

Article

A Framework for Standardizing the Development of Serious Games with Real-Time Self-Adaptation Capabilities Using Digital Twins

Spyros Loizou *  and Andreas S. Andreou 

Faculty of Engineering and Technology, Department of Electrical Engineering, Computer Engineering and Informatics, Cyprus University of Technology, Limassol 3036, Cyprus; andreas.andreou@cut.ac.cy

* Correspondence: spyros.loizou@cut.ac.cy

Abstract

Serious games are an important tool for education and training that offers interactive and powerful experience. However, a significant challenge lays with adapting a game to meet the specific needs of each player in real-time. The present paper introduces a framework to guide the development of serious games using a phased approach. The framework introduces a level of standardization for the game elements, scenarios and data descriptions, mainly to support portability, interpretability and comprehension. This standardization is achieved through semantic annotation and it is utilized by digital twins to support self-adaptation. The proposed approach describes the game environment using ontologies and specific semantic structures, while it collects and semantically tags data during players' interactions, including performance metrics, decision-making patterns and levels of engagement. This information is then used by a digital twin for automatically adjusting the game experience using a set of rules defined by a group of domain experts. The framework thus follows a hybrid approach, combining expert knowledge with automated adaptation actions being performed to ensure meaningful educational content delivery and flexible, real-time personalization. Real-time adaptation includes modifying the game's level of difficulty, controlling the learning ability support and maintaining a suitable level of challenge for each player based on progress. The framework is demonstrated and evaluated using two real-world examples, the first targeting at supporting the education of children with syndromes that affect their learning abilities in close collaboration with speech therapists and the second being involved with training engineers in a poultry meat factory. Preliminary, small-scale experimentation indicated that this framework promotes personalized and dynamic user experience, with improved engagement through the adjustment of gaming elements in real-time to match each player's unique profile, actions and achievements. Using a specially prepared questionnaire the framework was evaluated by domain experts that suggested high levels of usability and game adaptation. Comparison with similar approaches via a set of properties and features indicated the superiority of the proposed framework.

Keywords: serious games; digital twins; self-adaptation; real-time gaming experience; learning skills; training



Academic Editors: Oleksandr Kuznetsov and Cristian Randieri

Received: 16 April 2025

Revised: 6 August 2025

Accepted: 8 August 2025

Published: 18 August 2025

Citation: Loizou, S.; Andreou, A.S. A Framework for Standardizing the Development of Serious Games with Real-Time Self-Adaptation Capabilities Using Digital Twins. *Technologies* **2025**, *13*, 369. <https://doi.org/10.3390/technologies13080369>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Serious games (SGs) are considered significant learning and skill development tools in various fields, such as education, industry training and healthcare. In comparison to

traditional entertainment games, SGs are designed to achieve specific objectives, such as improving skills, teaching new concepts and delivering training through simulation using real-world scenarios [1]. This type of game offers interactive and more engaging experience as technology advances (e.g., based on augmented or virtual reality). However, SGs face several critical challenges related to personalization, adaptability and effectiveness when serving the needs of users [2]. Notable examples of such challenges are the need to address players' demands for adaptation, efficient management of the game's smooth progression and gameplay, and handling of the game's difficulty (levels or scenes), all in real-time. Tackling these issues will ensure that players are indeed benefited by the learning approach followed, they enjoy a personalized game environment tailored to their needs and skills, and their interest in the game is kept high thus maximizing the educational experience.

Real-time adaptation in modern software systems may be achieved through integration with Digital Twins (DTs). A DT is practically a digital replica of a physical object which allows experimentation with various parameters aiming to study their effect in a controlled environment before applying certain values to the real world [3]. This investigation also allows the collection of information related to the environment and how it is affected by changes that actuators perform through feedback cycles. Although mainly used in different contexts (e.g., smart manufacturing or healthcare delivery), a DT may enable a SG to respond dynamically to players' behavior and performance.

This paper presents a framework which offers a phased and structured SG development process supported by standardization of the steps to produce dynamically adjustable games using DTs. The standardization provided is not aligned with any public or industry wide standards for games development, such as, the IEEE P2948 (Cloud Gaming), or P3341 (Mobile Game Experience) [4,5], the IMS Learning Tools Interoperability (LTI) or Learning Design (LD) [6,7], or SCORM [8]; it aims primarily to facilitate portability across platforms, improve consistency and support integration of different game elements. The DT in our case has a special form that allows changing the way educational content is delivered and training activities are performed. This is achieved by making certain adjustments (e.g., changing the scenes or scenarios, or modifying the complexity/difficulty) for a specific sample user profile category and then measuring the response of the trainees. In this case the physical object is the educational tool used, that is, the training approach, content, performance assessment and adaptability according to types of users.

The framework is general and domain-agnostic as it is not tailored for or attached to a specific application area. Instead, it provides the means, activities and structures to serve practically any field, with experts supporting the transfer of domain knowledge which is reflected in scenarios, scenes, goals and gameplay. The way scenarios, levels and game elements are defined is flexible and independent of the application area, making this approach applicable in a wide range of domains, not just the industrial training or therapeutic cases used in this paper for demonstration purposes.

The proposed framework and associated methodology are demonstrated via two case studies, the first being utilized as a therapeutic game and the second being integrated into a smart manufacturing environment. More specifically, the former is built to support training delivered by the Rehabilitation Clinic of the Cyprus University of Technology (<https://www.cut.ac.cy/faculties/hsc/reh/rehab-clinic/?languageId=1>, accessed on 16 March 2025) to people (children) with intellectual disabilities. The latter is used in the Paradisiotis Group (PARG) poultry meat factory (<https://paradisiotis.com/>, accessed on 10 April 2025) and is related with training engineers for monitoring a device that controls climate within chick farms. The diversity of the games developed indicates the generalizability of the framework as the different settings, operational environment and educational targets are served equally well despite the application domain. In these case-studies the

game environment supports efficiently different user training needs, such as to overcome certain types of learning weaknesses or enhance their experience on the use of complicated machinery by allowing personalized and self-adaptive game experience. The interaction between the DT and the SG facilitates the automatic adjustment of the gameplay, educational material and level of difficulty by processing information collected during the use of the game (we call this the dynamic phase), but also after the training session has been concluded (static phase).

The main contributions of this paper may be summarized to the following: (i) Provides a phased, stepwise approach for developing SG from scratch, which is domain agnostic; (ii) Enables real-time, automatic self-adaptation of SGs via the integration with DT; (iii) Offers internal standardization by introducing a formalized structure for describing game elements, scenarios, and adaptation rules thus supporting consistency and portability.

The rest of the paper is structured as follows: Section 2 discusses related work and provides a brief overview of the technical background behind SGs and DTs. Section 3 presents methodology and the proposed framework and describes how the DT and analytical models work synergistically to allow the SG to adapt itself based on player experience. Section 4 demonstrates the framework through the two real-world case studies mentioned above, while Section 5 provides a case-studies design, short scale experimentation and performs a comparative analysis with other approaches that share similar characteristics with the proposed one. Finally, Section 6 concludes the paper and highlights future work steps.

2. Technical Background and Related Work

This section describes technologies that form the backbone of the framework and presents how these technologies can be blended to enhance user interaction and system efficiency. The core of the proposed framework combines SGs, DTs, semantic annotation and predefined rules to enable real-time self-adaptation and personalized learning experiences.

Literature review conducted in this study followed a short but focused and structured approach to ensure relevance and clarity. A selection of recent peer-reviewed journal articles, systematic reviews, and established frameworks in the fields of educational technology, serious games, and adaptive systems was performed aiming at identifying research gaps directly related to the development, standardization and adaptability of serious games, especially real-time self-adaptation and integration with digital twin technology. The review of these sources was brief, giving priority to works that best matched the main ideas of the suggested framework. The search process aimed at tracing articles indexed in Scopus, Science Direct, IEEE Xplore, ACM Digital Library, SpringerLink, Google Scholar and Wiley Online. Inclusion criteria focused on studies that addressed standardization in game development, integration of digital twins, and real-time adaptation mechanisms. This set of criteria consisted of the type of paper, publication year, publication venue and number of citations. Only scientific papers published in recognized venues with a significant number of citations were included in the final set of papers. This targeted approach allowed collecting findings and best practices from the relevant literature, ensuring that a compact, yet solid background was formed that enables the identification of gaps and challenges and act as an informative guideline for the design and validation of the proposed framework.

SGs are games designed with more than just entertainment features; they also serve informational and educational goals [9]. A few studies have been focused on integrating real-time feedback from DTs in the context of SGs, but several papers have discussed the need for more adaptive game user interaction. Although DTs have been used frequently for monitoring and simulation across a range of disciplines, it is still new to integrate them with interactive environments like games and analytics systems.

2.1. Serious Games

Serious gaming is used for purposes other than entertainment, also known as applied games. They are a type of gaming that serves many disciplines including education, emergency management, sports, engineering and healthcare. Furthermore, a range of applications for simulating users, which include tracking player behavior in a virtual environment, as well as detecting and analyzing anomalous behavior, can evaluate player performance using computer vision and deep learning-based techniques [1].

Several research studies address the problem of adapting SGs based on user needs and goals. These studies focus on user engagement, improving learning skills and personalizing game experience [1,9]. Other parts of the literature discuss the dynamic adjustment of game difficulty based on user performance or investigate how game mechanics targets for learning or therapeutic methods [10,11]. There is still a gap in integrating real-time adjustments based on DTs to further personalize game experience [10]. Combining DTs with SGs will create self-adaptive and effective strategies, which is exactly what the present paper aims at.

Damaševičius et al. [12] present challenges related to the lack of standardization in the creation and assessment of SGs. This work also reports lack of knowledge about basic processes underlying game-based actions and limited integration of these games into current healthcare systems.

The authors in [1] argue that gamification in SG continues to face several challenges, such as unclear effectiveness of game mechanics for health interventions, difficulties in maintaining long-term engagement and lack of empirical validation of outcomes. The authors also note that we still do not fully understand how SGs work to produce outcomes. As a result, they are sometimes incompatible with emotional engagement, motivation and learning. Customizing SGs to meet the needs of each individual player is another significant challenge. It is not easy to find the best combination of game mechanics to keep players motivated and promote behavior change. Moreover, they suggest that it is better to combine many theories like self-determination and flow, into an overall framework that can direct the creation of more effective gamified experiences. These challenges show how much more study is required to make these systems impactful and adaptable.

Kara [13] argues that different studies deliver different results because there are no standardized ways to measure efficacy. Additionally, that paper suggests that the potential of using trend technologies, like virtual reality (VR) and augmented reality (AR), are not yet usually used in SGs. It also notes that considerable research focuses on adventure games leaving out other types like puzzle and simulation games. These gaps suggest exploring more methodologies and game types to fully understand the potential of SG in educational tasks.

