons with results from the literature. These comparisons should also be interpreted with caution since adjustments were made to the items of the scale. On average, university students rated all four variables as moderately positive, slightly above the median value of the scale (50). Regarding the results obtained from the final questionnaire completed after at least one attempt at the Murderer game, students who participated in the second experiment during the 2024-25 academic year (N=201) gave more positive overall evaluations than in the previous study (N=156). Furthermore, the standard deviations decreased, suggesting that participants' responses were more consistent and varied less. These improvements may be related to the enhancements made to the game components, such as the improved voice-over feedback, the clearer introductory explanations, and the attention paid to making the characters more realistic. Both experiments

involving university students in serious games show a similar trend, with higher values obtained for the SUS and DGBL scales than for the TAM and ENG scales.
Despite a slight decline in the second experimentation (60.73 compared to 63.38 in the pilot study), the SUS remained one of the variables with the highest scores, alongside the DGBL. In their literature, Bangor et al. (2009) consider a SUS rating between 60 and 70 to be good. However, this value is still lower than the average SUS rating of 68 computed by Sauro (2011). The study by Tolentino et al. (2011), using the SUS in serious games, suggests that a score of 50, or the midpoint of the scale, would be sufficient to classify the software as usable. However, in their study different values for usability were obtained, namely 58.46 and 82.61. This may suggest that there is no fixed reference value for usability in serious games, as scores can vary depending on factors such as the complexity of the game, the target audience and the overall user experience. In any case, this result suggests that there are still areas for improvement in usability in the Murderer game.
The items of the TAM scale obtained lower mean values than those of the SUS scale, remaining below 60 in both experimentations, although the scores improved in the 2024-25 study (56.85 vs. 59.86). In both studies, the highest standard deviation was observed for the TAM variable, suggesting a considerable spread in students' opinions and experiences regarding the acceptance of the serious games, particularly in terms of perceived usefulness. The results obtained in relation to TAM were compared with those obtained when the model was used to evaluate other serious games. For example, in Deborah's serious game by Malaquias et al. (2018), users rated perceived usefulness using TAM with an average value of 4.0 on a scale from 1 to 5, which represents a very positive value. The results of this thesis regarding the perceived usefulness of the serious game are more consistent with the findings of Bourgonjon et al. (2010). They reported an average value of 2.74, which, on a scale from 1 to 5, is quite similar to the values obtained in the two experimentations. Nevertheless, it is important to reflect on which aspects of the game may have contributed to students perceiving it as less useful and how it might be improved.
As the items identified with the DGBL scale are newly developed, it is not possible to make comparisons with the literature. Nevertheless, it can be concluded that, on average, respondents had a fairly positive view of the use of DGBL for learning Mathematics, as indicated by the mean score of the DGBL items in both experiments (61.50 vs. 62.39).
Examining the descriptive statistics for the four factors extracted from the SUS, DGBL and TAM variables reveals that students found the game technically accessible and easy to use, and that they had a positive perception of the impact of DGBL on their Mathematics learning experience. Factor 2 "Perceived effectiveness for learning Mathematics" and Factor 4 "Game experience and serious game appeal" received lower evaluations. This suggests that, although students generally enjoyed the experience, they did not find the game itself particularly appealing. Consistently, the ENG scale produced the lowest values in both experiments. On average, the items of this scale received mean scores of 50.75 and 57.52 respectively in the first and second experiments with university students, indicating a modest level of engagement among participants. Morteover, correlation analysis revealed a strong positive Pearson correlation between engagement and Factor 4 "Game experience and serious game appeal". The engagement improvement from pilot to the main study could be because the ENG scale was halved, so it required less effort from the students. This may have led to lower levels of fatigue and boredom when filling in the questionnaire, thereby reducing the likelihood that students became impatient and provided more negative or random responses. Moreover, it was found that engagement is a key element in shaping the gaming experience and appeal of a serious game, as it significantly predicts Factor 4. Therefore, the low levels of engagement observed in the studies may have influenced the gaming experience and the overall appeal of the serious game. The ENG scale also showed the lowest standard deviation of all the scales, suggesting that participants' responses were consistent and showed little variation around the mean. The values obtained for the "Flow" factor in the pilot study were among the lowest of all the ENG factors. Similarly, the game developed by Ntoa et al. (2024), evaluated using the Game Experience Questionnaire (GexpQ), obtained low values for "Flow" and "Immersion", with mean values of around 1.9 corresponding to 47.5 on a 0-100 scale. Similarly, the category of "Flow" obtained the lowest score in the study by Cruz et al. (2023), at 2.8 out of 5. The authors suggested that lower scores were associated with experimental conditions, such as supervision by instructors, and aspects of game design that reduced emotional tension.

In the pilot study, the degree programme had a weak influence on the DGBL and ENG scales, but not on the other variables. Students on the Bachelor of Medical Radiology, Imaging and Radiotherapy Techniques programme had the lowest mean scores across all variables, while those on the Strategic and Security Sciences postgraduate programme provided more consistent and favourable responses. In the subsequent study, however, significant differences emerged between degree programmes for all variables. Biotechnology students gave the most positive evaluations, whereas Radiology and Prevention students reported the lowest scores again. However, no correlations were found between their upper secondary school grades and the variables and factors of the questionnaire.

In both experiments, female students rated the game more positively than male students. This pattern may be partly explained by the fact that male participants played video games more frequently and may therefore have adopted a more critical perspective. Moreover, the item "*To what extent do you think that the use of DGBL in Mathematics can reduce the gender gap in scientific disciplines*" obtained the highest positive score among all DGBL items across all experiments, for both secondary school and university students. The results showed no statistically significant differences in the answers given by male and female students, meaning that they show comparable levels of agreement in their responses to this item. This suggests that tools such as serious games may be perceived as effective instruments for reducing gender differences in scientific disciplines. This is particularly relevant given the traditionally male-dominated nature of the video game field. At the outset of this project, there was a concern that game-based learning approaches might be more readily embraced by male students than female students. Since female students tend to have less prior experience with video games on average, there was a risk that they could feel disadvantaged or excluded. However, the results suggest that female participants generally evaluated the serious games and their use in learning Mathematics more positively.

Qualitative feedback confirmed that the game was considered interesting and well structured. However, in the pilot study it also revealed that some students found the rules difficult to understand. This information was used to make subsequent improvements to the serious game. Qualitative comments confirmed that students largely viewed the game as a useful tool for revising and applying mathematical concepts, and in some cases as a valid tool of assessment

The Alien game was tested with 88 undergraduate students enrolled in Biotechnology Bachelor's degree programme. It received moderately positive evaluations across all group of items. Comparative analyses of the two serious games revealed that the Murderer game consistently achieved higher mean scores across all four variables (SUS, DGBL, TAM, and ENG), with differences of approximately five points compared to the results obtained in the second study. The trend for the Alien game was consistent with that observed for the Murderer game. The SUS and DGBL measures obtained the highest mean scores (58.25 and 57.75, respectively). ENG showed again the lowest mean score (52.77), while the TAM variable occupied an intermediate position (53.47). As with the Murderer game, TAM exhibited the highest standard deviation (22.04), indicating substantial variability in students' perceptions of perceived usefulness. The lower evaluation of the second serious game may be because it covers more specific and complex topics than the first one. To identify the critical aspects of the serious game, comments from students regarding perceived difficulties and suggested improvements were analysed. Apart from a few isolated remarks, such as the need to clarify the content of the questions (three students) and improve the animations (two students), no substantial or recurring issues emerged. As one student commented, an additional explanation is that, unlike the other serious game, this one cannot be completed by relying solely on prior knowledge of the covered topics. While the mathematical concepts involved in the Murderer game (functions of a single variable) are extensively covered in upper secondary school, the topics included in the Alien game are not usually studied beforehand.

In both serious games, the items "*The contents were easy to read*" and "*The contents were easy to understand*" received the highest mean values on the SUS scale. This suggests that serious games were perceived as being very accessible in this sense. From the perspective of readability and comprehension, therefore, the Alien game was evaluated more positively. When analysing the responses to questions about the extent to which DGBL can foster critical thinking (TAM_G) and problem-solving (TAM_H) skills, higher mean scores were obtained for the Alien game. This suggests that students perceived more opportunities to develop these skills in the Alien game. This is consistent with the initial design idea, given that this second serious game is more closely linked to the ability to read and interpret real-world data and news and therefore requires these skills.

Examining the correlations between students' responses to the questionnaire and their performance in the serious games revealed that there were typically no significant correlations, or only weak ones. The only weak correlations observed were those relating to scores obtained in the Murderer game, the SUS scale in the pilot study and ENG-related factors in the second study.
The serious games, particularly the Murderer game, were also experimented with upper secondary school students. A total of 46 Italian students from four upper secondary schools and four different classes participated in the 2023-24 pilot study. Descriptive statistics show that students' evaluations of the serious game were generally moderate and lower than those obtained for the university sample. According to informal interviews and feedback from the teachers involved, students often did not complete the activity carefully. The reduced motivation among upper secondary school students to play the serious game could be partly explained by the absence of bonus point incentives, which were offered to university students but not to this group. Consistently with university sample, the SUS and DGBL measures obtained the highest mean scores, while the TAM and ENG items had lower mean scores. Students who achieved higher mean scores in the serious game tended to provide more positive feedback about its usability, with a weak correlation ($r = 0.325$). A total of 74 Italian upper secondary school students, from five different classes in four different schools, participated in the second study. Further analysis of the data from the second experiment was possible, as it was conducted more thoroughly and rigorously. In this experiment, the average scores for the DGBL, TAM, SUS and ENG variables were, in most cases, around 10 points higher than in the pilot study. There were no significant differences between the four groups of items, which achieved similar values ranging from approximately 55 to 59 on a scale from 0 to 100. The results obtained from the factors extracted from the final questionnaire are consistent with previous findings for the total sample of 201 university students in the second pilot study. The $\text{Factor}_{\text{ENG}}$ 2 “Game experience and serious game appeal” had the lowest mean score (50.00), suggesting that participants considered the overall game experience and the appeal of serious games to be less impactful. Moreover, secondary school students with stronger mathematical skills (considering the Mathematics grades) tended to give more positive evaluations of its usability.
In this case, no statistically significant differences were found between the male and female groups. However, about the SUS variable, students who reported playing video games less frequently gave higher usability ratings. This indicates that the Murderer game is accessible also to those who don't usually play video games. Their teachers emphasised the ability of the serious games to engage students in challenging mathematical problems and promote autonomous learning. However, they were more cautious about their ability to simplify or speed up the learning of mathematics.
The Murderer game was also experiment with seven PhD students in Digital Humanities at the Universities of Genoa and Turin during the 2024-25 academic year. Overall, the game was well received: the students found the storytelling motivating, the embedded mathematical content engaging, and the subtitles appreciated for their inclusivity. The only reported criticism was that the educational component could become more explicit as the game progresses, which would detract from the playful element.

8.1.3 (RQ2) Do Serious Games Designed Improve Students’ Approach to, and Learning Outcomes in Mathematics?