Serious games often incorporate biofeedback and immersive technologies to reduce stress and enhance well-being [14]. Several studies report that serious games incorporating mindfulness, guided breathing, and light cognitive tasks can induce measurable relaxation effects [15]. For example, games designed with calm visual and auditory stimuli help users enter a parasympathetic state, reducing anxiety and muscle tension. Virtual Reality (VR) games like Deep and FlowVR immerse players in peaceful underwater or natural environments, requiring breathing-based controls that promote deep, slow respiration. These have been shown to significantly reduce self-reported stress and anxiety scores in both healthy and clinical populations [16]. In addition, researchers frequently use EEG (brainwaves), ECG (heart rate), and EDA (skin conductance) to evaluate the effectiveness of relaxation-focused serious games [17]. EEG studies show that increased alpha and theta activity—associated with relaxed wakefulness and meditative states—are more prominent during and after game play. For instance, a study using a neurofeedback game found a

rise in frontal alpha power, correlating with reduced anxiety levels [14]. Heart rate variability (HRV), a key indicator of autonomic nervous system balance, also improves after gameplay in many studies. Greater HRV indicates a shift toward parasympathetic dominance, suggesting reduced stress. Similarly, decreases in skin conductance reflect reduced arousal [15]. There are also studies in the literature that report the use of physiological data for adjusting game parameters. For example, RECOGNISED is a serious game that uses real-time EEG input to modify in-game environments based on user relaxation levels. Clinical trials showed improvement in sleep quality and reduced stress in participants [17]. Other examples are NeuroRacer and similar games that are tailored for older adults, which also demonstrated improved emotional regulation and stress recovery [15]. Finally, rehabilitation games that use motion sensors and biofeedback help patients manage pain and anxiety during recovery [14]. Digital games have become a popular tool in pediatric speech therapy because they offer innovative ways to activate children in treatment [18]. That review paper analyzed various studies on speech therapy games focusing on benefits and challenges. These games increase motivation for children to continue speech therapy providing an interactive experience. The use of digital games in this therapy is a promising solution to make therapy more accessible and fancier for children.

2.2. Digital Twins

A DT is a real-world representation of a physical product or process that operates in the same way. It is commonly used to optimize error detection and correction, reduce defects and manage the IoT lifecycle to save time and cut costs [19]. Moreover, it is a powerful technological tool as it can stream, optimize and analyze data in both the real and the digital world. The main benefit a DT offers is the ability to simulate real-world scenarios, predict outcomes and perform decision-making based on data. As DTs can help adaptive systems adjust in real-time, it remains challenging to use real-time data from to offer personalized user experience in interactive environments like SGs [10,20].

DT integration with adaptive systems face difficulties [21]: (i) there is no standardization for integrating data or third-party systems; (ii) current models do not support dynamic updates to reflect changes in their physical environment by processing large volumes of real-time data with accuracy; (iii) how to achieve self-adjustment and interaction with other systems is still an ongoing area of research.

Adaptive Digital Twins (ADT) face several open challenges, such as difficulty in real-time synchronization and optimization in dynamic systems. Furthermore, dynamic adaptability is quite complex when integrating tools and data sources due to lack of standardization [3].

Erkoyuncu et al. [22] suggest that integrating existing “brownfield” systems with DT architectures requires dynamic updates. Additionally, it is still difficult to maintain synchronization and real-time adaptability while promising compatibility across different data sources. The design for managing a wide range, volume and velocity of data presents an additional difficulty.

Various open challenges exist in the incorporation of the DT framework into personalized SGs based on therapy. Some of these are as follows: (i) the integration of a variety of data sources; (ii) enabling real-time analysis, simulation and feedback for personalized therapy requirements; (iii) the unique therapeutic needs of each patient; (iv) standardization for data collection, storage and sharing across different devices; and, (v) limits on the integration of DT frameworks into healthcare systems [10].

The combination of gamification, virtual reality (VR) and DTs is a powerful tool that enables users to visualize systems in a virtual world. The work in [23] presents the

challenges of such an approach and highlights the need for high quality real-time data synchronization between real and digital worlds.

None of the studies thus far have explored adequately the standardization in integrating DTs with SGs for creating self-adaptive games based on user interactions and abilities. The aforementioned gaps and challenges observed in the relevant research on DTs supporting interactive and adaptive environments are significant obstacles to the development of dynamic systems that enable real-time adaptability and personalized experience. Some of these challenges are addressed in the present paper focusing on providing a conceptual framework for creating efficient and effective SGs with self-adaptive capabilities by combining DTs with SGs offering at the same time standardization, real-time game data collection and semantic annotation.

3. Methodology and the Proposed Framework

To the best of our knowledge, no research has been documented thus far that shows how the technologies described in the previous section may be combined in a standardized approach to creating a self-adaptive gaming system able to personalize user experience. The target of the proposed framework is to offer a standardized approach to developing serious games with real-time adaptation using data-driven insights. These include goals definition, data collection and processing from game experience, expert consultation, analytical models, digital twin functionality and automatic gameplay adjustments. Figure 1 depicts the conceptual process schema of the framework, which is analyzed later in this section.

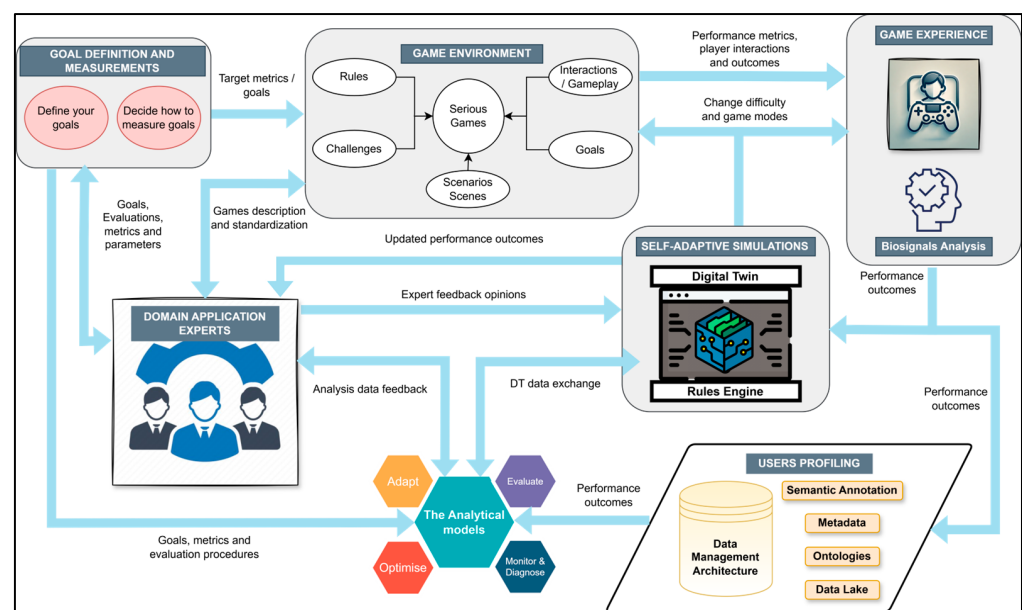


Figure 1. The conceptual architecture of the proposed framework.

The methodology used for defining the framework introduced in this paper was as follows: First, we identified three key pillars upon which we structured the phases of the framework, that is, Functionality, Stakeholders and Data. Functionality pillar enclosed all game elements that give birth to the scenarios, scenes, tasks and gameplay. These elements constitute the environment of the game and as such a critical part of the framework is devoted on describing them effectively and efficiently through dedicated descriptors and standardized formats that will be analyzed later on. Stakeholders pillar is considered the cornerstone of the framework as it serves two purposes. First, it includes domain experts whose valuable guidance and knowledge drive the development of the game. Second, it involves the end-users, that is, the enablers of the game (e.g., administrators that initiate the

game and make sure it will execute as planned) and the actual users that engage with the game and produce performance results. Data is the final pillar and encompasses all related information, starting with game elements and ending with the outputs of the game during user interaction. These three pillars were thought of as living interconnected parts of a larger ecosystem, the integration of which serves the following purposes: The functionality of a game is described in its environment, which is driven by a set of goals. These goals are defined from the beginning utilizing domain experts. Game tasks are oriented by rules and gameplay which must also be part of the environment. The execution of the game by its end-users produces game experience data which should be recorded in profiles and analyzed further. Therefore, corresponding mechanisms for profiling and managing data should be in place, allowing for models and techniques to monitor, analyze and evaluate it. The core of the framework is its ability to offer to a game self-adaptive properties in real-time. To serve this purpose there should be a game engine in place governing adaptability decisions, which are taken based on the processing of the data and feedback from experts where necessary. This engine would be built to combine elements of Digital Twins, a powerful technology not fully exploited in the past for SG, and a rule-based approach that should enable bridging game requirements and goals with gameplay. This was in a nutshell the philosophy behind the structuring of the framework introduced in this work.

The proposed framework is divided into a series of phases that start with setting the goals of the game according to players' needs, and define the gaming environment (e.g., scenarios, gameplay, rules, etc.). This environment is built with the support of domain experts, and it is formally described using standardized semantic forms, such as ontologies and blueprints. These forms are designed to make it easier to describe, parse and transfer game elements and data described within the proposed framework. A key feature of our methodology is its hybrid nature, with the potential of utilizing domain experts to assist in defining and/or refining game rules and scenarios, while the DT adapts the game in real-time based on player data (e.g., progress) and game rules. Progress is measured through performance metrics, such as completion times, interactions and performance outcomes. Every interaction is tracked and valuable data is collected that helps understand how the players respond to distinct game scenes. Analytical models process the data to evaluate a player's progress and feed the adjustment phase of the game. Specifically, the output of these models is used to adapt the game experience in real-time, adjusting scenarios based on the observed players' behavior and patterns. This is feasible as the game structure is designed to respond to players' actions in real-time, that is, the game adapts itself by adjusting the complexity of the educational content delivery ensuring that the player is challenged at the correct level of difficulty.

The process starts by defining the system objectives: training, performance or skill improvements and goals of the game in terms of user education and service, as well as the way to assess the results (upper left corner). For example, with the goal being to improve memory skills, a game may include a scene for solving a puzzle or matching hidden tiles that periodically open/close randomly. Measurements in this case provide an evaluation of the game's effectiveness, including completed task time, interaction frequency, error rates, etc.

The goals and metrics orient the *Game Environment*, the latter including the core aspects of the game, such as the gaming rules, game objects (images, videos, text, music, etc.), scenarios/scenes and challenges to be faced, associated with partial goals for each scenario/scene. The structure of the game is built in a way that is adaptable to different roles or requirements of the players corresponding to their condition (e.g., patients that have suffered stroke), or skills/abilities (e.g., newly hired factory engineers under training). More specifically, the framework supports three levels of adaptation: (i) Parameter-Level,

that adjusts values for certain parameters (e.g., difficulty, time limits), (ii) Scenario-Level, that modifies game tasks, rules or environmental elements such as switching from syllable recognition to word formation, and (iii) Structural-Level that modifies core gameplay mechanics and content by introducing new storylines or interactive objects. Adaptation at each level is performed using appropriate rules as described in the subsequent paragraphs.

Expert consultation is utilized for designing the game context and structure. Expert support has been extensively used in the literature: Hocine et al. [24] report the engagement of experts in the design process to ensure that the game precisely meets user needs, as well as the ability for experts to select the game scenarios and the number of targets and also update game parameters. Alcover et al. [25] employ experts to validate and test game elements, or define the characteristics of the interaction mechanism. Experts are also used for configuring games, assist in composing the sequences of actions organized in levels and define parameters and runtime behavior for these actions, allowing the personalization of game exercises [26,27].