To identify any changes in students’ attitudes towards Mathematics before and after engaging with the serious game, attitude-related questions were taken from validates questionnaires. The same questions were included in the initial (before playing the game) and final (after playing the game) questionnaire to enable a pre-post comparison. These items considered the following dimensions: Perception of Mathematics, Interest and Motivation in Mathematics, Math anxiety, Persistence in Mathematics, Persistence in problem solving, Development of critical thinking, Perception of the course, and Extrinsic motivation.
Notably, math anxiety and extrinsic motivation decreased among biotechnology students in both didactical experimentations, with a stronger effect observed in the pilot study. This indicates that students experienced less pressure and were less driven by external rewards when engaging with Mathematics in the game context. Overall, math anxiety decreased across all degree programmes, with reductions of up to one point. The findings regarding math anxiety are consistent with the descriptive statistics, which revealed more favourable scores in areas relating to negative experiences. Specifically, $\text{Factor}_{\text{ENG}}$ 2 “Negative feelings and pressure” had one of

the highest positive values of all the factors extracted in the first pilot study. These findings suggest that students did not experience significant negative events or feel overwhelmed by pressure when playing the serious game. Note that questions that were worded in such a way that a positive response was given for low scores were rephrased so that a positive response was given for high scores on the Likert scale. The results are consistent with those of the study by Ayyasy and Asrul (2024), which found a significant decrease in math anxiety from the initial to the final questionnaire. This substantial reduction in math anxiety may also be related to the findings of Verkijika and De Wet (2015), who reported that math anxiety tends to increase with age. This is possibly because older students are generally exposed to more complex mathematical problems than younger students. Consequently, higher initial levels of anxiety are expected among university students, with the effect being more evident by the end. The pilot study revealed a slight decline in persistence in Mathematics for the Prevention Techniques in the Environment and Workplaces and the Medical Radiology, Imaging, and Radiotherapy Techniques sample, suggesting lower self-reported persistence after the game. This result was not confirmed in the second experiment. Slight improvements of around 0.2 were observed in the second study for undergraduate and postgraduate students on the Strategic and Security Sciences programme in the Perception of Mathematics, Interest and motivation, and Extrinsic motivation dimensions.

The results of the exam sessions, which compared the learning outcomes of the experimental group (students who played serious games) with those of the control group (students who did not play serious games in previous or the same academic years), generally indicate better performance by the experimental group than the control group, where mainly traditional methods were employed. Students from the Biotechnology and Medical Radiology, Imaging and Radiotherapy Techniques degree programmes were considered for comparison in the exam. The level of difficulty of the assessments was the same for experimental and control groups. In both the 2023-24 and 2024-25 studies, the average results obtained for both experimental samples were significantly higher than in the previous academic year, when the serious game was not employed. For the Biotechnology sample, it was also possible to compare the results of students in the same 2024-25 academic year who completed the serious game before the exam with those who did not. The former group achieved an average score that was approximately 0.7 points higher than the latter. However, it should be noted that students who did not participate in the serious game may have had a lower level of interest in Mathematics prior to the study, and their poorer exam performance may not be solely attributable to their lack of participation in the game. Overall, better performances were observed in the 2024-25 academic year. However, the results for the 2023-24 academic year were higher than those for the academic year in which serious games were not used. An analysis of the students' Mathematics background revealed no statistically significant differences between the 2023-24 and 2024-25 cohorts based on their upper secondary school grades. It was not possible to make the same comparison for students from previous academic years, as no information about their secondary school grades is available. It should be noted that improvements were made to the serious game during the 2024-25 academic year. Therefore, it is hoped that the enhanced performance is also a result of these improvements.

To further evaluate the impact of the Alien game on learning outcomes, the exam results relating to the topics of the game were considered. Once again, the experimental group performed better, achieving a mean score approximately 0.3-0.5 points higher than the control group in the previous academic years. Similarly, comparing students in the 2024-25 academic year who played the serious game before taking their exams with those who did not, it was found that the former achieved an average score that was approximately 0.2 points higher.

The results obtained are consistent with the findings of the systematic literature review presented in Chapter 2. This review showed that the most part of the selected studies provided evidence of the positive impact of GBL and gamification on Mathematics learning outcomes (Pan et al., 2022; Gusti, 2025; Crnković et al., 2022; Jiménez-Hernández, 2021; Magat, 2023; Gil-Doménech, 2017; Christopoulos et al., 2024; Wisconsin & Kasemsukpipat, 2022; Ayyasy & Asrul, 2024; Kebritchi et al., 2010). Albano et al. (2020) identified additional mathematical benefits within the context of game-based activities, including the development of problem-solving strategies. This may suggest that the higher exam performance observed in this study may partly be due to students being able to apply strategies more effectively during the exam after receiving training through the serious game sessions. However, some studies have reported no statistically significant differences in

Mathematics performance between game-based and traditional approaches, or between experimental and control groups (Maisey et al., 2022; Chen et al., 2023; Crocco et al., 2016).
Examining the results from the game sessions, the data show that a higher average number of attempts at the Murderer game were made during the 2023-24 pilot study than in the 2024-25 second study (2.67 vs 2.05 per student). However, the average score achieved in the various attempts was lower than in the second study (70.27% vs. 72.90%). Even though a small number of attempts were made in the second study, the highest exam performance was also reflected in the best performance in the serious game. Data from students enrolled in the Biotechnology degree programme who subsequently took the exam showed a weak positive correlation (approximately $r = 0.20$) between the average scores achieved across multiple attempts at the serious game and the average scores on the two exam questions related to functions, in both studies. This correlation was also observed in the second study for students enrolled in the Medical Radiology, Imaging and Radiotherapy Techniques programme with a stronger correlation coefficient ($r = 0.43$). This suggests that students who performed better in the serious game also tended to perform slightly better in the exam. However, no statistically significant correlation was found between exam results and the number of attempts made at the serious game. This suggests that the number of attempts at the serious game may be less indicative of subsequent success in the exam. Therefore, it cannot be concluded that more attempts led to better exam performance. On average, students attempted the Alien game slightly below two times (1.91 attempts for person), achieving an average score of 72.6%. This is comparable to the mean score obtained in the Murderer game. However, unlike the other serious game, no correlation was found between the number of attempts at the game or the average scores obtained and the average scores achieved in the two exam questions related to descriptive and inferential statistics. Notably, the mean score achieved by secondary school students in the second study for the Murderer game was found to be 75.73%. This average score is higher than those obtained by university students, and it is slightly above the 75% threshold required to win the game.

8.1.4 (RQ3) How Can Game-Based Activities Be Designed to Be Effectively Employed by Teachers in Their Instructional Practices and by Students in Their Learning Process?

This project began immediately after the pandemic, at a time when new learning strategies were required. As Harrington and Mellors (2021) have highlighted, students who engage extensively in video and computer games and other IT activities have become accustomed to more frequent access to digital resources during the pandemic. They tend to find more traditional learning methods lacklustre and unengaging. In the post-COVID-19 pandemic context, which is characterised by the widespread use of technology and extensive online and blended learning in education, two serious games for Mathematics learning were integrated into the Learning Tools Interoperability (LTI) framework within the Moodle LMS. Their main aim is to facilitate and enhance students' approach to Mathematics, to stimulate problem-solving and critical thinking skills, by providing an engaging and interactive way to reinforce learning and to put into practice the concepts introduced in the classroom. Serious games were designed to reinforce knowledge and skills that have already been introduced in the classroom. The games are accessible to students anytime and anywhere, and they are intended as a supporting tool to learn Mathematics. One reason for not sharing the solutions to the serious games was to encourage multiple attempts to successfully complete them. In this way, students could attempt the games as many times as they wished to improve their performance. They were conceived as a tool to enhance development, rather than as a means of initially acquiring these concepts. However, from the analysis of GenAI usage, it was found that students used the second serious game to familiarise themselves with new concepts that they had not yet studied. The games were made accessible to students through an editor integrated into Moodle. This meant that students could access the game directly from the platform they normally use for Mathematics courses without having to navigate to additional webpages or use different login credentials to those they already use daily to access university services. Furthermore, at the end of the game, students could immediately view their results and identify where errors had been made directly on the game interface. They could also subsequently review the outcome of their attempt through the Moodle gradebook. This feature is particularly useful for students in terms of monitoring and reflecting on their learning process. Based on the descriptive statistics, students particularly appreciated the usability of the serious game, considering it a valuable learning tool for learning Mathematics. The results on usability confirm that students found the game easy to access. However, some students had different opinions about the perceived usefulness of the game, which may be because they were not accustomed to this type of approach in previous school years. As expected, engagement was the least appreciated aspect. This may be because today's students are familiar with highly polished and sophisticated games.

From a teacher's perspective, the tool employed to develop the two serious games is characterised by its flexibility and ease of use. No advanced technical or programming skills are required to design a serious game, but it should nevertheless be specified that the maintenance of the editor necessitated the involvement of a programmer to ensure the proper and optimal functioning of the system. This means that, with a little training and adequate resources, any teacher can create a serious game, as the interface is highly intuitive. This means that teachers can effectively integrate DGBL into their teaching practices with minimal effort, regardless of their digital competencies (Fissore et al., 2023a). Furthermore, even if the game were designed by more experienced game designers, the integration of LTI with Moodle enables instructors to monitor the learning process directly within the course. They can access and review gameplay data directly within the Moodle course, as this data is not exclusively available to the game designer but are shared in the Moodle register. The evaluations provided by teachers who had used serious games in the classroom highlighted very positive usability ratings, thus confirming that the tools are easy to use in educational settings. In this case, the teachers were not acting as game designers. Making the game accessible to students and teachers can facilitate the adoption of DGBL and overcome the challenges associated with implementing digital tools in educational settings (Huang et al., 2013; Qasim et al., 2024).

8.1.5 (RQ4) What Advantages Can New GenAI Tools Offer in Terms of Developing and Using Serious Games?

In recent years, artificial intelligence has become an increasingly integral part of everyday life, permeating a wide range of educational and professional contexts. The widespread adoption of GenAI in many different areas has raised questions about its potential application in digital game-based learning.

Thus, the integration of GenAI tools was explored in this study, and preliminary insights were identified to inform future research. However, the integration of GenAI in this study remains largely exploratory as it was introduced during the project rather than being part of the original research design. From a methodological perspective, GenAI was introduced for two main purposes:

- To explore its potential to support the development of serious games and provide practical advantages for teachers during the design phase, particularly by reducing the time and effort needed to create instructional materials.
- Students were given the opportunity to use GenAI tools for support while playing the game. For instance, they could use the tools to recall or clarify concepts related to the game content that they did not fully understand or had forgotten. The study also aimed to investigate how students used GenAI tools and to what extent they did so during gameplay.

In this study, new Generative Artificial Intelligence (GenAI) tools offered several advantages in the development and use of serious games. Open online tools were used for image generation, scenarios and to create the alien avatar, Allan-Meta. These tools were also employed in developing the storyline of the serious game. This experience shows that, from a design perspective, GenAI can significantly speed up the creation of educational materials and game resources. Through the accessibility of current tools, designers and teachers can quickly create content, scenarios, feedback and supporting materials, reducing development time and overcoming technical and design barriers. This is a significant advantage for teachers who want to incorporate gamification into their teaching but are discouraged by the time it takes to develop instructional materials. For example, GenAI makes it possible to create engaging narratives and badges that can be awarded to students when they achieve specific goals (Fissore et al., 2024a; Fissore et al., 2024b).

Responses to the final questionnaire revealed that only a small proportion of students declared that they had used GenAI tools while playing serious games. Specifically, 38% of students reported using GenAI tools during the game session of the Murderer game, with usage distributed across the different degree programmes. An even smaller proportion (17%) reported using GenAI tools for the Alien game. The most used tools were AI chatbots and assistants such as ChatGPT, Copilot and Gemini. For the Murderer game, mean scores achieved by students who used GenAI were higher compared to students who didn't use GenAI tools. However, it was not possible to infer an improvement in performance on the serious game resulting from the use of GenAI. This may be due to the small sample of students who reported using GenAI tools. Further studies with larger samples are therefore needed to investigate this relationship more effectively. From a usage perspective, GenAI tools can support students in learning and reviewing content, as the qualitative comments collected in this study also revealed. Students reported using GenAI-based tools to help them clarify and revise content while engaging with the serious games. It also emerged that, in the Alien game, it was occasionally used to explain concepts that students had not yet studied or could not recall.

These findings should be considered preliminary within the broader context of integrating GenAI tools into gamified learning environments. Further research is needed to improve the understanding of its potential and limitations. Future work should aim to integrate GenAI more systematically and intentionally within game-based learning environments. For example, GenAI could support the development of personalised gamification strategies where game narratives, challenges and feedback adapt dynamically to students' progress and prior knowledge. GenAI could also be used to provide students with immediate feedback to guide them throughout the learning process. However, the effective implementation of these approaches requires further investigation through more structured research designs and larger-scale studies.

8.1.6 The Development of a Digital Game-Based Learning Model and Its Implementation into a DLE

A new model for DGBL that can be implemented into a DLE was developed through an evolutionary process. Starting with a review of existing literature, the model was iteratively refined based on feedback collected during experimental phases. The practical experience gained from using and evaluating the developed serious games progressively informed and shaped the model, ultimately resulting in the framework presented in Chapter 7. The model is intended to support teachers lacking advanced digital competencies, design expertise or research experience. It places a strong emphasis on ease of use, and it incorporates key principles of gamification and DGBL. The model is based on the ADDIE instructional design framework (Peterson, 2003), which provides a clear and sequential structure. Its phases - Analysis, Design, Development, Implementation and Evaluation - closely reflect the steps taken within the design of the two serious games presented above. Each phase comprises tasks and responsibilities for each stage and every actor involved.

The developed model is technology-independent, meaning it can be used to create digital game-based activities regardless of the technology used. The E-core editor (available at https://github.com/EntropyKN/ecore3.0) was used in the design of the two serious games, the Murderer game and the Alien game, and this can be integrated into a DLE via LTI. However, any other tool that is accessible or reachable within a DLE can be employed. The model incorporates Gen AI tools to support and simplify the design of game-based activities. For example, it can be used to create scenarios and characters. This does not diminish the fact that it requires ongoing rethinking and adaptation as technology evolves. As new tools, platforms and digital capabilities emerge, the model may need updating to ensure it remains effective and relevant. The model developed demonstrated that well-structured, game-based activities can enhance learning in Mathematics and improve attitudes towards the discipline, particularly regarding math anxiety and learning outcomes. However, more studies are needed to understand the impact on engagement and perceived usefulness. A critical reflection on the work carried out in this project reveals that greater attention should be paid to game appeal during the development phase, as this is crucial for achieving high levels of student engagement.

One of the main limitations of the model is its ability to ensure personalised learning. Although it offers a structured framework, it is limited in its ability to adapt learning pathways to the individual needs, preferences and prior knowledge of learners. In the pilot study of the Murderer game, the aspects relating to personalised learning were the least appreciated. In response, the design of this second serious game was given a branched structure to explore whether it could provide a higher level of personalisation. However, results indicate that personalisation did not improve. Therefore, it is impossible to conclude that either structure promotes personalisation more effectively. Another important point to note is that the model does not explicitly encourage collaboration between students. While teachers have the flexibility to design the game as either single-player or multiplayer, the model does not offer specific guidance on designing, implementing or actively

encouraging student collaboration. Consequently, key collaborative learning elements such as peer interaction and collaborative learning are not systematically embedded in the model. Another important point to note is that the model requires access to digital tools. In educational contexts where technological resources and infrastructure are limited, implementing the model may be difficult.
The key features of the model include flexibility and adaptability in any DLE, allowing customization not only for Mathematics but also for a broad range of academic disciplines. The proposed model offers a structured yet flexible way of integrating gamification and digital game-based learning in a DLE. Its five-phase structure - Analysis, Design, Development, Implementation and Evaluation - ensures alignment between learning objectives, game mechanics and technological implementation. It also supports iterative refinement based on student feedback and learning outcomes. This flexibility means the model can be applied and adapted effectively across different disciplines and educational contexts. Teachers are granted a high degree of autonomy throughout the model's phases: during the Analysis phase, they define topics and learning objectives, making the model applicable to a wide range of subjects; during the Design phase, they select and implement game-based components, tailoring them to the specific educational context. Furthermore, as the model is technology-independent, educators can select the most suitable tools according to the available resources. The model's modular structure, combined with the use of reusable design templates, allows for easy adaptation, updating and transfer across courses and disciplines, without the need for a complete redesign. Consequently, the model is flexible, scalable, and sustainable, supporting its long-term adoption and applicability in diverse educational settings.
The model is designed to facilitate the implementation of digital game-based activities by both students and teachers. It provides teachers with guidance and monitoring tools to support the integration of DGBL into their didactical practices. It also provides students with innovative tools to help them learn and apply their mathematical skills in an original, game-based context. As well as supporting skill development, it enables students to monitor their own learning process by providing feedback on where mistakes were made and the scores achieved in each game session. One of the model's key strengths lies in its easy access to resources and sustainability. As noted by Qasim (2024), game strategies must be sustainable in different educational contexts and over the long term. This entails a robust design, flexible implementation approaches, and continuous evaluation and adaptation.
In response to the low levels of engagement observed in the studies, the model was refined to place greater emphasis on engagement and interactivity. Particular attention was devoted to the target analysis phase to ensure closer alignment with students' needs, real-life situations and expectations. During the design phase, the selection and integration of game mechanics was given increased focus, with a view to promoting the use of more interactive and engaging elements, such as video content and storytelling techniques.
These findings also prompted a stronger emphasis on aligning learning activities with meaningful, recognisable real-world applications to enhance students' perceived usefulness of the learning experience.
The model also includes an evaluation phase, during which decisions can be made regarding modifications and improvements to the initial version of the game. The implementation phase is therefore essential for developing a practical, hands-on approach, as well as for gathering feedback that can inform potential revisions and improvements. To this end, the final questionnaire presented in paragraph 8.1.2 was developed to evaluate serious games and assess their impact on Mathematics learning. However, it could be interesting to extend its application to other domains.

8.1.7 Theoretical implications of the results

The experiments conducted and their results allow to draw several relevant theoretical implications for DGBL, particularly in the context of Mathematics education. Firstly, whereas previous research has typically concentrated on individual factors such as motivation, engagement and learning outcomes, this study adopts a broader perspective, examining multiple factors and exploring how they interact with one another in some cases. One of the main outcomes of this study is the development of a final, ready-to-use questionnaire designed to evaluate serious games. It represents a reliable tool for gaining in-depth insights into students' experiences and perceptions of a serious game. It allows for the assessment of multiple dimensions: the impact of the serious game on learning, perceived usefulness, technical complexity and usability, game experience, game appeal and engagement. Although the questionnaire was developed and validated in the context of Mathematics education, it can easily be adapted for use in other disciplines to evaluate serious games in a broader educational context. The results obtained from the final questionnaire confirm the multidimensional nature of the game-based learning experience by demonstrating the presence of multiple factors covering the dimensions mentioned above. The results show that the two serious games performed better on scales related to learning impact and usability than on those related to engagement and perceived usefulness. This pattern can be explained by the fact that serious games are often designed with a strong focus on educational effectiveness and ease of use to ensure learning content is accessible and clear for a wide range of users.

Using serious games in Mathematics education has a positive impact also on learning outcomes, particularly among non-Mathematics students, who may have a weaker background in the subject. This suggests that serious games, which were designed as reinforcement tools aligned with the topics covered in class, have effectively support the learning process. The results clearly demonstrate an improvement in learning outcomes, highlighting the potential of the serious games developed as an innovative, interactive approach to strengthening students' understanding of Mathematics. However, no meaningful correlation was found between the number of attempts at the serious game or the scores obtained and students' performance in the exam. Compared to usability and learning outcomes, the results relating to engagement and perceived usefulness of the games appear weaker. One possible reason for this is that the games were designed with a strong emphasis on instructional clarity and task efficiency, which optimises usability and measurable learning gains. Another possible explanation could lie in user expectations. Learners may approach serious games with preconceived ideas formed through commercial entertainment games. If the experience does not meet these expectations in terms of immersion, interactivity or reward systems, engagement ratings may suffer, even if the game performs well in terms of its educational purpose. Consequently, elements such as interface clarity, task structure, and feedback mechanisms tend to be optimised, which has a positive influence on both learning outcomes and the overall user experience. In contrast, engagement and perceived usefulness depend more heavily on subjective factors, such as individual preferences, user expectations, and familiarity with digital, game-based environments. The results show that high levels of engagement do not automatically arise from the use of game-based approaches. Furthermore, a strong positive correlation was found between engagement and the appeal of the serious games developed, meaning that user engagement increases as the serious game is perceived as more appealing. Furthermore, engagement often requires more sophisticated, immersive game mechanics, which are not always prioritised in educational contexts where instructional effectiveness remains the primary goal. Similarly, perceived usefulness may be less immediate if users do not clearly recognise how they can apply the knowledge they have acquired to real-world or practical contexts. From a theoretical perspective, these findings suggest a potential misalignment between the priorities of instructional design and the motivational dimensions of serious games. Specifically, they support the idea that optimising usability and learning effectiveness does not automatically lead to greater engagement or perceived usefulness. This emphasises the importance of integrating motivational and engagement design frameworks with didactic approaches more effectively in the development of serious games.

The study also included the development of an initial questionnaire containing items related to students' attitudes towards Mathematics, the same items of the final questionnaire where Mathematics learning was contextualised within the serious game. This was done to examine whether engaging in subject-related activities within a game-based context could lead to improvements in attitudes towards Mathematics. The results suggest that carrying out didactical activities in a game-based environment can significantly reduce students' anxiety towards Mathematics.