Expert input is critical to align the game scenarios with the educational goals of the potential players (e.g., therapeutic/rehabilitative or machinery hands-on experience). For example, a domain expert may recommend calming visuals and music for players/patients with stress. While the current version of the therapeutic game leverages expert input to provide supportive elements, such as calming visuals and music for users with heightened stress, no biosignals or physiological measurements are recorded as part of the adaptation process. However, it is recognized that including real-time biosignals (such as heart rate or skin conductance) in future iterations could provide deeper, objective insights into stress and relaxation responses, further enhancing the game's ability to tailor interventions to individual needs. There is a feedback cycle between the game environment and the experts, which enables adding new or modifying existing game artifacts that will be accessed by the DT to conduct the self-adaptation process. The experts define also the rules which will guide the DT to perform adjustments in terms of complexity and difficulty. The integration between experts and DT is a feedback loop starting with the initial setup and continuing with periodical re-definition of rules and scenarios taking place when necessary (i.e., based on new educational targets). These rules are incorporated in the DT's Rule Engine and relevant feedback may be provided to the experts to support their re-definition when necessary. As previously mentioned, the rules can vary depending on the level of adaptation, ranging from simple scene adjustments (objects, difficulty, gameplay) to complex, deep structural changes to the content, educational approach (e.g., increase in learning curve) and game style (e.g., from simple tile or image matching to incorporating AR or VR elements depending on the game implementation). The multi-layer adaptation capabilities are therefore driven by DT's rule engine, which supports expert-defined XML representations to align with real-time players' data.

Users interact with the game (phase *Game Experience*), and their sessions produce data, which is continuously updated. Player interactions, behaviors, decisions and performance metrics, such as scores and completion times, potentially enhanced by biosignals of the players recorded during playing sessions (e.g., ECG, heart rate, blood pressure) are examples of game experience data. This information essentially feeds the self-adaptive part of the framework performed by the DT, while it is also stored in the user profiles. The adaptability feature is ensured by using a specific, formal and standard description of the current rules that are active within the game environment. This description is expressed using a dedicated XML format structure, accompanied by expert consultation (see Appendix A, Figures A1 and A2).

As mentioned earlier, the dedicated XML structure formalizes the description of game scenarios and parameters. This standardized format may be considered internal

and it is used to define all related elements, such as game scenes, player difficulty levels, game tasks, and performance metrics. The term “internal” is used here to differentiate between the use of a globally accepted format, such as XML or JSON files, to express game elements and does not refer to universal standards for educational technology or game development as explained previously. At the same time, this standardized format offers on one hand a unique and unambiguous structure of the game, and, on the other, the means for the proper communication between the game environment and the DT, thus enabling self-adaptation and real-time adjustments. For example, in the game with the therapeutic purpose mentioned in the Introduction and utilized later in this paper for experimentation, the XML defines the relevant scenarios and parameters, such as task type (syllable recognition), difficulty level (beginner, intermediate, advanced) and success criteria (pass, fail, repeat). Furthermore, this structured XML relates to metadata that describes the above parameters so that their dynamic adjustment, based on player feedback and performance data collected during gameplay, is made feasible.

Figure A1 in the Appendix A demonstrates the XML structure used for describing game scenarios and parameters for the therapeutic case-study game. Each scenario utilizes this XML for defining tasks with specific difficulty levels, game components, challenges, performance attributes and for guiding the gameplay progress. Figure A2 (Appendix A) describes the dynamic elements of real-time performance tracking, player interactions data and feedback loops. These elements are used by the DT to modify the game parameters based on real-time performance metrics. Essentially, the DT parses the corresponding XML file and recognizes the rules set. For example, the DT assesses a player’s stress levels, scores, response time and error rates. This information triggers adaptation rules to increase or reduce task complexity or provide hints to assist users successfully completing a scene or game task.

Data collection describing player interactions may include structured, unstructured and semi-structured data related to performance, decision-making patterns and engagement. For example, images of game scenes or the user’s eyes (eye tracking system), audio files of the player’s recorded voice answering questions during gameplay, numerical interaction scores, etc. Semantic annotation techniques tag this data (i.e., generate metadata information) and enable its easy future retrieval for further analysis. For example, historical metadata like game mode and difficulty level can be used to assess a player’s reaction time to graphical signals. Ontologies and metadata play an organizing role for managing the collected data during gameplay. Ontologies provide a formal representation of knowledge because they define the relations between data, such as different scenarios, tasks, or game objectives in the game environment. This enables a classification of game objects, players interactions and performance metrics. Metadata provides full descriptions of information about data collected, such as the type of data (structured, unstructured, semi-structured), sources and its relevance to specific game objectives. The framework ensures that data is categorized and is easily accessible for real-time adaptation and analysis. This integration offers DT the ability to make decisions about game adjustments and provide a more personalized and effective gaming experience.

The framework uses four analytical models, namely *Monitor & Diagnose*, *Evaluate*, *Adapt* and *Optimize*, to perform processing of players’ interactions and guide self-adaptation. The *Monitor & Diagnose* model inspects in real-time the game environment and player data making it available for processing. The *Evaluate* model assesses playing behavior and detects deviations or faults in specific tasks compared to the targets. The *Adapt* model suggests specific changes to improve gaming experience, like modification of game scenes or gameplay modes. The *Optimize* model uses historical data to forecast future player performance and guide possible adaptations. In general, the last two models provide the DT with

insights for applying and monitoring potential adaptations thus delivering personalized experience. The XML structure described earlier also operates as the interface between the game environment and the analytical models. It organizes data in hierarchical forms and ensures that every gameplay element is annotated and linked to the corresponding rules. This link facilitates the real-time execution of the models, allowing for monitoring, evaluation, adaptation and gameplay optimization to be performed dynamically based on the XML-defined parameters. For example, a game scenario where players match syllables to words can include XML annotations for difficulty (e.g., two-syllable words), expected completion times, and error thresholds. Player interactions are stored and updated also in XML format. Updates of the interactions trigger analytical models to analyze player behavior and support adjustments. By adding adjustment rules into the XML schema, the framework ensures that the game is self-adaptive and consistent across different scenarios and player profiles. This adaptability is key to achieving personalized learning.

The *Game Experience* phase records interaction data to be used for further analysis and self-adaptation. Player interactions, performance metrics and outcomes are collected here including every selection made and the response time within which it was made, all user actions (e.g., clicking correctly or not, canceling or re-loading stage), and completed challenges. Also, Game Experience offers the potential to exploit biosignals via a dedicated Biosignal Analysis module. This component is designed to acquire in real-time and process physiological data such as heart rate, skin conductance, or EEG signals, when and if such information is available. Its purpose is to add an additional layer of insight into the players' state such as stress, relaxation or emotional engagement thus enabling more precise adaptation of the game environment. Although biosignals data are not currently utilized for real-time adaptation in the case studies described later in this paper, the proposed framework structure allows for their integration as an additional input stream, enhancing personalization and supporting a more holistic evaluation of user experience.

To protect user privacy, all collected data is first anonymized, meaning that any information that could directly identify users is removed, while each user is associated with a unique internal ID. Other sensitive data, such as the profiling of a user for game adaptation (e.g., a syndrome) or their performance data, is encrypted during acquisition and storing so that the relevant data remains unreadable to unauthorized users even if they gain access. Additionally, this information is logged and stored in each player's profile using semantic annotation and metadata. The metadata may include the player's code, game details (game ID, scenario ID, scene ID), session (timestamp), type of data (structured (e.g., numerical), unstructured (e.g., voice), semi-structured (text and video), task type, difficulty, player performance and attributes, etc. (see Appendix A, Figure A4 for XML representation). This data management scheme enables building a detailed dataset for each player to facilitate analysis and future automatic adaptation of the game. The datasets are stored in a Data Lake (DL) architecture, and allows for the semantic separation of the information stored in the DL based on the aforementioned metadata, and by using ponds and paddles [28]. The ponds correspond to the type of data (e.g., structured, unstructured), while the paddles lead to identifying the part of the semantic annotation that a portion of information is described by. This ensures that no matter how big the volume of data is produced by multiple, simultaneous players/users, the information can be efficiently extracted from the DL [29]. The latter is able to efficiently handle vast amounts of data that is being generated from multiple simultaneous users at different frequencies and formats resembling Big Data. Therefore, this metadata scheme may also be used to facilitate big data management [30]. The DL architecture of our framework is flexible to support the definition and management of any game environment, that is, scenarios, tasks, gameplay and central data. This reflects to practically all known game types or genres, with the

proposed framework being able to accommodate all distinct characteristics of a game type without any dependencies or limitations hindering the setup of the environment for any such kind of game, including those that give birth to big data game environments.

Semantic annotation provides a standardized description of the gameplay components, which, on one hand, ensures data quality, that is, the information stored in the DL is always complete based on the metadata, and on the other hand, it enables organizing this information in a straightforward and easy manner using approaches such as ontologies, knowledge graphs and tagging techniques. A dedicated ontology describing the game environment is depicted in Figure 2. A game consists of several different scenarios, each comprising one or more scenes. Each scene requires the completion of various tasks. A task is completed using a specific set of actions or gameplay, while the process of executing it is monitored and assessed based on different performance indicators using specific metrics. This stage guarantees that the game environment is in line with the learning goals and effectively adjusts to player needs by modifying the structured metadata (i.e., adding new or updating existing game artifacts).

The framework utilizes a set of rules encoded in the DT rule engine to guide the adaptation process. These rules dynamically modify the game environment based on players' performance and selections and can vary from simple changes in the graphical layout (e.g., colors or lines), to deep structural changes involving scenarios and educational content. The pseudocode depicted in Figure A3 (Appendix A) shows the steps for monitoring players' behavior, evaluating performance based on predefined thresholds and applying adaptation rules. For example, the input data is constituted by the player performance metrics recorded during gameplay, evaluation compares these metrics to detect patterns or values/thresholds (e.g., high error rates), and, finally, adaptation rules are triggered based on the evaluation results and the relevant actions.

Self-adaptation is performed in real-time via DT simulations, which create a feedback loop integrating player data, suggestions from the analytical models and expert guidance. This is performed using a two-stage approach, (i) Temporary, and (ii) Permanent. According to the former stage, the game environment is adapted partially in terms of complexity and difficulty (i.e., in some scenes or gameplay modes) based on the current goals and the player's performance thus far. For example, the game may reduce thinking tasks or add more visuals to help players understand them better. Temporary changes are then assessed and if a player (or group of players) shows sufficient alignment and smooth progress (these are reflected in the adaptation of the partial goals which accompanies changes in the game environment), then they are considered permanent. If not, then they are dropped and reversed. Permanent changes mean that the changes are enforced to more scenes and game characteristics, which gradually adjust the game globally to make it more/less difficult. These self-adaptive adjustments are thus performed in line with players skills and game objectives, keeping players interested and challenged, while delivering the best possible support of the educational goals.