A theoretical DGBL model was developed, based on the ADDIE (Analysis, Design, Development, Implementation, Evaluation) instructional design framework, and refined iteratively through empirical feedback. The model emphasises modularity, flexibility, and technological independence, enabling adaptation to different disciplines, contexts, and educational levels. This study found that DGBL activities can be effectively implemented by both teachers and students when integrated into digital learning platforms such as Moodle due to their accessibility, ease of use and alignment with existing digital infrastructures. This reduces technological and cognitive barriers, enabling educators without advanced technical or programming skills to design, implement and monitor game-based activities and allowing students to seamlessly access the content within their familiar learning environment. The study also shows that GenAI tools can significantly support and facilitate the development of serious games. GenAI can rapidly generate content, scenarios, narratives, thereby reducing the time and effort required for instructional design. Further studies are needed to understand the impact of GenAI on students in game-based context and how it can be used to create highly personalised game-based learning pathways. Further research is needed to improve personalisation in digital, game-based learning pathways and boost student engagement. Nevertheless, this model provides a robust, evidence-based framework for integrating DGBL in upper secondary and higher education, showing how technological and didactical elements can be effectively aligned to create meaningful and impactful learning experiences. It provides a structured framework to guide future research and the design and application of serious games in Mathematics education.

8.2 Limitations and Future Developments

The proposed model was developed and tested within the Italian national context, specifically in the field of Mathematics. The sample involved in the study primarily consisted of students enrolled in non-Mathematics degree programmes, as well as secondary school students. However, the model is not restricted to a single educational level since one of its key strengths is its flexibility and adaptability in designing digital, game-based activities. It can indeed be applied to a wide range of educational settings by adjusting key elements appropriately across its various phases, particularly during target analysis and design phases. For instance, teachers and instructional designers can use the model to create materials of different levels of complexity. They can also adapt and contextualise content flexibly across different subject areas to align with specific learning objectives. In light of these considerations, future research should aim to test the model in a broader range of educational contexts and disciplines. Such efforts would contribute to validating its applicability, robustness, and effectiveness more broadly. This approach paves the way for further research in other fields and offers the potential to extend the investigation to a wide range of subjects. The aim is to share and implement the two developed serious games across various digital learning environments (DLEs) and degree programmes. The overall model will be shared more widely as an open educational resource and experimented with in other contexts, not only at the University of Turin but also within the UNITA Alliance (Corino et al., 2025). The intention is also to share these resources with secondary schools, such as those in the PP&S network (Fissore et al., 2024a), where the two serious games have already been shared among teachers. From a research and educational perspective, the project would have a significant impact by supporting teaching and learning through innovative, engaging and meaningful digital resources. While these two games represent only an initial step, the initiative could be expanded to include a wider range of game-based learning activities. This would result in the creation of a Mathematics and other disciplines repository, transforming resources developed in isolation into adaptable solutions that can easily be adopted in different classes. This would make these resources more reusable and adaptable in different educational settings.

The results regarding the extent to which serious games support personalised learning were not satisfactory. Overall, the findings suggest that the ability to the serious games to adapt to individual learning needs was limited. Therefore, future studies should focus on developing targeted gamification strategies based on personal student characteristics, such as initial motivation or different types of gamification users (Dumas Reyssier et al., 2023). Investigating different gamification user types could be interesting. Understanding how users with different motivations interact with and respond to gamified systems could provide valuable insights into engagement, enjoyment and learning outcomes. Analysing these user types would make it possible to identify which game mechanics are more effective for specific groups, enabling the design of more adaptive, inclusive and user-centred serious games (Diamond et al., 2015).

As the results regarding the impact of serious games on engagement were not particularly compelling, future research will focus on conducting more targeted studies in this area.

A third experiment involving the two serious games is taking place during the 2025-26 academic year. Based on the data collected from the two previous experiments and the current one, it is intended to assess whether they confirm, extend, or enrich the findings presented in this thesis and to enable refinement of the model. In this study, items from various validated questionnaires in the literature were used in the initial and final questionnaires to compare students' attitudes towards Mathematics before and after playing the serious game. Exploratory factor analyses were not conducted as the study's primary aim was to develop a broader tool for evaluating serious games that could be extended and applied across other subject areas, rather than being strictly tied to Mathematics. Therefore, questions specifically related to attitudes towards Mathematics were inconsistent with this purpose. However, such analyses could be performed on the initial questionnaire in the future to explore, for example, whether additional factors emerge in the questionnaire. Another direction for future research could be to explore the interconnected causal relationships between the factors in the final questionnaire. Examining the interactions between these factors would reveal how one factor influences or is influenced by another (Cardona et al., 2023; Chauhan et al, 2021). The readability and clarity of the content were among the most appreciated aspects of both serious games. These features paved the way for investigating the accessibility of the games, particularly as potential support tools for students with specific learning difficulties. Future research could explore how such tools can support learners' specific needs, for example those with learning disorders, and contribute to more inclusive learning (Vidakis et al., 2023). The evidence of this study shows that the DGBL_M item "*To what extent do you think that the use of digital game-based learning in Mathematics can reduce the gender gap in scientific disciplines?*", achieved the highest mean and median scores. Furthermore, female participants generally evaluated serious games and their use in Mathematics learning more positively. Further investigation could therefore explore whether DGBL could help to reduce gender gaps in Mathematics by fostering interest and engagement in the subject among female students.

One aspect that did not yield particularly conclusive results was the relationship between the impact of serious games on learning outcomes and the evaluations of the tools provided by students. This relationship was unclear in several existing studies (Cruz et al., 2023; Christopoulos et al., 2024). The present experiments did not clarify it either, so it would be interesting to explore this line of research further in future.

The Alien game received moderately positive evaluations across all item groups. One possible reason for this is that, in contrast to the first game, it cannot be completed by relying solely on prior knowledge of the topics covered. While the mathematical concepts in the first game (functions of a single variable) are widely taught in the final years of secondary school, the topics included in the second game are not usually studied beforehand. It could be tested again to assess its reception and identify areas for improvement.

One limitation of the study is that a pre and post-test comparison could not be conducted. Only the post-test exam results were analysed. When available, these results were occasionally compared with the students' upper secondary school grades to provide insight into their Mathematics background. Therefore, future studies will aim to conduct a targeted pre-post comparison to obtain a direct measure of learning outcomes. Furthermore, it was not possible to determine whether serious games were better received by students with stronger or weaker mathematical abilities. Future research will analyse the relationship between students' mathematical abilities and their appreciation of serious games. One approach could be to start from the results of a pre-test. Another study planned for future research aims to investigate the relationship between students' attitudes towards Mathematics and their appreciation of serious games. The goal is to understand whether learners who are more motivated and confident in Mathematics value serious games more, or whether this group may perceive such tools as redundant, while they may be more useful and appreciated by students who are less motivated or confident in the subject.

9. Conclusion

This study explores digital game-based methods to learn and teach Mathematics. A systematic literature review was conducted to examine the current state of research on game-based learning (GBL) and its potential. The review revealed a lack of research on this topic, particularly at higher educational levels. While the results reported in the literature are mixed, most identified studies indicate positive effects of GBL on learning outcomes. Subsequently, digital game-based activities were developed and experimented to support Mathematics learning. Two serious games were designed as educational tools that translate key mathematical concepts into interactive scenarios. The serious games cover topics taught in the classroom and encourage students to apply their knowledge into practice. One is related to functions in a real variable, while the other covers more specific topics related to probability and statistics. The game-based activities were experimented with approximately 360 university students enrolled in non-mathematics degree programmes, and 120 upper secondary school students in the final or penultimate year of high school during 2023-24 and 2024-25 academic years.
The research questions that guided this study are:
(RQ1) How can the effectiveness and impact of the developed serious games be evaluated, both in terms of technical and game-related aspects (such as usability, game appeal, engagement, and perceived usefulness) and discipline-related dimensions (such as motivation, attitude, and math anxiety)?
(RQ2) Do serious games designed improve students' approach to, and learning outcomes in, Mathematics?
(RQ3) How can game-based activities be designed to be effectively employed by teachers in their instructional practices and by students in their learning process?
(RQ4) What advantages can new GenAI tools offer in terms of developing and using serious games?

The study resulted in the creation of a final, reliable questionnaire that measures the following dimensions:

1) The impact of digital game-based learning on the Mathematics learning experience.
2) Perceived effectiveness for learning Mathematics.
3) The complexity of the game and possible difficulties.
4) Game experience and serious game appeal.
5) Engagement, in terms of students' enjoyment and involvement, negative feelings, pressure, competence and goal achievement. There is also an extended version of the engagement scale that measures also flow and immersion while playing the game.

Considering the results obtained, it emerged that serious games were generally perceived quite positively in both higher and secondary education, with scores around or above the midpoint of the scale. In the second experimental phase of the "The Murderer at the High-Tech Institute" (Murderer game), more positive evaluations and improved performance were observed among university and upper secondary school students than in the pilot study. These improvements are expected to be due to the revision and refinement of the game. Based on the descriptive statistics of the final questionnaire, it was observed that students particularly valued the usability of the serious games, considering them a positive learning tool for Mathematics. However, some students had differing opinions about their perceived usefulness, possibly due to a lack of prior exposure to this approach in earlier school years. As expected, engagement and game appeal received the lowest ratings, possibly because today's students are accustomed to highly polished, sophisticated commercial games. The low levels of engagement observed in the studies may have had a negative impact on the gaming experience and the serious game's overall appeal. This result suggests that there is still room for improvement, particularly regarding engagement. It is important to note that the serious games were designed to be highly challenging, with a high threshold for success. The performance scale ranges from 0% to 100%. Only scores between 75% and 100% are considered winning. In most cases, the correlations between responses to the final questionnaire and the gameplay data revealed no significant relationships. The only notable finding was that students who achieved higher scores in the serious game tended to perceive it as slightly more usable for university students and slightly more engaging and useful for secondary school students.
Differences emerged across degree programmes. Students on healthcare programmes were found to appreciate the Murderer game the least, scoring up to 10 points lower than students on other programmes. Conversely, the game was well received by students of Strategic and Security Sciences, and Biotechnology. Moreover, female students evaluated the game more positively than male students. This pattern may be partly explained by the fact that male participants are used to play video games more frequently and therefore may have adopted

a more critical perspective. In general, however, no substantial differences were observed in the responses based on gameplay frequency, except for usability. Students who reported playing video games less frequently tended to rate usability more highly.

Even though it was tested only with 88 students from the biotechnology degree programme, the Alien game, which covered probability and statistics topics, received moderately positive evaluations respect to the other serious game. It also emerged that this serious game was used as a tool for learning new concepts. Overall, the students' comments suggest that they saw both games as an opportunity to assess their level of understanding and apply what they had learnt in class.

One of the main findings of the study concerns learning outcomes. Exam results were consistently higher for all experimental groups in the biotechnology and healthcare fields who used one or both serious games compared to the control group, who primarily followed a traditional approach and did not use these tools. Students who performed better in the serious game also tended to perform slightly better in the exam. However, no statistically significant correlation was found between exam results and the number of attempts made at the serious game. Therefore, it cannot be concluded that more attempts led to better exam performance. The correlation found between exam scores and responses to the final questionnaire is mixed. Therefore, it cannot be concluded that students who provided more positive responses to the final questionnaire performed better in the exam, or that students with higher or lower levels of mathematical achievement appreciated the serious game more.