User profiling and data management phase records and manages the behavior of a player when using the game (game objects selections, response time, error and success rates). This entails storing historical data (e.g., past interactions time and performance metrics) in a storage architecture (in our case a Data Lake). Profiling enables the processing of historical and real-time data to achieve personalized gaming experiences evolved with player progress. For example, with a player exercising memory strengthening there exists a timer challenge for solving puzzles. According to the progress of the user, future scenarios are adjusted to provide less/more time to solve the puzzle or increase/decrease the number of each puzzle's parts.

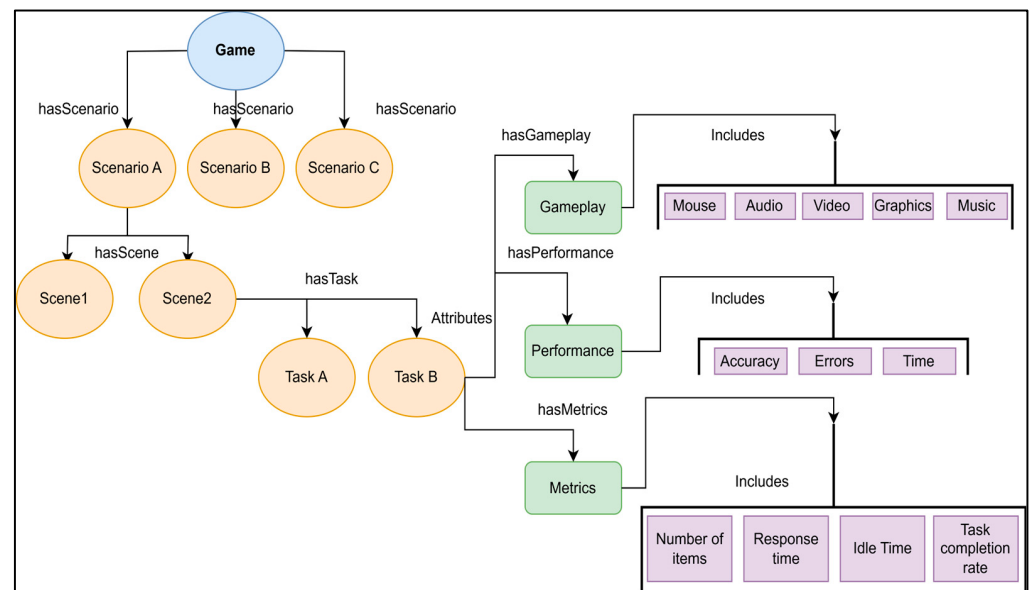


Figure 2. An ontology for the game environment.

4. System Demonstration

The proposed framework is demonstrated in this section using one of the two real-world case-studies that were developed for experimentation purposes (see next section). This example will illustrate the steps followed using Figure 1 to provide the required functionality and gameplay adaptability level of a SG integrated with a DT. Specifically, the SG supports the activities of the Rehabilitation Clinic of the Cyprus University of Technology for the education of children with syndromes that hinder their learning abilities. The game scenarios developed aimed to assist these children learn how to process syllables.

Goal definition and measurement focus on defining the goals based on each child/patient learning needs. These goals are the basis for evaluating progress and guiding the automatic game adaptation process. The primary goal was to improve the child's/user's ability to segment words into syllables accurately. Sub-goals included improving recognition accuracy to 85–90% and gradually reducing response times for identifying syllables by 5–10%. The target metric for accuracy was to achieve 4 correct answers out of 5 during a session. Base metrics in this stage also included initial and final reaction times (i.e., the first and last time the game task was executed within the same time frame—day, week), error rates and task completion accuracy. These metrics allowed tracking progress over time and dynamically adjusted the task or its gameplay characteristics to increase performance.

The game environment was designed with the aid of speech pathologists at the rehabilitation clinic so as to be simple but also attractive (e.g., colors, icons, figures, etc.), and at the same time be interactive to motivate children to participate in the educational tasks. This case study highlights the hybrid workflow of our approach, where domain experts guide the initial setup and ongoing redefinition, while the adaptive system manages real-time adjustments based on players' performance. This stage included also the production of audio and visual components that were relevant to each child's level or category (i.e., syndrome, capacity, abilities). The game environment presented a scenario where the player undertakes tasks to segment words, recognize syllables and select the correct answer between multiple options. The responses and interactions were recorded in logs for further analysis. The tasks were designed to align with the goals of the previous stage.

The game basis was to ask players to select the number of syllables from options that appeared in blocks (e.g., 2, 3 or 4) when a word was presented in a colorful image (see

Figure 3a for the word “balloon”). If the player selected the correct answer (2 blocks), the game provided positive feedback like a happy sound or balloon animation.

The contribution of domain experts was quite important. Structuring the data allows standardizing game descriptions provided by the experts without the need for technical knowledge as in the case with the speech therapists of our example.

Instead, using semantic annotations the experts were supported to refine, validate and align the game context with the therapeutic goals. Furthermore, this annotation facilitated the definition of rules, thresholds and adjustments while evaluating gameplay results. Specifically, the experts set the starting point to be the easy level with words of 2 syllables (e.g., “balloon”), the threshold to 4 correct answers for completing the game task successfully, and if the child achieved 85% accuracy (number of correct answers out of the total words displayed) then to advance to 3-syllables words.

As previously mentioned, the self-adaptive simulation (DT and rules engine) dynamically adjusts the game difficulty in real-time based on player performance. The DT processes gameplay data and applies certain rules set by domain experts to modify game tasks at a certain pace and level. Essentially, the DT continuously monitors the response of users to these modifications based on performance data. Then, it applies the relevant rules for lowering or increasing the difficulty/complexity of the current game task either partially or fully. In our example, if the player faced difficulties, then the adaptation process would trigger a slower or marginal differentiation and enable the provision of hints, such as underlining the correct answer, or splitting the word into the correct syllables and displaying them on screen. Otherwise, if progress showed that it was too easy for the player to answer (this was measured by counting the number of correct answers and the time it took to provide them) then the adaptation would make the task harder at higher levels of change, for example, by increasing the words to 3 syllables, or by using words having less common sounds like “br” as in “bridge”. Two more tasks were additionally developed in our case-study game with increasing difficulty for adaptability purposes, which involved (i) replacing one or more letters of a given word and creating new words, and, (ii) listening to the pronunciation of various words and choosing the correct image out of the options displayed (see Figure 3b). The adaptation process followed the logic that is described in the pseudocode of Figure A3 in Appendix A.

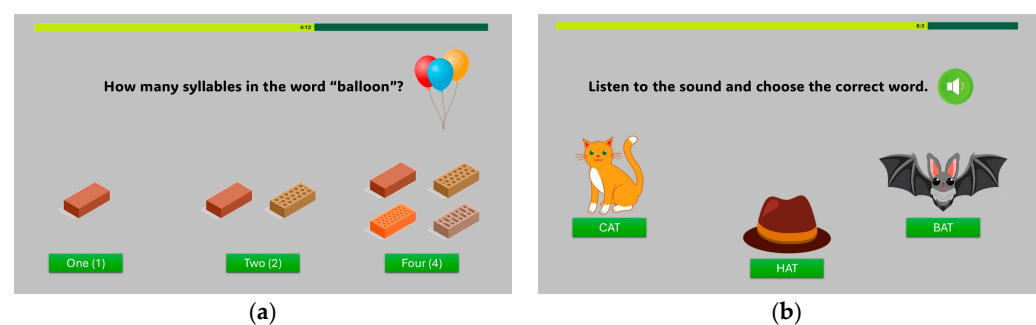


Figure 3. (a) Game environment sample screen; (b) Different types of game elements.

User profiling included a detailed record of every player/patient gameplay history, storing trends, performance and preferences to inform future sessions. Visualization of performance metrics was used by our experts (speech therapists) to monitor progress and adjust (manually) the rules of the game, when necessary, so that adaptation in real-time was performed efficiently. In this preliminary demonstration, five users/players (children) up to the age of twelve years old with learning difficulties due to various syndromes (e.g., Down, Williams) were engaged in the gaming process and used the game task of recognizing the syllables of words. The private nature of the participants’ data was preserved through

mechanisms to prevent unauthorized access to the system, as well as the utilization of data anonymization and encryption techniques. The task was carried out for one hour with system adaptation for the difficulty level (from 2 to 3 syllables and by using words with “br” and “ch”) and under supervision by our experts. Some of the performance metrics that were recorded during the process are depicted in Figure 4, which shows performance metrics, such as accuracy, error rates and completion times, across all players. It should be noted that in the case in which the player did not correctly identify the syllables of a word, the system repeated the display of that word two more times, the first with an indication “Try again” and the second with a hint. (c) Task Success Rates illustrates the success ratio of players that finished the task and recognized the syllables for all words correctly, thus offering a measure of the overall performance.

Therapeutic Game Evaluation Dashboard



Figure 4. Graphical dashboard for domain experts.

By processing the information above, the analytical models yielded visualization elements incorporated into dashboards for the team of experts and game developers (e.g., the graphs of Figure 4). These visualizations helped the experts to inspect progress and decide where and if current rules of adaptation should be modified for adjusting gameplay and improve overall outcomes. In our example, analysis of results for the task with the recognition of syllables displayed trends in time improvements, performance percentage of accuracy for different syllables, and type of recommendations for the next stage of the game (going up to four-syllable words) which suggested that no change was needed thus far in the rules. While Figure 4 provides examples of graphs that domain experts can utilize and presents the fully developed therapeutic game evaluation dashboard using Python 3.11.5 (<https://www.python.org/>, accessed on 24 March 2025) and Dash (<https://dash.plotly.com/>, accessed on 24 March 2025). This dashboard presents multiple data visualization elements

into a uniform interface, enabling domain experts to monitor real-time metrics and historical trends. The dashboard shows performance metrics, such as accuracy, error rates and completion times, across all players. Historical data tracking using pie charts and bar graphs offers insights into long-term trends and performance behavior. Adaptation logs are also displayed in a tabular form, listing adaptation rules, something that provides transparency into game dynamic adjustments. The main purpose of this dashboard was to give domain experts actionable insights and tools for supporting them refine game rules and optimize therapeutic outcomes.

It should also be noted that the experts who participated in the short user experimentation session (two speech therapists) acknowledged the flexibility offered by the proposed framework to change the rules and the game environment as they deemed necessary. They were also very satisfied with the self-adaptation capabilities of the game as they were freed from continuous monitoring of the children's progress so as to change the difficulty levels by hand. Figure 5 presents an excerpt of the tool offered to domain experts for adjusting the settings for each game task. As shown in this figure, the tool allows experts to modify variables such as complexity of words, timer settings and accuracy requirements, and to enable/disable hints. Equally importantly, the experts were committed to collaborating with the authors in the future to develop further the rule-basis of the game and extend the experimentation process to cover more children for longer periods of time.

Adjust Difficulty Settings

Select Stage: Stage 1

Select Word Complexity: ☒ Easy ☐ Medium ☐ Hard

Set Timer for Stage (Seconds): 20 25 30 35 40 45 50 55 60

Number of Words in the Stage: 2 3 4 5 6

Minimum Accuracy Requirement (%): 50 55 60 65 70 75 80 85 90 95 100

Enable Hints: ☒ Enable Hints

Figure 5. Tool for adjusting various parameters via the expert's dashboard.

5. Preliminary Experimentation and Evaluation

5.1. Case-Studies Design

The scope of this section is first to provide a preliminary experimentation process for assessing the proposed framework and then perform a comparative evaluation against similar approaches.

For experimentation purposes, two SGs were developed that cover distinct key domains: Healthcare and Industry 4.0. The diversity of the games demonstrates the flexibility and generalizability of the framework, and its potential to adapt to a wide range of application areas. The two game examples have been chosen so as to show the framework's applicability across these two very different application domains with distant user groups characteristics and requirements. The speech therapy game targets children with special needs, requiring therapeutic personalization. The factory game targets adult engineers, focusing on operational skill training and real-time adaptation to performance. Therefore, through these case-studies it is demonstrated that the proposed approach is not tied to a specific application area but is instead general and flexible enough to be applied to various domains as long as the targets, rules, scenarios and other game elements are tuned to serve a particular domain and user group. By successfully applying the framework to these

contrasting settings, its domain-agnostic design and adaptability is highlighted. In both cases, requirements and design guidelines were sourced from domain experts ensuring relevance and effectiveness in real-world contexts. The selection of the specific case studies was motivated by an additional factor, the ongoing collaboration with domain experts in the relevant application areas which would allow utilizing and exploiting groups of users for real-time experimentation and evaluation, and working within a real-world environment and not a simulated one.

The development of the two games followed a short cycle of requirements analysis, which resulted in the following:

1. Speech Therapy Game for Children

Purpose: Develop in collaboration with speech therapists at the Rehabilitation Clinic of the Cyprus University of Technology and aims at supporting children with syndromes or learning difficulties, focusing on improving phonological awareness and syllable segmentation skills.

Requirements: Sourced by Speech Therapy Experts. Target Group shall be children with learning difficulties, including those with syndromes affecting phonological awareness.

Therapeutic Goals: Improve syllable segmentation and phonological processing, maintain high engagement and motivation during therapy.

Game Structure and Mechanics: The main task is to identify the correct number of syllables in words presented with colorful images, receiving immediate feedback. Correct answers should trigger positive reinforcement (e.g., happy sounds, animations). Increasing difficulty: starting with two-syllable words, advancing to more complex words and higher syllable counts as accuracy improves. Additional tasks include replacing letters to form new words and matching spoken words to corresponding images.

Design Guidelines: Shall begin with simple two-syllable words, progressing to more complex structures as mastery is demonstrated. Utilize feedback and trigger mechanisms that use positive reinforcement (sounds, animations) for correct answers; offer hints or visual cues when errors occur.

Scenarios: Support a variety of task types (syllable segmentation, letter replacement, image-word matching) to address different therapy needs.

Adaptation and Personalization: The game should dynamically adjust difficulty based on individual player performance (accuracy, response time, error rates). Also, should a child struggle, the system will provide immediate, supportive feedback and hints as needed (e.g., underlining the correct answer or splitting words into syllables). Progress shall be tracked, and metrics like accuracy and completion time shall guide real-time adjustments.

Expert Involvement: Speech therapists shall define the initial rules, thresholds, and adaptation strategies. Experts shall be able to manually adjust game parameters and review performance dashboards to monitor progress and refine the game experience.

Evaluation: Preliminary trials shall involve at least five children with various syndromes. The game shall be assessed in terms of flexibility, adaptability, and ability to reduce the need for constant manual intervention by therapists.

Data Collection: Record accuracy, response times, error rates, and progression. Also, enable therapists to review session data for individualized planning.

Performance Monitoring: Integrate real-time dashboards for therapists to monitor and adjust therapy dynamically.

2. Factory Training Game for Engineers

Purpose: Implement at the PARG poultry meat factory. Design to train engineers and factory staff in machinery operation and climate control within poultry farms.

Requirements: Target Group is factory engineers and technical staff in a Paradisiotis production environment. The training objectives are to simulate real-world machinery operation and climate control scenarios, support skill development in diagnosing faults and maintaining optimal environmental conditions.

Game Structure and Mechanics: Two main types of tasks: (i) Simulate the factory environment where users diagnose and resolve machinery issues by selecting correct images based on textual descriptions. (ii) Control the climate system: players adjust parameters such as temperature, humidity, and ventilation to maintain optimal conditions during breeding cycles.

Design Guidelines: Scenarios shall create realistic simulations of machinery faults and climate management, reflecting actual factory challenges. Furthermore, adaptation shall increase task complexity as users demonstrate proficiency that offer additional support for users who struggle or have difficulties.

Scenarios: Support a variety of task types (handle knobs, gauges, control panels) to address different training needs.

Adaptation and Personalization: The game shall increase complexity of scenarios as the player progresses (e.g., more challenging machinery faults, more nuanced climate control tasks). Real-time adaptation shall be based on performance metrics like task accuracy and completion times. The system shall provide feedback and adjust the level of challenge to match the trainee's skill and progress.

Expert Involvement: Factory managers and domain experts shall contribute to scenario design, rule definition, and evaluation of training effectiveness. Experts shall use dashboards to track user performance and adjust training parameters as needed.

Evaluation: The game shall be tested by domain experts and shift managers to provide feedback on usability, adaptability, and the relevance of training scenarios. Also, the approach shall be assessed in terms of supporting both individual and group training, adapting to various skill levels, and providing actionable performance insights.

Data Collection: Record task accuracy, completion times, and decision patterns, and enable supervisors to track individual and group progress.

Performance Monitoring: Use dashboards to visualize progress, highlight areas for improvement, and inform future training adjustments.

Two game tasks were implemented that focus on speech therapy (phonological training), while another two focus on machinery training within the poultry meat factory described earlier.

The Phonology game was designed for children with speech and language difficulties, aiming at improving their phonological awareness. The first task presented different words to players and trained them to identify the correct number of syllables. The game dynamically adjusted the difficulty of this task by increasing the number of syllables or complexity of phonetic patterns according to performance as the player progressed through the stages of the game. The second task provided textual descriptions and associated images that corresponded to those descriptions for the user to identify the correct match. The game adjusted the task's difficulty by increasing the size of a description and/or the number of accompanied images. Figure 6 presents a collage of game tasks with adjustments, flow of the game and some hints for the Phonology game.

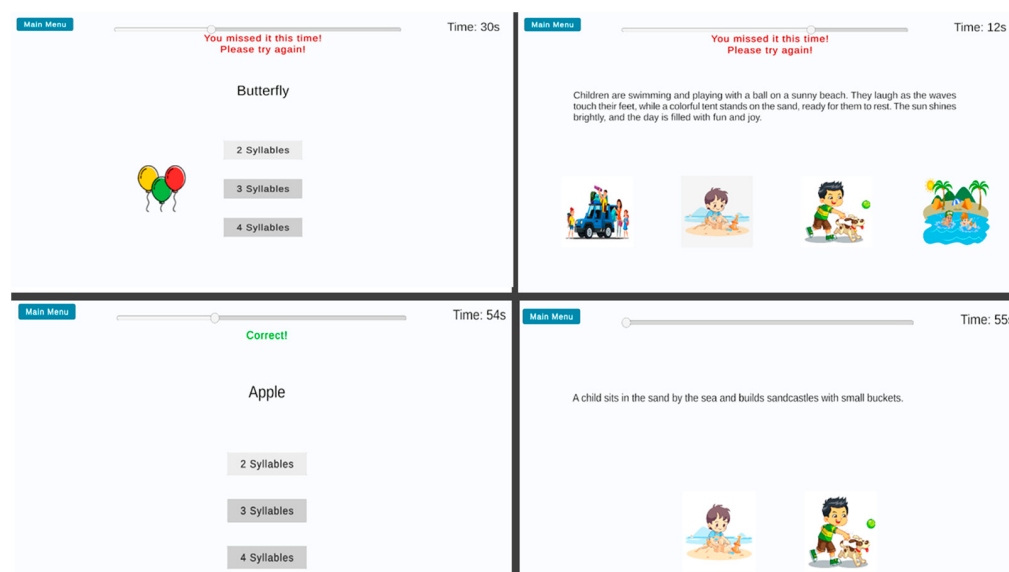


Figure 6. Speech therapy game scenes.

The Industry 4.0 training game tasks served two purposes. The first was to simulate the factory environment where users should be able to diagnose and resolve issues on the use of machinery within the production line using descriptions presented and selecting the correct images. The game increased the difficulty and granularity of the issues presented as the player advanced the scenes and required identifying and solving more complex problems. The second task focused on the factory's climate management system where players adjusted parameters like temperature, humidity and ventilation within a poultry farm during breeding cycles. The goal here was to assist trainees in learning how to maintain better climate conditions of the farm while responding to various challenges and scenarios. The game task essentially simulated the climate conditions in the farm, while training was based on modifying and adapting these conditions based on player actions and personalized learning experiences. Figure 7 presents the training game tasks of the factory with real-time adjustments and challenges scenarios.

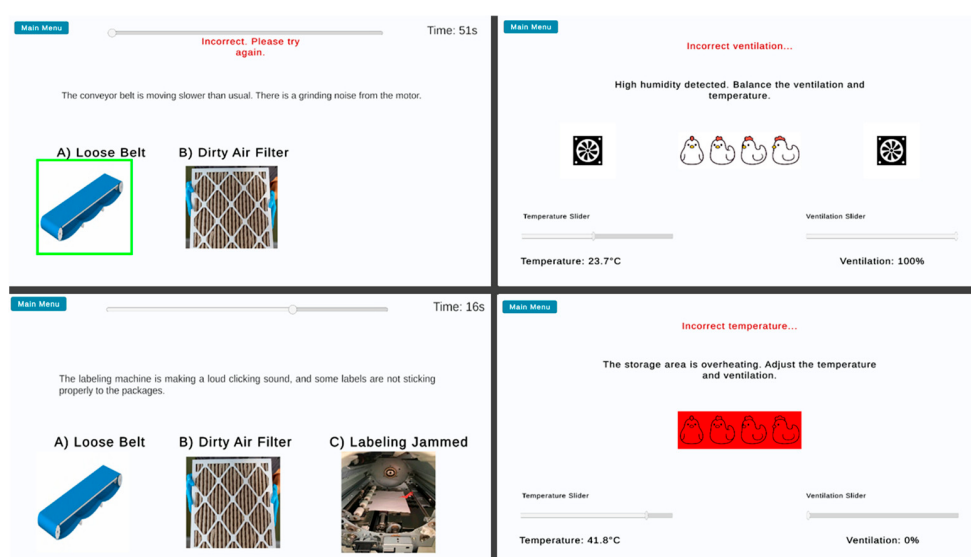


Figure 7. Excerpts of the factory training game tasks.