Both studies observed across all degree programmes a substantial reduction in math anxiety between the initial and final questionnaires, which were completed before and after playing the game, respectively. This suggests that students' math anxiety decreases when Mathematics is taught in a game-based setting rather than in a traditional learning context. Among biotechnology students, a reduction in extrinsic motivation was also observed. It is notable that these results concern courses in which Mathematics was considered a secondary subject.

The readability and clarity of the contents were among the most appreciated aspects of both serious games. These features were also investigated to assess the games' accessibility, particularly as potential support tools for students with specific learning difficulties. Students also valued the tools as an opportunity to develop their problem-solving and critical thinking skills. This was particularly evident in the Alien game.

The study was conducted during the rapid expansion of GenAI tools, and the second experimentation was therefore adapted to incorporate these new technologies. The widespread adoption of GenAI in many different areas has raised questions about its potential application in digital game-based learning. From a methodological perspective, GenAI was introduced for two main purposes. Firstly, GenAI tools were used during the design phase of the serious game "Don't open that spacecraft" (the Alien game) and they were proved useful for quickly developing digital, game-based activities. GenAI tools were also made available to students to support them in solving the game and it emerged that students used GenAI-based tools to help them clarify and revise content while engaging with the serious games.

One of the main outcomes of the research is the identification of a model to be made available as an open educational resource for designing digital game-based activities in a DLE. The proposed is based on the ADDIE instructional design framework (Peterson, 2003) and provides a structured yet flexible way of integrating gamification and digital game-based learning in a DLE. Its five-phase structure - Analysis, Design, Development, Implementation and Evaluation - ensures alignment between learning objectives, game mechanics and technological implementation, while supporting iterative refinement based on student feedback and learning outcomes. The model's flexibility means it can be adapted for use in different disciplines and educational contexts. Throughout its various phases, the model gives teachers a high degree of freedom and autonomy. During the Analysis phase, teachers can select the topics and learning objectives, thereby enabling the model to be applied across different disciplines and contents. In the Design phase, they determine which game-based components to adopt and how to implement them, tailoring their choices to the specific educational context. As the model is technology-independent, teachers have flexibility in choosing the tools for develop digital game-based activities. The model's flexibility makes it scalable and sustainable. It is designed to be reusable and adaptable, allowing it to be continuously refined. Its phases and associated design templates can be applied to different courses and disciplines, and its modular structure means that updates and modifications can be made without the need for a complete redesign.

The model incorporates an evaluation phase. To this end, a questionnaire was designed to evaluate the effectiveness of serious games for learning Mathematics. Although an experimentation would be necessary, this tool can also be adapted and applied to other subject areas. As the literature review highlights, it is crucial that teachers receive adequate training in designing and adopting game-based learning resources. Although the model is designed to be accessible and relatively easy to use without advanced technical skills, a minimum level of training is still necessary to ensure effective use and provide students with quality, innovative learning tools.

References

1. Afari, E., Aldridge, J. M., Fraser, B. J., & Khine, M. S. (2013). Students' perceptions of the learning environment and attitudes in game-based mathematics classrooms. Learning Environments Research, 16(1), 131-150. https://doi.org/10.1007/s10984-012-9122-6
2. Al Fatta, H., Maksom, Z., & Zakaria, M. H. (2018). Game-based learning and gamification: Searching for definitions. *International Journal of Simulation: Systems, Science and Technology, 19*(6), 41.1–41.6. https://doi.org/10.5013/IJSSST.a.19.06.41
3. Albano, G., Arzarello, F., & Dello Iacono, U. (2020). Digital inquiry through games. Technology, Knowledge and Learning, 26(3), 577-595. https://doi.org/10.1007/s10758-020-09459-1
4. Algayres, M. G., Timcenko, O., & Triantafyllou, E. (2022). Development and use of a Playful Learning Observation Tool (PLOT) for active game-based learning in physical classroom situations. In *Proceedings of the 16th European Conference on Games Based Learning (ECGBL 2022)* (pp. 28–37). Academic Conferences International. https://doi.org/10.34190/ecgbl.16.1.713
5. Aragonez, T., Saur-Amaral, I., & Gouveia, M. (2021). Game-based learning in higher education: A systematic literature review. In *EDULEARN21 proceedings* (pp. 1849–1856). IATED. https://doi.org/10.21125/edulearn.2021.0432
6. Ardani, R. A., Suwarno, M., Rezky, R., & Utari, T. (2024). Game-based edutainment media: How practical is it as a medium for learning mathematics? In *AIP Conference Proceedings* (Vol. 3046, No. 1, p. 020031). AIP Publishing LLC. https://doi.org/10.1063/5.0194544
7. Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science*, *11*(5), 181-185. https://doi.org/10.1111/1467-8721.00196
8. Ayyasy, H. Y., & Asrul, A. (2024). The effect of game-based learning with math bingo in reducing math anxiety to improve students' Mathematics learning outcomes in data presentation materials. *Desimal: Jurnal Matematika, 7*(2), 383-394. https://doi.org/10.24042/djm.v7i2.23546
9. Bakri, S. R. A., Liew, C. Y., Chen, C. K., Tuh, M. H., & Ling, S. C. (2020). Bridging the gap between the derivatives and graph sketching in calculus: An innovative game-based learning approach. *Asian Journal of University Education, 16*(4), 121-136. https://doi.org/10.24191/ajue.v16i4.11962
10. Bangor, A., Kortum, P., & Miller, J. (2009) Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, 4(3), 114-123.
11. Barana, A., & Marchisio, M. (2021). *Analyzing interactions in automatic formative assessment activities for mathematics in digital learning environments*. In *Proceedings of the 13th International Conference on Computer Supported Education (CSEDU 2021)* (Vol. 1, pp. 497–504). SciTePress. https://doi.org/10.5220/0010474004970504
12. Barana, A., Conte, A., Fradiante, V., Marchisio Conte, M., & Rabellino, R. (2025). Serious game for higher mathematics education: Evaluation of the learning tool. In *Proceedings of the 2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 2367–2376). IEEE. https://doi.org/10.1109/COMPSAC65507.2025.00333
13. Barana, A., Marchisio, M., & Roman, F. (2023). Fostering problem solving and critical thinking in mathematics through generative artificial intelligence. *In Proceedings of the 20th International Conference on Cognition and Exploratory Learning in the Digital Age* (CELDA 2023) (pp. 377-385). IADIS Press.
14. Barana, A., Marchisio, M., Sacchet, M., & Salusso, D. (2022). *Teaching online EMI mathematics courses: A proposal to combine gamification and adaptive learning.* In N. Adamuz-Povedano, E. Fernández-Ahumada, & C. A. Huertas-Abril (Eds.), *Handbook of Research on International Approaches and Practices for Gamifying Mathematics* (pp. 304–324). IGI Global. https://doi.org/10.4018/978-1-7998-9660-9.ch015
15. Barbieri, G. G., Barbieri, R., & Capone, R. (2021). Serious games in high school mathematics lessons: An embedded case study in Europe. *Eurasia journal of Mathematics, Science and Technology Education*, *17*(5), em1963. https://doi.org/10.29333/ejmste/10857

16. Barboros, Ş., Saygı, E., & Akar, Ş. Ş. (2024). Games in mathematics education: Systematic literature review. *Journal for the Mathematics Education and Teaching Practices, 5(2)*, 45-56. https://doi.org/10.5281/zenodo.14599876
17. Becke, K. (2021). What's the difference between gamification, serious games, educational games, and game-based learning? *Academia Letters*, Article 209. https://doi.org/10.20935/AL209
18. Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger P., & Berta, R. (2013) Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction, 2013*, 136864. https://doi.org/10.1155/2013/136864
19. Bente, J., & Breuer, J. (2010). Why so serious? On the relation of serious games and learning. *Journal for Computer Game Culture*, *4(1)*, 7-24. https://doi.org/10.7557/23.6111
20. Bertram, L. (2020). Digital learning games for mathematics and computer science education: The need for preregistered RCTs, standardized methodology, and advanced technology. *Frontiers in Psychology, 11*, 02127. https://doi.org/10.3389/fpsyg.2020.02127
21. Bourgonjon, J., Valcke, M., Soetaert, R., & Schellens, T. (2010) Students' perceptions about the use of video games in the classroom. *Computers & Education, 54*(4) 1145-1156. https://doi.org/10.1016/j.compedu.2009.10.022
22. Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhart, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624-634. https://doi.org/10.1016/j.jesp.2009.02.016
23. Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
24. Bunchball. (2016). *Gamification 101: An introduction to game dynamics* [White paper]. https://www.bunchball.com
25. Casasso, F., Fissore, C., Fradiante, V., & Marchisio, M. (2022). *Developing computational thinking and problem-solving skills in early childhood and primary school*. In *EDULEARN22 Proceedings* (pp. 9316–9325). IATED. https://doi.org/10.21125/edulearn.2022.2245
26. Chatzea, V. E., Logothetis, I., Kalogiannakis, M., Rovithis, M., & Vidakis, N. (2024). Digital educational tools for undergraduate nursing education: A review of serious games, gamified applications and non-gamified virtual reality simulations/tools for nursing students. *Information, 15(7)*, 410. https://doi.org/10.3390/info15070410
27. Chauhan, S., Mittal, M., Woźniak, M., Gupta, S., & Pérez de Prado, R. (2021). A technology acceptance model-based analytics for online mobile games using machine learning techniques. *Symmetry, 13*(8), 1545. https://doi.org/10.3390/sym13081545
28. Chen, M. F., Chen, Y. C., Zuo, P. Y., & Hou, H. T. (2023). Design and evaluation of a remote synchronous gamified Mathematics teaching activity that integrates multi-representational scaffolding and a mind tool for gamified learning. *Education and Information Technologies*, *28*(10), 13207-13233. https://doi.org/10.1007/s10639-023-11708-6
29. Chen, M. F., Chen, Y. C., Zuo, P. Y., & Hou, H.T. (2023). Design and evaluation of a remote synchronous gamified mathematics teaching activity that integrates multi-representational scaffolding and a mind tool for gamified learning. *Education and Information Technologies, 28*, 13207–13233. https://doi.org/10.1007/s10639-023-11708-6
30. Chen, P. Y., Hwang, G. J., Yeh, S. Y., Chen, Y. T., Chen, T. W., & Chien, C. H. (2021). Three decades of game-based learning in science and mathematics education: An integrated bibliometric analysis and systematic review. *Journal of Computers in Education*, *9*(3), 455-476. https://doi.org/10.1007/s40692-022-00220-3
31. Chen, S., & Michael, D. (2005). *Serious games: Games that educate, train and inform.* Muska & Lipman/Premier-Trade.
32. Chen, S., & Michael, D. (2006). *Serious games: Games that educate, train, and inform*. Thomson Course Technology.