5.2. Evaluation Criteria

As previously mentioned, the key aspects of the present paper are the standardization of the process to create SGs, and at the same time, the dynamic, real-time adaptation and integration with DT to facilitate adjustments based on users' actions and the learning objectives. The evaluation of the process to create successful SG, as well as its educational impact and level of user engagement, is conducted in this section based on a set of criteria. The criteria used stem from the relevant literature and were adapted to reflect the properties to be assessed. Mayer et al. [31] provide a SG evaluation between educational content and gameplay mechanics. That work focuses on evaluating how well the educational objectives are integrated in the game and how effectively gameplay mechanics engage users. Pacheco-Velazquez et al. [32] presented a framework that highlights the importance of adaptive learning systems in SG. The authors explain that the effectiveness of a SG is when the game can be dynamically adjusted based on user progress and performance. This systematic review of evaluation factors in SGs highlights key aspects such as user profiling to evaluate each learning progress and tailor experiences, and real-time feedback ensuring that players receive actionable guidance throughout gameplay. Finally, Martinez et al. [33] introduced the Gaming Educational Balanced model, which integrates mechanics, dynamics and aesthetics with educational goals to ensure the game design. Their model emphasizes the balance between gameplay enjoyment and educational content. The aforementioned studies guided our evaluation process and the selection of specific criteria which were included in a specially prepared questionnaire to evaluate the two SGs developed, and the proposed approach in general. The questionnaire was distributed to domain experts, including speech therapists and factory management professionals/shift managers, and their opinions on usefulness and efficacy were recorded.

The use of a customized questionnaire and not a standardized one like the System Usability Scale (SUS) [34], Post-Study System Usability Questionnaire (PSSUQ) [35], or the Game Experience Questionnaire (GEQ) [36] as it would serve better the purposes of the evaluation study. The aforementioned approaches offer well-established questionnaires that provide valuable insights into usability, user satisfaction, and general game experience, but they do not fully capture the unique and critical aspects that the evaluation of the framework aimed at. As previously mentioned, the proposed framework emphasizes real-time self-adaptation, offers the ability to define, store and process rules via a dedicated engine, uses semantic annotation, and provides dynamic personalisation based on detailed performance metrics. Therefore, the customized questionnaire was designed to directly evaluate these specific components, focusing on the effectiveness of rule definition, adaptability based on user needs, accuracy and performance through data collection. It also allowed for the collection of a specific kind of feedback on key performance indicators. By aligning the evaluation closely with the features of the framework, it was ensured that the assessment was both relevant and actionable, providing insights that directly inform the validation of the framework's novel capabilities.

The questionnaire was designed based on the pillars mentioned above and having the goal of obtaining feedback from domain experts on the following areas: (i) *Usability*, focusing on how easy it was to understand the delivered content and utilize the supported gameplay to complete the tasks [31]; (ii) *Configuration and Rule definition*, assessing how easy and efficient was the process of defining rules and configuring game parameters and settings; (iii) *Adaptability*, that is, how well the games provided feedback and adjusted to user needs; (iv) *Accuracy and Usefulness of Results*, evaluating the proper use of performance metrics and their usefulness in assessing user progress and interactions [33]; (v) *User experience*, measuring the overall user experience including enjoyability; and (vi) *General*

satisfaction and suggestions, assessing overall impression or general feedback on the platform and recording possible suggestions [32].

The list of questions per area was as follows:

Usability of the platform

- Q1. *Was it easy to navigate in the game platform interface?*
- Q2. *Were the game settings clear?*
- Q3. *Did you find the user interface friendly and understandable?*

Configuration and Rule definition

- Q4. *Was it easy to add new rules or adjust the game parameters?*
- Q5. *Did the available options meet your needs?*

Game Adaptability

- Q6. *Were the games adapted according to diverse anticipated needs?*
- Q7. *Did the settings you made affect the gaming experience as you expected?*

Accuracy and Usefulness of results

- Q8. *Were the results collected from the games accurate and representative?*
- Q9. *Were the performance metrics (e.g., times, success rates) clear and useful for analysis?*

User experience

- Q10. *Will end-users find the games understandable and joyful in your opinion?*
- Q11. *Were the difficulty levels appropriate?*

General satisfaction and suggestions

- Q12. *Are you satisfied with the overall experience?*
- Q13. *Do you believe that the platform can be improved? If yes, how and where*

5.3. Evaluation

A Likert scale with five possible answers was utilized: “Strongly Agree”, “Agree”, “Neutral”, “Disagree”, “Strongly Disagree”. The responses of 15 domain experts were recorded after a session of 2 h using the game from the perspective of the experts and from that of a simple user. The experts were selected so as to reflect both therapeutic and industrial expertise relevant to the framework’s application areas. Of these, 9 were professionals from the therapeutic domain. This group included 3 highly experienced individuals holding PhDs or recognized as senior experts in their field, while the remaining 6 were either graduate-level specialists or students at the master’s level. The other 6 participants came from the PARG company, representing the industrial training context. This subgroup included 2 shift managers, each with around 10 years of experience in the company, and 4 Electrical Engineers. Among the engineers, 2 had similar experience (~8–10 years), while the other 2 were newer to the organization, with approximately 2 years prior experience in other industries. This mixture of senior and junior staff ensured that feedback collected a variety of perspectives, from long-term operational knowledge to new viewpoints from newer employees. This diverse participants’ profile provided a strong basis for evaluating the framework’s usability, adaptability, and relevance across both therapeutic and industrial training aspects.

These responses provided valuable insights in terms of the usability, adaptability and overall outcomes of the game platform (see Table 1). In general, the results suggested a highly positive reaction to the platform, with strengths in terms of usability and how well the game adapts to multiple players’ needs. Most of the participants expressed their satisfaction with the ease of use, which is one of the most important findings. Almost

everyone who used the platform found its steps clear and guiding, while the game parameters were easy to understand and set. This suggests that the platform design satisfies the requirements for ease of use and accessibility. Feedback regarding the adjustment of games was also positive. Most experts indicated that the games met the original needs, and they liked the fact that gameplay and challenging stages were modified in real-time according to scenarios they tested which varied the skills, knowledge and competences of the users/players. The majority of the experts also expressed the opinion that players will receive full gaming experience (e.g., navigation, audio and graphics support, hints), but, most importantly, acknowledged how well the games performed self-adjustments to customize certain features of the tasks. However, some experts believed that creating new rules or changing settings was a little challenging. There is certainly room for improvements to make the platform simpler in terms of setting rules (e.g., selecting rules from a pre-defined list, browsing parameters and changing their values graphically through gauges, inspecting current rules without being forced to change their parameters, etc.).

Table 1. Number of answers to the evaluation questionnaire received per scale.

Question	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
Q1	6	7	2	0	0
Q2	10	4	1	0	0
Q3	12	2	1	0	0
Q4	2	7	3	3	0
Q5	7	3	3	2	0
Q6	9	5	1	0	0
Q7	3	6	4	2	0
Q8	11	4	0	0	0
Q9	10	5	0	0	0
Q10	9	4	1	1	0
Q11	7	6	2	0	0
Q12	11	3	1	0	0
Q13	1	8	5	1	0

The accuracy and reliability of the performance metrics were highly rated. Moreover, the experts found the results of the games descriptive, accurate and helpful in monitoring progress. This was characterized as very significant in evaluating educational games based on performance data as it can be utilized by both the experts and the trainers that are monitoring users' progress, allowing the assessment of how users are performing thus far and making better decisions on what to do next during the learning process.

The overall user experience assessment was generally positive, but some participants suggested improvements in user instructions and difficulty levels. Some experts found the configuration of scenarios difficult and felt it needed more fine-tuning, while others thought it was sufficient. The platform adaptability to user needs suggested how to handle (increase) complexity for making training in the factory more challenging. Overall, users proposed new features that could be incorporated in upcoming updates (e.g., allowing experts to adjust training scenarios dynamically through a visual editor, adding guided directions for first-time users, enabling the system to recommend difficulty adjustments based on long-term performance trends, etc.). For the platform to effectively meet the needs of domain experts and players, this input was considered essential.

The feedback collected is analyzed graphically in Figure 8. Figure 8a shows the questions that received the highest/lowest responses in total, which gives a clear picture of the overall trend across questions. Figure 8b displays how evaluators felt about each aspect of the platform.

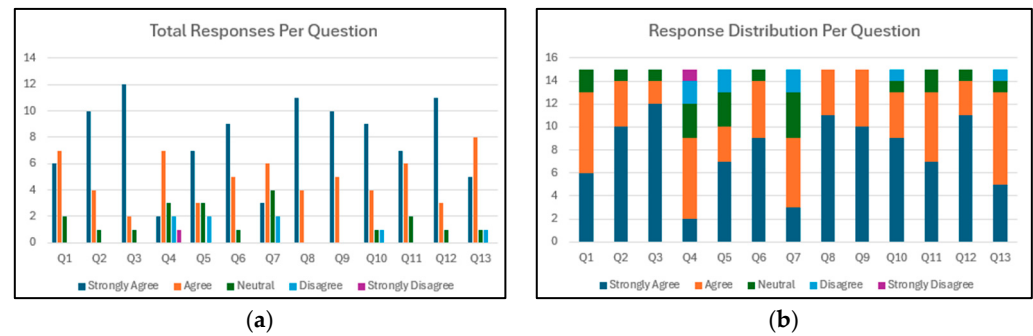


Figure 8. (a) Total number of responses across all questions; (b) Stacked bar chart on distribution of responses.

In conclusion, the preliminary evaluation results indicate that the framework is promising with respect to adaptability and usability. Nevertheless, it is important to note that the pilot results reported may be considered as preliminary evidence rather than proof of efficacy. To this end, additional large-scale and long-term user studies are required to fully assess the framework's efficacy, scalability and generalization. The use of a DT for real-time self-adjustment appears to be a strong advantage, creating an efficient process for a personalized and dynamic environment. Still, there is ample room for improvement on scenarios configuration and on enhancing the difficulty adjustment.

5.4. Comparison with Other Approaches

Various frameworks or approaches have been proposed across the literature to streamline development and ensure effectiveness. Table 2 provides an overview of similar approaches analyzed in terms of their focus, key features and limitations, juxtaposed with the proposed framework.

Table 2. Overview of similar approaches.