33. Cheng, M.-T., Chen, J.-H., Chu, S.-J., & Chen, S.-Y. (2015). The use of serious games in science education: A review of selected empirical research from 2002 to 2013. *Journal of Computers in Education, 2*, 353–375. https://doi.org/10.1007/s40692-015-0039-9
34. Chorianopoulos, K. & Giannakos, M., (2014). Design principles for serious video games in mathematics education: From theory to practice. *International Journal of Serious Games*, *1(3)*, 51-59. https://doi.org/10.17083/ijsg.v1i3.12
35. Chou, Y. K. (2014). *Actionable gamification: Beyond points, badges, and leaderboards.* Createspace Independent Publishing Platform.
36. Christopoulos, A., Mystakidis, S., Kurczaba, J., Laakso, M. J., & Stylios, C. (2024). Is immersion in 3D virtual games associated with mathematical ability improvement in game-based learning? *International Journal of Science and Mathematics Education, 22*(7), 1479-1499. https://doi.org/10.1007/s10763-023-10440-4
37. Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79-122.
38. Clarke, S., Arnab, S., Keegan, H., Morini, L., & Wood, O. (2016). *Escaped: Adapting live-action, interactive games to support higher education teaching and learning practices*. In A. De Gloria & R. Veltkamp (Eds.), *Games and Learning Alliance* (pp. 144–153). Springer. https://doi.org/10.1007/978-3-319-50182-6_13
39. Corino, E., Costa, M., Fiorentino, A., & Garbarino, S. (2025). Language alliances and language policy: A focus on UNITA-Universitas Montium. *European Journal of Language Policy*, *17*(1), 141-159.
40. Correia, R. R., Ferreira, J. D. S., Teixeira, S., & Teles, A. S. (2023). A serious game for students with math learning difficulties: Co-design-based development and user experience evaluation study. *Revista Novas Tecnologias na Educação, 21*(2), 286-296.
41. Cosgrove, J., & Cachia, R. (2025). *DigComp 3.0: European digital competence framework* (5th ed.; JRC144121). Publications Office of the European Union. https://doi.org/10.2760/0001149
42. Crnković, V. M., Traunkar, I., & Crnković, B. (2022). *Treasure hunt as a method of learning mathematics*. In ECGBL 2022: 16th European Conference on Game-Based Learning (pp. 349–357). Academic Conferences and Publishing International. https://doi.org/10.34190/ecgbl.16.1.740
43. Crocco, F., Offenholley, K., & Hernandez, C. (2016). A proof-of-concept study of game-based learning in higher education. *Simulation & Gaming, 47*(4), 403-422. https://doi.org/10.1177/1046878116632484
44. Cruz, J. T., Coluci, V. R., & Moraes, R. (2023). ORUN-VR2: A VR serious game on projectile kinematics: Design, evaluation, and learning outcomes. *Virtual Reality*, *27*(3), 2583-2604.
45. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008
46. De Carvalho, C. V., & Coelho, A. (Eds.). (2022). *Game-Based Learning, Gamification in Education and Serious Games*. Basel, Switzerland: MDPI AG. https://doi.org/10.3390/books978-3-0365-3809-9
47. De Freitas, S. (2006). *Learning in immersive worlds: A review of game-based learning*. Joint Information Systems Committee (JISC).
48. Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). Gamification: Toward a definition. In *Proceedings of the CHI 2011 Gamification Workshop* (pp. 12–15).
49. Di Blasi, M., De Angelis, I., & Postiglione, A. (2022). *The usage of Kahoot! during activities for schools on physics*. *Il Nuovo Cimento C,* Article 220. https://doi.org/10.1393/ncc/i2022-22220-7
50. Diamond, L., Tondello, G. F., Marczewski, A., Nacke, L. E., & Tscheligi, M. (2015). The HEXAD gamification user types questionnaire: Background and development process. In *Proceedings of the Workshop on Personalization in Serious and Persuasive Games and Gamified Interactions* (CHI PLAY 2015). ACM.
51. Ding, Y. (2016). *Measuring developmental students' mathematics anxiety. Research and Teaching in Developmental Education, 33*(1), 31-47.
52. Dondio, P., Gusev, V., & Rocha, M. (2023). Do games reduce maths anxiety? A meta-analysis. *Computers & Education, 194*, 104650. https://doi.org/10.1016/j.compedu.2022.104650

53. Dowker, A., Sarkar, A., & Looi, C. Y. (2016). Mathematics anxiety: What have we learned in 60 years? *Frontiers in Psychology*, *7*, 508. https://doi.org/10.3389/fpsyg.2016.00508
54. Dumas Reyssier, S., Serna, A., Hallifax, S., Marty, J.-C., Simonian, S., & Lavoué, E. (2023). *How does adaptive gamification impact different types of student motivation over time? Interactive Learning Environments, 32*(10), 6043–6062. https://doi.org/10.1080/10494820.2023.2248220
55. Duggal, K., Gupta, L., & Singh, P. (2021). Gamification and machine learning inspired approach for classroom engagement and learning. *Mathematical Problems in Engineering*, 2021, Article 9922775, 1–18. https://doi.org/10.1155/2021/9922775
56. Estrada-Molina, O., Fuentes-Cancell, D. R., & García-Hernández, A. (2022). Evaluating usability in educational technology: A systematic review from the teaching of mathematics. LUMAT: *International Journal on Math, Science and Technology Education, 10*(1), 65-88. https://doi.org/10.31129/LUMAT.10.1.1686
57. European Commission, Directorate-General for Education, Youth, Sport and Culture. (2019). *Key competences for lifelong learning*. Publications Office of the European Union. https://doi.org/10.2766/569540
58. Fadda, D., Pellegrini, M., Vivanet, G., & Zandonella Callegher, C. (2021). Effects of digital games on student motivation in mathematics: A meta-analysis in K-12. *Journal of Computer Assisted Learning*, *38*(1), 304-325. https://doi.org/10.1111/jcal.12632
59. Fahlgren, M., Barana, A., Brunström, M., Marchisio Conte, M., Roman, F., Sacchet, M., Vinerean, M., & Wondmagegne, Y. (2024). *Example-generation tasks for computer-aided assessment in university mathematics education: Insights from a study conducted in two educational contexts*. International Journal of Research in Undergraduate Mathematics Education, *10,* 732-753. https://doi.org/10.1007/s40753-024-00252-4
60. Fante, C., Ravicchio, F., & Manganello, F. (2024). Navigating the evolution of game-based educational approaches in secondary STEM education: A decade of innovations and challenges. *Education Sciences, 14*(6), 662. https://doi.org/10.3390/educsci14060662
61. Filatro, A., & Cavalcanti, C. C. (2016). Structural and content gamification design for tutor education. In *eLearn: World Conference on EdTech* (pp. 1152-1157). Association for the Advancement of Computing in Education (AACE).
62. Fiorese, M., Macrì, A., & Mandorin, C. (2019). Un editor per creare serious game. In *Atti del MoodleMoot Italia 2019* (pp. 105-112). ISBN 978-88-907493-5-3
63. Fissore, C., Floris, F., Marchisio, M., & Rabellino, S. (2021). *Digital learning environment embedded in a larger nationwide one: A case study of a high school*. In *Lessons from a pandemic for the future of education – EDEN Conference 2021* (pp. 124–134). European Distance and E-learning Network. https://doi.org/10.38069/edenconf-2021-ac0012
64. Fissore, C., Fradiante, V., Marchisio, M., & Pardini, C. (2023a). Design didactic activities using gamification: The perspective of teachers. In M. B. Nunes, P. Isaías, T. Issa, & T. Issa (Eds.), *Proceedings of the 17th E-Learning and Digital Learning Conference 2023 (EDL2023)* (pp. 11–18). IADIS Press. https://doi.org/10.33965/EL_STE_2023 (Outstanding Paper Award).
65. Fissore, C., Fradiante, V., Marchisio, M., & Pardini, C. (2023b). Teachers' strategies and difficulties in designing gamification activities. *IADIS International Journal*, *21*(2), 86–100.
66. Fissore, C., Floris, F., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (2024a). From theory to training: Exploring teachers' attitudes towards artificial intelligence in education. In *Proceedings of the 16th International Conference on Computer Supported Education* (Vol. 2, pp. 118–127). SCITEPRESS–Science and Technology Publications. https://dx.doi.org/10.5220/0012734700003693
67. Fissore, C., Floris, F., Fradiante, V., Marchisio Conte, M., & Sacchet, M. (2024b). Involving teachers in gamified learning activities using generative artificial intelligence tools. In A. Schönbohm, F. Bellotti, A. Bucchiarone, F. de Rosa, M. Ninaus, A. Wang, V. Wanick, & P. Dondio (Eds.), *Games and Learning Alliance: 13th International Conference, GALA 2024, Berlin, Germany, November 20–22, 2024, Proceedings* (pp. 36–46). Springer. https://dx.doi.org/10.1007/978-3-031-78269-5_4
68. Floris, F., Fradiante, V., Marchisio Conte, M., & Rabellino, S. (2023). Design gamification strategies in a digital learning environment: The impact on students. In P. Dondio, M. Rocha, A. Brennan, A.

Schönbohm, F. de Rosa, A. Koskinen, & F. Bellotti (Eds.), *Games and Learning Alliance: 12th International Conference, GALA 2023, Dublin, Ireland, November 29–December 1, 2023, Proceedings* (pp. 464–469). Springer. https://doi.org/10.1007/978-3-031-49065-1_50