Paper/Framework	Main Focus/Domain	Key Features	Limitations
Hocine et al. [24]	Upper-limb rehab games	Dynamic difficulty adaptation, modular server	Single domain, not generalisable
Saeedi et al. [37]	Speech therapy for children	User-centered design, usability evaluation	Language and disorder-specific, not adaptable
Antunes et al. [10]	Therapy (multi-domain)	Modular DT platform, multi-game/sensor integration	Focus on therapy, less on hybrid expert automation
Peery [38]	Conceptual/methodological overview	Theory, taxonomy, adaptivity models	Not an implementable framework
Calvo-Morata et al. [11]	Bullying/cyberbullying prevention	Systematic review, mechanics analysis, target groups	No framework, not technical
Amengual Alcover et al. [25]	Motor rehab therapy	Process framework, validated in balance therapy	Rehab-focused, not cross-domain
Antunes & Madeira [26]	Therapy, multi-domain (PLAY)	Model-based DT platform, web-based configuration, therapist tools	Clinical orientation, less focus on industrial/other
Proposed Framework	Domain-agnostic (education, industry, therapy)	DT-driven real-time adaptation, rule engine (driven-experts rules), semantic standardisation, demonstrated in two different fields	Highlight applicability, hybrid adaptation, strong reproducibility (domain-agnostic)

This section provides a short comparison of the proposed framework with four of the approaches listed in Table 2, which were selected based on the level of similarity to the proposed approach. The selection was made having in mind primarily that an approach should aim at offering support for serious games development and at the same time provide features such as adjustment and user performance evaluation. Therefore, the list of approaches selected may not by all means be regarded as full or exhaustive, but instead it represents a good sample of available approaches to compare with. The approaches considered were the following: (i) Hocine et al. [24] that present a dynamic difficulty adaptation (DDA) technique to enhance stroke rehabilitation outcomes through serious games. Their approach customizes game levels and adjusts difficulty based on patients' motor abilities and performance. The study evaluates this technique using PRehab, a serious game designed for upper-limb rehabilitation. The findings highlight DDA's ability to balance challenge and effort, making it a promising tool for stroke rehabilitation programs. Their approach is domain-specific, centered on physical rehabilitation, and primarily validated in a single context.; (ii) Saeedi et al. [37], who developed "Ava", a smartphone-based serious game designed to assist speech therapy for preschool children with speech sound disorders (SSD). The game teaches consonants, syllables, words, and sentences through four interactive levels. Results showed high satisfaction among speech-language pathologists and positive feedback from children. The game demonstrated potential as a tool for home-based therapy under parental supervision. The work emphasizes user-centered design and usability, but is still specific to speech therapy and does not address any adaptability or domain transfer; (iii) Alcover et al. [25], that introduce PROGame, a structured framework for developing serious games aimed at motor rehabilitation therapy. The framework integrates agile methodologies (Scrum), web application development principles, and clinical trial processes to ensure systematic and validated game development. PROGame demonstrates potential for broader use in rehabilitation game development. The model is process-oriented and repeatable within rehabilitation contexts, but its application and validation are limited to motor therapy; (iv) Antunes and Madeira [26], who introduce the PLAY platform, a model-driven framework for designing serious games tailored to children with special needs. It focuses on physiotherapy and cognitive rehabilitation by gamifying therapeutic exercises into structured levels and actions. The platform integrates patient profiles, enabling personalized game recommendations based on therapeutic needs and progress. Therapists can monitor performance and adapt exercises in real time, while data analytics tools support decision-making. The PLAY platform is modular and domain-agnostic, but focuses only on therapy and clinical settings.

The comparison was performed using the following set of features:

- (a) Development guidance: This refers to a structured, step-by-step framework for developing game scenarios and tasks, gameplay logic, and user interaction mechanisms.
- (b) Standardization: Corresponds to the use of standardized data formats and interfaces for interoperability (e.g., using XML/JSON for content definition and system integration) so that games become portable and playable on different platforms, hardware or operating systems.
- (c) Adaptability: This feature evaluates if this property is offered, how and where assessing also whether the game produced may be self-adjusted or not. It basically describes the system's ability to modify game parameters dynamically based on player performance or context.
- (d) Scaling capabilities: Assesses the system's ability to scale in complexity, levels, user types, user volume and content volume without degradation in performance or usability.

- (e) DT integration: This corresponds to the ability to integrate the game with a DT and enjoy the benefits of controlling the game environment through simulations prior to making adjustments or any other form of DT support.
- (f) Performance Assessment: Relates to the assessment of users playing the game, including how (i.e., the metrics) and where this is performed, possibly including domain expert judgment.
- (g) Data Management: Concerns how data (e.g., user logs, performance data) is collected, stored, organized and used within the game system, including the description of the game environment.
- (h) Domain Agnosticism: This feature assesses whether the approach is tightly connected to a specific application domain or if it is able to generalize game development across domains.
- (i) AI integration/Prediction: Evaluates the ability to make use of Artificial Intelligence or Machine Learning to provide predictions and use them in the game in some way.
- (j) Evaluation: This feature differs from performance assessment as it relates to the evaluation of each approach (if any), the evaluation scale and its characteristics.
- (k) Game Production: This feature deals with how the game is produced as an end-result with the possible ways including automatic, expert based, or hybrid.

Table 3 summarizes the comparison findings. In the following, a discussion is provided that particularly emphasizes the strengths and advancements brought by the proposed approach, while drawing meaningful comparisons with the four representative studies selected from prior literature.

(i) *Development Guidance*

A well-defined development methodology is crucial for producing reliable and effective serious games. All five approaches, including the proposed one, employ a structured or phased methodology. While Hocine et al. [24] and Saeedi et al. [37] include phased guidance with usability testing and step-by-step development, our approach distinguishes itself by seamlessly integrating both these elements, delivering a clearly phased, modular structure that ensures adaptability and continuous feedback from users and experts.

(ii) *Standardization*

Standardization is a key enabler of interoperability and cross-platform functionality. While all approaches incorporate some form of standardization, ranging from procedural content generation [24] to JSON-based data exchange [38], our approach stands out through its comprehensive use of both XML and JSON to define rules, game logic, scenarios, metrics and thresholds. While supporting portability and modular updates in our framework is not claim that these represent full standardization as defined by industry-wide standards or public. This provides a structured way to manage and reuse game elements, which can ease future integration or migration to standards if needed.

(iii) *Adaptability*

As previously mentioned, one of the most impactful features in modern serious games is adaptability, that is, the ability to dynamically adjust game difficulty and content. All reviewed approaches offer this to varying degrees. However, our approach delivers self-adaptive behavior, reacting to real-time user performance to alter difficulty and pacing. Unlike in the case of Saeedi et al. [37] which focus on children's progression, or the work of Antunes and Madeira [26] that offers real-time personalization, our approach generalizes adaptability across user profiles and scenarios, making it more versatile.

(iv) *Scaling Capabilities*

Scalability is often overlooked but essential when expanding usage across institutions or user demographics. While only Hocine et al. [24] and Antunes and Madeira [26] highlight scalability, our approach delivers two-way scaling, both in game content (scenes and tasks) and in user management (simultaneous users or independent sessions). This dual scaling capacity ensures that the game system can evolve alongside growing or diversifying user bases without performance trade-offs.

(v) *Digital Twin (DT) Integration*

A key differentiator of our approach is its integration of DT technologies, a feature absent in all four other models compared. This allows for real-time simulations and environment modeling before applying changes, enabling predictive testing and adaptive gameplay adjustments. The DT component significantly enhances system reliability and personalization, offering a cutting-edge advantage in serious game development.

(vi) *Performance Assessment*

Monitoring user performance is integral to evaluating game effectiveness. All five approaches include performance assessments using varied methods, from clinical metrics [25,27] to usability surveys [37]. Our approach offers a multi-layered system of metrics (accuracy, error rates, completion time) and projects them on dedicated expert dashboards, merging real-time analytics with domain expert feedback for a more subtle view of user engagement and progress.

(vii) *Data Management*

Data collection, organization, and reuse are vital for continuous improvement and research. While the compared models like Hocine et al. [24] and Saeedi et al. [37] seem to lack clear data strategies, our approach presents a more concrete data management component through Data Lakes and semantic annotation, enabling long-term storage, contextual retrieval, and structured data reuse. This robust infrastructure supports scaling, personalization and offline analysis.

(viii) *Domain Agnosticism*

All four of the compared approaches are domain-specific: Hocine et al. [24] targets stroke patients; Saeedi et al. [37], speech therapy; Alcover et al. [25], motor rehabilitation; and, Antunes and Madeira [26] therapeutic games for children. In contrast, the proposed approach is domain agnostic, capable of adapting to various contexts from education to rehabilitation and simulation as demonstrated in the two sample games developed for experimentation. This flexibility is crucial for wide-scale adoption and reuse across disciplines.

(ix) *AI Integration and Prediction*

AI and machine learning are powerful tools for personalization. Only Saeedi et al. [37] and Antunes and Madeira [26] report integration with these technologies, primarily for speech recognition and adaptive feedback. While our approach does not yet embed predictive models, it is designed to accommodate future AI integration, particularly through its modular and data-rich infrastructure. This makes it ready for next-generation predictive enhancements.

(x) *Evaluation*

Evaluation of the reviewed approaches typically rely on performance-based assessment with users experiencing the game and collecting metrics. All reviewed approaches report such an evaluation, with our approach at present being evaluated only via a custom expert questionnaire and with limited user engagement. The questionnaire enabled focused and relevant feedback from domain professionals, while the small group of children that experienced the speech therapy game is of course

lacking compared to extensive feedback with user trials. The latter is already planned, and the first phase will involve testing the phonetic game with users (children) in the Rehabilitation Clinic of the Cyprus University of Technology. Feedback collected through experts, though, may be considered also very critical as it supports iterative refinement during early stages of deployment.

(xi) *Game Production*

All reviewed approaches follow a hybrid production method, balancing automated generation with expert input. Our approach follows the same path, benefiting from expert-curated content while incorporating automation in game logic and content generation and aligning with best practices in the field.

Concluding this brief comparison, we may argue that the proposed approach embodies a forward-looking framework that bridges traditional development with cutting-edge technologies like DT, data storage architectures and semantic data systems. It preserves the proven strengths of other methodologies, that is, structured guidance, adaptability, hybrid development, while setting itself apart through scalability, interoperability, and domain flexibility. Although some of the other approaches contribute valuable domain-specific insights and integration methods, our approach represents a robust and adaptable alternative for the evolution of serious game development.

Table 3. Synopsis of the approaches compared against a set of features.

Features	Hocine et al. [24]	Saeedi et al. [37]	Alcover et al. [25]	Antunes and Madeira [26]	Our Approach
Development guidance	YES, phased approach with usability tests	YES, follows a step-by-step approach	YES, follows structured development with phases	YES, follows a model-driven framework	YES, phased approach, step-by-step
Standardization (means)	YES, procedural content generation is used to create adaptable game levels for reusability across different environments	YES, uses structured game design documents (GDD) and integrates speech recognition systems for systematic feedback	YES, Kinect-based tracking modules	YES, uses JSON for data exchange	YES, uses XML and ontologies for rules, scenarios, game tasks, metrics, thresholds
Adaptability (how, where)	YES, dynamic difficulty adjustment adapts difficulty based on patient performance	YES, adapts difficulty levels based on children progress and abilities	YES, adjust interaction elements and game parameters	YES, real-time data to personalize games and adjust parameters	YES, self-adapts based on performance, changes difficulty levels
Scaling capabilities	YES, scalable in terms of game levels	NO	NO	YES, scalable across multiple users and clinics contexts	Yes, two-way: (i) game scenes and tasks, (ii) number of users (simultaneous or not)
DT integration	NO	NO	NO	NO	YES, integrates DT for real-time adaptation and personalized gameplay adjustments

Table 3. Cont.