69. Floris, F., Fradiante, V., Marchisio Conte, M., & Rabellino, S. (2024). Strategie di gamification con Moodle per lo sviluppo di competenze di problem solving e per uno sviluppo sostenibile. *Rivista Bricks*, *2*, 25–34.
70. García-López, I. M., Acosta-Gonzaga, E., & Ruiz-Ledesma, E. F. (2023). Investigating the impact of gamification on student motivation, engagement, and performance. *Education Sciences, 13*(8), 813. https://doi.org/10.3390/educsci13080813
71. Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation and learning. *Simulation & Gaming: An Interdisciplinary Journal of Theory, Practice and Research,33*(4), 441-467.
72. Gentile, M., La Guardia, D., Dal Grande, V., Ottaviano, S., & Allegra, M. (2014). An agent-based approach to design serious game. *International Journal of Serious Games*, *1*(2). Article 3.
73. Gil-Doménech, D., & Berbegal-Mirabent, J. (2017). Stimulating students' engagement in mathematics courses in non-STEM academic programmes: A game-based learning. I*nnovations in Education and Teaching International, 56*(1), 57-65. https://doi.org/10.1080/14703297.2017.1241471
74. González-Fernández, A., Revuelta-Domínguez, F. I., & Fernández-Sánchez, M. R. (2022). Models of instructional design in gamification: A systematic review of the literature. *Education Sciences*, *12*(1), 44. https://doi.org/10.3390/educsci12010044
75. Gounaridou, A., Siamtanidou, E., & Dimoulas, C. (2021). A serious game for mediated education on traffic behavior and safety awareness. *Educational Sciences, 11*(3), 127. https://doi.org/10.3390/educsci11030127
76. Gris, G., & Bengtson, C (2021). Assessment measures in game-based learning research: A systematic review. *International Journal of Serious Games*, *8*(1). https://doi.org/10.17083/ijsg.v8i1.383
77. Gusti, V. Y. K. (2025). Teaching algebra through games: A literature review. *Al-Irsyad Journal of Mathematics Education, 4*(2), 209–225. https://doi.org/10.58917/ijme.v4i2.177
78. Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education*, *80*, 152-161. https://doi.org/10.1016/j.compedu.2014.08.019
79. Harrington, I., & Mellors, M. J. (2021). Utilising online gamification to promote student success and retention in tertiary settings. *International Journal of Higher Education, 10*(7), 45-52. https://doi.org/10.5430/ijhe.v10n7p45
80. Hernández, A. F. V., Padilla, D. S. B., & Newball, A. A. N. (2024). Videogames and open feedback systems to enhance probabilistic reasoning and engagement. *International Journal of Serious Games*, *11*(1), 3-24.
81. Holmes, W., Persson, J., Chounta, I. A., Wasson, B., & Dimitrova, V. (2022). *Artificial intelligence and education: A critical view through the lens of human rights, democracy and the rule of law.* Council of Europe.
82. Huang, W., Johnson, T. E., & Han, S. (2013). Impact of online instructional game features on college students' perceived motivational support and cognitive investment: A structural equation modeling study. *The Internet and Higher Education*, *17*, 58-68. https://doi.org/10.1016/j.iheduc.2012.11.004
83. Hunicke, R., LeBlanc, M., & Zubek, R. (2004). *MDA: A formal approach to game design and game research. In Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, No.1). http://www.cs.northwestern.edu/~hunicke/MDA.pdf
84. Hussein, M. H., Ow, S. H., Elaish, M. M., & Jensen, E. O. (2022). Digital game-based learning in K–12 mathematics education: A systematic literature review. *Education and Information Technologies, 27*(2), 2859–2891. https://doi.org/10.1007/s10639-021-10721-x
85. IJsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2013). *The Game Experience Questionnaire.* Technische Universiteit Eindhoven
86. Ilić, J., Ivanović, M., & Klašnja-Milićević, A. (2024). Effects of digital game-based learning in STEM education on students' motivation: A systematic literature review. *Journal of Baltic Science Education, 23*(1), 20-36. https://doi.org/10.33225/jbse/24.23.20

87. Jaipal-Jamani, K., & Figg, C. (2018). *Application of gamification to blended learning in higher education*. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology* (4th ed., pp. 3238–3247). IGI Global. https://doi.org/10.4018/978-1-5225-2255-3.ch282
88. Jiménez-Hernández, E. M., Jiménez-Murillo, J. A., Segura-Castruita, M. A., & González-Leal, I. (2021). Using a serious video game to support the learning of tree traversals. In *Proceedings of the 2021 9th International Conference in Software Engineering Research and Innovation* (CONISOFT) (pp. 238-244). IEEE.
89. Kapp, K. M. (2012). *The gamification of learning and instruction: Game-based methods and strategies for training and education*. John Wiley & Sons.
90. Kapp, K. M., Blair, L., & Mesch, R. (2014). *The gamification of learning and instruction fieldbook: Ideas into practice*. John Wiley & Sons.
91. Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & education, 55*(2), 427-443. https://doi.org/10.1016/j.compedu.2010.02.007
92. Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Springer. https://doi.org/10.1007/978-1-4419-1250-3
93. Kim, S., Song, K., Lockee, B., & Burton, J. (2018). *Gamification in learning and education: Enjoy learning like gaming.* Springer International Publishing. https://doi.org/10.1007/978-3-319-47283-6
94. Krath, J., Schürmann, L., & von Korflesch, H. F. (2021). Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Computers in Human Behavior, 125*, 106963. https://doi.org/10.1016/j.chb.2021.106963
95. Lewis, J. R. (2018). The System Usability Scale: Past, present, and future. *International Journal of Human–Computer Interaction, 34*(7), 577–590. https://doi.org/10.1080/10447318.2018.1455307
96. Lee, J. Y., Pyon, C. U., & Woo, J. (2023). Digital twin for math education: A study on the utilization of games and gamification for university mathematics education. *Electronics, 12*(15), 3207. https://doi.org/10.3390/electronics12153207
97. Li, L., Hew, K. F., & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, has minimal impact on competence: A meta-analysis and systematic review. *Educational Technology Research and Development*, *72*, 765-796. https://doi.org/10.1007/s11423-023-10337-7
98. Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *The International Journal of Management Education, 21*(2), 100790. https://doi.org/10.1016/j.ijme.2023.100790
99. Magat Jr, R. B. (2023). Gamification of statistics and probability education: A mobile courseware approach. *Journal of Information Technology Education: Innovations in Practice, 22*, 159-177. https://doi.org/10.28945/5179
100. Maisey, J., Thoma, G., Moeller, K., Kiili, K., & Ninaus, M. (2022). Effects of a game-based fraction estimation task on math anxiety. In K. Kiili, K. Antti, F. de Rosa, M. Dindar, M. Kickmeier-Rust, & F. Bellotti (Eds.), *Games and Learning Alliance: 11th International Conference, GALA 2022, Proceedings* (pp. 137–146). Springer. https://doi.org/10.1007/978-3-031-22124-8_14
101. Malaquias, R. F., Malaquias, F. F., & Hwang, Y. (2018). Understanding technology acceptance features in learning through a serious game. *Computers in Human Behavior, 87*, 395-402. https://doi.org/10.1016/j.chb.2018.06.008
102. Malvasi, V., Gil-Quintana, J., & Bocciolesi, E. (2022). The projection of gamification and serious games in the learning of mathematics: A multi-case study of secondary schools in Italy. *Mathematics, 10*(3), 336. https://doi.org/10.3390/math10030336
103. Marchisio, M., Remogna, S., Roman, F., & Sacchet, M. (2020). *Teaching mathematics in scientific bachelor degrees using a blended approach*. In 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC) (pp. 190–195). IEEE. https://doi.org/10.1109/COMPSAC48688.2020.00034

104. Marchisio, M., Remogna, S., Roman, F., & Sacchet, M. (2022). Teaching mathematics to non-mathematics majors through problem solving and new technologies. *Education Sciences, 12*(1), 34. https://doi.org/10.3390/educsci12010034
105. Marchisio, M., Roman, F., & Sacchet, M. (2021). Basic mathematical modelling competencies for non-STEM higher education. *In Proceedings of the 15th International Conference on e-Learning* (pp. 54–61). IADIS Press.
106. Marques, M. M., & Pombo, L. (2019). Game-based mobile learning with augmented reality: Are teachers ready to adopt it? In *Project and Design Literacy as Cornerstones of Smart Education: Proceedings of the 4th International Conference on Smart Learning Ecosystems and Regional Development* (pp. 207-218). Springer.
107. Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370-396. https://doi.org/10.1037/h0054346
108. McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.
109. Ministero dell'Istruzione e del Merito. (2025). *Indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo di istruzione* (D.M. 30 giugno 2025, n. XXX). Gazzetta Ufficiale della Repubblica Italiana.
110. Ministero dell'Istruzione, dell'Università e della Ricerca. (2010a). *Schema di regolamento recante indicazioni nazionali riguardanti gli obiettivi specifici di apprendimento concernenti le attività e gli insegnamenti compresi nei piani degli studi previsti per i percorsi liceali di cui all'articolo 10, comma 3, del Decreto del Presidente della Repubblica 15 marzo 2010, n. 89, in relazione all'articolo 2, commi 1 e 3, del medesimo regolamento* (Decreto 7 ottobre 2010, n. 211). Gazzetta Ufficiale della Repubblica Italiana.
111. Ministero dell'Istruzione, dell'Università e della Ricerca. (2010b). *Linee guida per il passaggio al nuovo ordinamento degli istituti tecnici a norma dell'articolo 8, comma 3, del Decreto del Presidente della Repubblica 15 marzo 2010, n. 88*. Gazzetta Ufficiale della Repubblica Italiana.
112. Ministero dell'Istruzione, dell'Università e della Ricerca. (2018). *Regolamento recante la disciplina dei profili di uscita degli indirizzi di studio dei percorsi di istruzione professionale, ai sensi dell'articolo 3, comma 3, del Decreto Legislativo 13 aprile 2017, n. 61 (Decreto Ministeriale 24 maggio 2018, n. 92)*. Gazzetta Ufficiale della Repubblica Italiana, Suppl. Ord. n. 35, 27 luglio 2018.
113. Nautiyal, V. V., Silverio, S. A., & Salvador, E. E. P. (2024). Let's get on-board: A practical framework for designing and implementing educational board games in K-12 classrooms. *Frontiers in Education 9*, 1420515. https://doi.org/10.3389/feduc.2024.1420515
114. Norman, K. L. (2013). GEQ (game engagement/experience questionnaire): A review of two papers. *Interacting with computers*, *25*(4), 278-283. https://doi.org/10.1093/iwc/iwt009
115. Ntoa, S., Ntagianta, A., Flores, F., Kolek, L., Petrova, A., Apostolakis, K., Stamou, S., Margetis, G., & Stephanidis, C. (2024). Serious games beyond entertainment and learning: An evaluation methodology for assessing awareness raising, empathy, and social change. In A. Marcus, E. Rosenzweig, M. M. Soares, P. L. P. Rau, & A. Moallem (Eds.), *HCI International 2024 – Late Breaking Papers* (pp. 141-164). Springer. https://doi.org/10.1007/978-3-031-76821-7_11
116. Organisation for Economic Co-operation and Development. (2014). *PISA 2012 technical report.* OECD Publishing. https://doi.org/10.1787/6341a959-en
117. Organisation for Economic Co-operation and Development. (2024a). *PISA 2022 technical report*. OECD Publishing. https://doi.org/10.1787/01820d6d-en
118. Organisation for Economic Co-operation and Development. (2024b). *Do adults have the skills they need to thrive in a changing world? Survey of adult skills 2023*. OECD Publishing. https://doi.org/10.1787/b263dc5d-en
119. Organisation for Economic Co-operation and Development. (2024c). Explanatory memorandum on the updated OECD definition of an AI system. *OECD Artificial Intelligence Papers*, 8.
120. Pais, S., & Hall, A. (2024). When games get in the way: Gamification approach for teaching maths—A case Study with undergraduate tourism students. In *Proceedings of the European Conference on Games Based Learning (ECGBL 2024)* (pp. 685-694A). Academic Conferences International Limited.

121. Pais, S., Sousa, A. E., & Pires, A. (2023). *Using an escape room activity to enhance the motivation of undergraduate life science students in mathematics classes: A case study*. In Proceedings of the 17th European Conference on Games Based Learning (pp. 490–496). Academic Conferences and Publishing International. https://doi.org/10.34190/ecgbl.17.1.1431
122. Pan, Y., Ke, F., & Xu, X. (2022). A systematic review of the role of learning games in fostering mathematics education in K-12 settings. *Educational Research Review*, 36, 100448. https://doi.org/10.1016/j.edurev.2022.100448
123. Perini, S., Luglietti, R., Margoudi, M., Oliveira, M., & Taisch, M. (2018). Learning and motivational effects of digital game-based learning (DGBL) for manufacturing education: The life cycle assessment (LCA) game. *Computers in Industry, 102*, 40-49. https://doi.org/10.1016/j.compind.2018.08.005
124. Perrotta, C., Featherstone, G., Aston, H., & Houghton, E. (2013). *Game-based learning: Latest evidence and future directions*. National Foundation for Educational Research.
125. Peterson, C. (2003). Bringing ADDIE to Life: Instructional design at its best. *Journal of Educational Multimedia and Hypermedia, 12*(3), 227–241.
126. Piñero Charlo, J. C., Canto López, M. D. C., & Caballero Leiva, C. (2024). Treating trainee teachers' mathematical anxiety by using game-based learning: A case study. *Bolema: Boletim de Educação Matemática, 38*, e220218. https://doi.org/10.1590/1980-4415v38a220218
127. Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
128. Plass, J. L., Homer, B. D., & Kinzer, C. K. (2016). Foundations of game-based learning. *Educational Psychologist, 50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533
129. Plass, J. L., Mayer, R. E., & Homer, B. D. (2019). *Handbook of game-based learning*. Mit Press.
130. Poels, K., de Kort, Y. A. W., & IJsselsteijn, W. A. (2007). *D3.3: Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games* (Technical report).Technische Universiteit Eindhoven.
131. Prensky, M. (2003). Digital game-based learning. *Computers in entertainment (CIE)*, *1*(1), 21.
132. Qasim, S. H. (2024). *Beyond the classroom: Emerging technologies to enhance learning*. Book Bazooka Publication.
133. Qian, M., & and Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, *63,* 50-58. https://doi.org/10.1016/j.chb.2016.05.023
134. Ranganathan, P., Caduff, C., & Frampton, C. M. (2024). Designing and validating a research questionnaire—Part 2. *Perspectives in Clinical Research, 15*(1), 42-45. https://doi.org/10.4103/picr.picr_318_23
135. Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (EUR 28775 EN). Publications Office of the European Union. https://doi.org/10.2760/159770
136. Richardson, F. C., & Suinn, R. M. (1972). The mathematics anxiety rating scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551-554 https://doi.org/10.1037/h0033456
137. Romero, M., Frosig, T., Taylor-Beswick, A. M. L., Laru, J., & Bernasco, B. (2024). *Manifesto in defence of human-centred education in the age of artificial intelligence*. In A. Urmeneta & M. Romero (Eds.), *Creative Applications of Artificial Intelligence in Education* (pp. 157–178). Springer Nature. https://doi.org/10.1007/978-3-031-55272-4_12
138. Sánchez-Mena, A., & Martí-Parreño, J. (2017). Drivers and barriers to adopting gamification: Teachers' perspectives. *Electronic Journal of e-Learning*, *15*(5), 434-443.
139. Sánchez-Ruiz, L. M., López-Alfonso, S., Moll-López, S., Moraño-Fernández, J. A., & Vega-Fleitas, E. (2022). Educational digital escape rooms footprint on students' feelings: A case study within aerospace engineering. *Information, 13*(10), 478. https://doi.org/10.3390/info13100478
140. Sauro, J. (2011). *Measuring usability with the system usability scale (SUS).* MeasuringU. https://measuringu.com/sus/
141. St Omer, S. M., Evers, K., Wang, C. Y., & Chen, S. (2025). Technology-enhanced Mathematics learning: Review of the interactions between technological attributes and aspects of mathematics

education from 2013 to 2022. *Humanities and Social Sciences Communications, 12*(1), 1–13. https://doi.org/10.1057/s41599-025-05475
142. Sümer, M. and Aydın, C. H. (2022). Design principles for integrating gamification into distance learning programs in higher education: A mixed method study. *International Journal of Serious Games*, *9(2),* 79–91. https://doi.org/10.17083/ijsg.v9i2.494
143. Tcha-Tokey, K., Christmann, O., Loup-Escande, E., & Richir, S. (2016). Proposition and Validation of a questionnaire to measure the user experience in immersive virtual environments. *The International Journal of Virtual Reality*, *16*(1), 33-48. https://doi.org/10.20870/ijvr.2016.16.1.2880
144. Ting, F. S. T., Lam, W. H., & Shroff, R. H. (2019). Active learning via problem-based collaborative games in a large mathematics university course in Hong Kong. *Education Sciences, 9*(3), 172. https://doi.org/10.3390/educsci9030172
145. Tolentino, G. P., Battaglini, C., Pereira, A. C. V., de Oliveira, R. J., & de Paula, M. G. M. (2011). Usability of serious games for health. In *Proceedings of the 2011 Third International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES 2011)* (pp. 172–175). IEEE.
146. Tomczyk, D., & Teckchandani, A. (2023). Using gamification strategies to create effective experiential exercises. *Management Teaching Review*, *10*(1). https://doi.org/10.1177/23792981231190590
147. United Nations General Assembly. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development* (Resolution A/RES/70/1). https://undocs.org/A/RES/70/1
148. Valencia, D. C., & Duque, F. A. B. (2023). Technology acceptance model (TAM): A study of teachers' perception of the use of serious games in the higher education. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 18*(1), 123-129.
149. Verkijika, S. F., & De Wet, L. (2015). Using a brain-computer interface (BCI) in reducing math anxiety: Evidence from South Africa. *Computers & Education, 81*, 113-122. https://doi.org/10.1016/j.compedu.2014.10.002
150. Vidakis, N., Syntychakis, E., Kalafatis, K., Varhalamas, P., & Triantafyllidis, G. (2015). Concealing education into games. In *European Conference on Games Based Learning* (p. 554). Academic Conferences International Limited.
151. Vidakis, N., Barianos, A. K., Trampas, A.M., Papadakis, S., Kalogiannakis, M., & Vassilakis, K. (2019). Generating education in-game data: The case of an ancient theatre serious game. In *Proceedings of the 11th International Conference on Computer Supported Education (CSEDU 2019)* (Vol.1, pp. 36-43). SciTePress. https://doi.org/10.5220/0007810800360043
152. Vidakis, N., Christinaki, E., Syntychakis, E., & Triantafyllidis, G. (2023). Designing a General Open Authorable Digital Ecosystem for Educational Games to Support Special Learning Needs. In *Virtual Reality Games for Rehabilitation* (pp. 191-208). New York, NY: Springer New York. https://doi.org/10.1007/978-1-0716-3371-7_9
153. Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens* (EUR 31006 EN). Publications Office of the European Union. https://doi.org/10.2760/115376
154. Weir, K. (2023). How to solve for math anxiety? Studying the causes, consequences, and prevention methods needed. *Monitor on Psychology*, *54*(7), 44.
155. Wen, R., & Dubé, A. K. (2022). A systematic review of secondary students' attitudes towards mathematics and its relations with mathematics achievement. *Journal of Numerical Cognition, 8*(2), 295-325. https://doi.org/10.5964/jnc.7937
156. Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.
157. Wiboonsin, K., & Kasemsukpipat, W. (2022). Wealth Creation: Serious game development to improve financial literacy for high school students. In *Proceedings of the European Conference on Games Based Learning (ECGBL 2022)* (pp. 637-658). Academic Conferences International Limited.
158. Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of educational Psychology, 80*(2), 210-216. https://doi.org/10.1037/0022-0663.80.2.210

159. Wijaya, T. T., Cao, Y., Bernard, M., Rahmadi, I. F., Lavicza, Z., & Surjono, H. D. (2022). Factors influencing microgame adoption among secondary school mathematics teachers supported by structural equation modelling-based research. *Frontiers in psychology, 13,* 952549. https://doi.org/10.3389/fpsyg.2022.952549
160. World Health Organization. (2003). *Skills for health: Information series on school health (Document No. 9).* World Health Organization. https://iris.who.int/handle/10665/42818
161. Yong, S. T., Harrison, I., & Gates, P. (2016). Using digital games to learn mathematics: What students think*? International Journal of Serious Games, 3*(2). https://doi.org/10.17083/ijsg.v3i2.113
162. Zabala-Vargas, S. A., García-Mora, L., Arciniegas-Hernández, E., Reina-Medrano, J., de Benito Crosetti, B., & Darder-Mésquida, A. (2022). Didactic strategy mediated by games in the teaching of mathematics in first-year engineering students. *Eurasia Journal of Mathematics, Science and Technology Education, 18*(2), em2082. https://doi.org/10.29333/ejmste/11707
163. Zabala-Vargas, S., de-Benito, B., Darder-Mesquida, A., Arciniegas-Hernandez, E., Reina-Medrano, J., & García-Mora, L. (2021). Strengthening motivation in the mathematical engineering teaching processes: A proposal from gamification and game-based learning. *International Journal of Emerging Technologies in Learning*, *16*(6), 4-10. https://doi.org/10.3991/ijet.v16i06.18633
164. Zabala-Vargas, S., García-Mora, L. H., Ardila-Segovia, D., & de Benito-Crosetti, B. (2019). Motivation increase of mathematics students in engineering: A proposal from game-based learning. In *Proceedings of the 2019 International Symposium on Engineering Accreditation and Education (ICACIT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICACIT46824.2019.9130297
165. Zaki, N. A. A., Zain, N. Z. M., Noor, N. A. Z. M., & Hashim, H. (2020). Developing a conceptual model of learning analytics in serious games for STEM education. *Jurnal Pendidikan IPA Indonesia, 9(3)*, 330-339. https://doi.org/10.15294/jpii.v9i3.24466
166. Zhonggen, Y. (2019). A meta-analysis of use of serious games in education over a decade. *International Journal of Computer Games Technology, 2019*, 4797032. https://doi.org/10.1155/2019/4797032
167. Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. O'Reilly Media.

Acknowledgements

I would like to express my deepest gratitude to everyone who contributed to, or provided support for, the completion of this PhD thesis. Without their support, guidance and insights, I would not have grown as much personally or professionally.

First and foremost, I would like to thank my thesis supervisors. Professor Marina Marchisio Conte welcomed me into her research group and guided me throughout my doctoral studies, encouraging me to grow and strive for excellence. Her expert guidance helped me to navigate my research path and to identify the most appropriate direction for my work. I am also deeply grateful to Engineer Sergio Rabellino, whose incisive and targeted insights, together with his experience, played a decisive role in my academic development. Thanks to his contributions, I was able to expand my knowledge beyond my initial background, making my work more rigorous, in-depth and accurate. Thanks to their help, I was able to refine and strengthen this research, which focuses on the intersection between Mathematics, gaming and digital technologies.

Special thanks also go to the entire Delta Research Group at the University of Turin for their continuous support and guidance throughout these years. The many opportunities for discussion, comparison and collaboration, as well as their moral support, were invaluable to my academic journey.

I would also like to extend my gratitude to Professor Gianni Vercelli for his precious support. Throughout my PhD studies, he acted as discussant and provided insightful guidance that helped shape my research. He introduced me to the field of research with which I was initially unfamiliar, and he also contributed significantly by facilitating networking opportunities.

I would also like to express my sincere gratitude to the university and secondary school students and teachers who participated in experimenting with the game-based activities. By allowing me to conduct this important phase of the research in their classes, they made this study possible. Without their contribution, this research could not have been carried out.

Finally, I would like to express my sincere thanks to Andrea and my family, whose constant encouragement helped me overcome the many challenges I faced during this period.