Features	Hocine et al. [24]	Saeedi et al. [37]	Alcover et al. [25]	Antunes and Madeira [26]	Our Approach
Performance Assessment	YES, through metrics like task success rates and distance covered	YES, using Nielsen’s principles and PSSUQ	YES, through clinical studies measuring improvements in balance	YES, through game results and patient progression tracking	YES, evaluates player progress using metrics like accuracy, error rates, and completion times; includes dashboards for expert analysis
Data Management	NO	NO	YES, data storage modules	Yes, using a repository for storing all patient data	YES. Through Data Lakes and semantic annotation
Domain Agnosticism	NO, only for stroke patients	NO, only for speech therapy	NO, designed for motor rehabilitation therapy	NO, focuses on therapeutic games for children	YES, demonstrates applicability across different domains
AI integration/Prediction	NO	YES, uses speech recognition technology (Google)	NO	YES, uses ML for adaptation	NO, potential of integration with analytical models
Evaluation (scale, performance based, other characteristics)	Performance-based evaluation using statistical analysis	Performance-based evaluation through usability testing	Performance-based evaluation through clinical studies	Performance-based evaluation using patient progression	Short-scale (experts), custom questionnaire for expert assessment
Game Production (automatic, expert based, hybrid)	Hybrid	Hybrid	Hybrid	Hybrid	Hybrid

5.5. Discussion

The proposed framework addresses key challenges that are described in Section 2. By integrating semantic annotation, expert-defined rules and feedback loops personalized, and dynamically adjustable gaming experiences are enabled that respond in real-time to user interactions. This paper has shown that real-time feedback mechanisms enhance player motivations and learning outcomes on customized levels of difficulty rather than following a fully dynamic adaption [18]. This is the major advancement over traditional SGs, which typically work based on predefined difficulty levels or periodic updates rather than continuous adaptation.

One of the key contributions of this framework is the introduction of an internal standardization process for adaptive SG making them scalable and reusable across different domains. This process formalizes SG development using ontologies and metadata for the description of game elements, scenarios, scenes, goals and gameplay. Standardization is therefore framework-specific and does not target at providing compliance with external standards for game development or educational technology. Our goal was mainly to provide better integration and consistency within our framework and facilitate portability. The concept of DT in manufacturing [18] has demonstrated the scalability of digital models for production environments, but their application in SG and adaptive learning remains underexplored. Similarly, the DT framework for personalized therapy [10] focuses on individual patient rehabilitation, but lacks a domain-expert approach. Moreover, the flexibility of the framework was demonstrated in two different domains: (i) Speech therapy

for children with learning difficulties, where phonological training tasks are adjusted based on performance data, and (ii) Industrial training for manufacturing environments, where machine operation scenarios are simulated at a poultry meat factory with dynamic difficulty scaling. These applications showcase the framework's adaptability and potential, demonstrating that it is not limited to one specific domain and that it can be applied to a range of training and educational contexts. Nevertheless, as the current evaluation is limited to small-scale, short-term demonstrations, future work will focus on large-scale and long-term experiments in order to fully evaluate the framework's efficiency, scalability and impact on a range of user groups. The mechanisms provided by the framework for setting up and creating games are modular, allowing developers to target different game genres and styles without difficulty.

The emphasis of the proposed framework on data-driven personalization is also one of its strongest points. Many of the SGs that are currently available either do not have adaptive features or offer only manual configuration of difficulty levels. Studies on speech therapy games [18] show that static difficulty levels can lead to disappointment and disengagement among children and research on goal-setting in pediatric rehabilitation [26] emphasizes the need for structured yet adaptable learning objectives. The proposed approach addresses these concerns by leveraging DTs to dynamically adapt gameplay to real-time metrics performance while continuously evaluating user interactions. Additionally, expert guidance supports the development of adaptation rules ensuring that the modifications made to the game are in full alignment with domain-specific learning objectives, keeping educational value and player engagement. It is important to note that our framework balances between automation and human intervention. It may be considered a hybrid solution that brings together the strengths of both worlds, that is, expert input and automated processes, allowing us to maintain quality and relevance while enabling scalability and flexibility. Nevertheless, the structure of the framework supports also modifying the levels of contribution between expert support and automation, even to the extreme ends with only experts being user, or only automated rules, the latter being directed by generative AI models.

Despite the many advantages, there are still specific issues that need to be investigated further. Scalability is one topic that needs more work, especially when managing large-scale usage with multiple users. Although the challenge of managing data produced at high speeds from many different sources (users) is already addressed in the framework via the use of Data Lakes and semantic annotation, large-scale real-time adaptation and high-volume data processing needs further analysis, with possible study of computational optimizations. Furthermore, adding machine learning methods could improve decision-making autonomy.

Another possible area for improvement is the incorporation of deep analytics on player behavior and engagement to discover possible patterns and trends that could be incorporated in the rule-engine and be treated as "ground-truth" for deciding in real-time if, when and where adaptations should take place. In addition, exploring AI-driven content generation may provide further automatic adaptations and support deeper structural changes.

Summing up, this work presented a standardized game development process and an adaptive learning environment that offer an innovative and practical addition to the field of SGs. This framework provides a strong foundation for future developments in adaptive gaming, education and DT-enabled training systems.

6. Conclusions and Discussion

6.1. Summary

This paper introduced a framework that supports a phased approach for the development of SG. Its major components include setting the game's environment through the

description of scenes, tasks and gameplay, and the assessment of user interactions via a set of metrics which are stored in user profiles. The framework facilitates the utilization of experts and the definition of rules that guide adjustments performed based on user history. It also integrates DT into SG, enabling real-time self-adaptation tailored to players' needs. Through this integration, the framework describes capabilities to provide a personalized and dynamic learning or training environment, where scenario difficulty and game objects adjust according to user profiling, performance metrics, and domain experts' input. This approach was demonstrated in two different case studies, a therapeutic game for children with learning difficulties and a factory training simulation. Both examples highlighted the framework's features and effectiveness in improving engagement and learning outcomes, confirming its value in very different domains.

The proposed framework standardizes the process for building SG by offering meta-data semantic annotation and ontologies for both the game environment (including scenarios, scenes, tasks, gameplay elements, etc.) and player data (such as performance metrics). It also defines rules for performing automatic adjustment based on expert consultation. These rules are parsed by the DT, which, with the support of the framework's analytical models that process related information and scores (e.g., user interactions, accuracy, and errors), performs controlled simulations to create a dynamic and self-adaptive environment.

The evaluation of the framework, which is based on well-defined criteria pillars, confirmed that real-time adaptability and ease of use are key strengths. Most experts are satisfied with the platform's ability to adjust to users' needs and suggested further improvements on providing a simple process for rule insertion and definition of difficulty level adjustments. A short comparison with similar approaches met in the literature suggested that the proposed framework differentiates by offering essential features necessary for the successful development of SGs, including scalability, domain flexibility, and strong personalisation. Therefore, it may be considered an innovative solution that addresses current challenges in the SG area and facilitates future integration with artificial intelligence and other modern technologies, based on its evolutionary (modular) approach, flexibility, and structured data processing.

6.2. Limitations and Future Work

The proposed framework was designed using a modular architecture so as to be able to address current and future SG challenges. Nevertheless, due to size limitations on one hand, and the fact that the aim of this paper was to introduce, explain and evaluate the proposed framework on the other hand, it was not feasible to address multiple issues and elaborate on them. Below we provide a short discussion on some limitations of the current work and how these may be addressed in the future.

The framework offers the ability to integrate physiological signals such as heart rate, skin conductance and EEG, for strengthening personalization. Nevertheless, these biosignals were not part of the case studies presented as the main target of experimentation was to showcase the applicability and efficiency of the framework in real-world settings. It should also be noted that operational and technical constraints, like complexity, requirements for real-time signal acquisition and ethical parameters related to privacy and data protection, particularly for groups such as children, make it more challenging and time consuming to implement. Biosignals offer critical insights into user emotional and cognitive states and are expected to significantly strengthen the real-time adaptations.

Another challenge is the small scale of experimentation and the short duration of the evaluation activities. The experimental phase primarily aimed to demonstrate feasibility and collect expert feedback using brief evaluation involving a limited number of users. As such, the framework's performance under long-term usage, across different user groups

and different contexts, remains unexplored. This restricts our ability to provide generalized conclusions about user engagement, learning outcomes, or therapeutic effectiveness. Larger-scale studies are required to validate the adaptability and educational impact of the system under real-world operational conditions.

The framework is designed to be domain-agnostic, and this was demonstrated in two different application domains (therapy and industrial training). Extending it to more complex contexts or domains introduces additional challenges, especially in cases which can be considered as “complex”. Theoretically, in the context of serious games, a complex environment refers to an environment characterized by the following:

- Multiple interconnected elements and/or layers (scenes, mechanics, systems);
- Realism, authenticity, or a multifaceted representation of aspects of reality;
- Variety in interaction methods, choices, and problem-solving approaches;
- Continuous feedback, interaction, and adaptations based on the player’s actions;
- Player engagement in situations where consequences are not immediate but are influenced by multiple factors (emergent gameplay, systems thinking).

The integration of heterogeneous resources and datasets keep real-time adaptability in new scenarios and align semantic representations across areas. Erkoyuncu et al. [22], highlights the complexity of merging DT architectures with “brownfield” systems and the need for continuous updates and compatibility across heterogeneous infrastructures. Our ontology-driven and rule-based design, as well as the ability to collect, store and process performance data, helps address some of these challenges, facilitating the successful transfer to new domains, with possible needs for domain-specific customization, improvement of the adaptation logic and evaluation of ethical or legal limits in every scenario.

The rule-based adaptation engine is currently based on expert-knowledge rules using structured formats (e.g., XML), supported by semantic annotations. While this offers transparency and consistency, it limits flexibility in complex environments that demand rapid reaction to unexpected or evolving data. Especially at a large scale where the volume of user interactions changes unexpectedly, the rule engine may face challenges in processing real-time performance and personalization data, as in the case with all big data applications. To achieve self-adaptation in different and complex environments, the framework must evolve towards dynamic rule learning, more autonomous processing and enhanced support for data-driven decision-making models.

Future work will also focus on extending the framework’s application to additional domains while refining adaptability mechanisms and optimizing user profiling capabilities. Standardization efforts will also continue, with development of formal methodologies for collecting and storing gameplay data from diverse sources and types, thereby improving the consistency and robustness of analytics. The integration of advanced AI and machine learning techniques is planned to deliver features such as predictive user behavior modeling and dynamic learning curve adjustment, thereby enabling more nuanced personalization. Subsequent phases will target improved scalability via AI-driven adaptation, accompanied by long-term, large-scale empirical validation across varied populations. Additionally, emerging AI technologies, including generative AI and large language models (LLMs), will be explored as tools to accelerate and enhance expert-driven rule creation and management. Depending on results, LLMs may ultimately support fully automated or hybrid rule generation approaches leveraging embodied domain knowledge. Moreover, large-scale deployments with many simultaneous users will be conducted to rigorously assess system impact over extended periods. Finally, the integration of biosignal analysis as an active input stream for real-time adaptation will be advanced to enable precise monitoring of user states and implementation of novel dynamic adjustment strategies, particularly in therapeutic and stress-management game scenarios.

Author Contributions: Conceptualization, S.L. and A.S.A.; methodology, S.L. and A.S.A.; software, S.L.; validation, S.L. and A.S.A.; formal analysis, S.L.; investigation, S.L.; resources, S.L. and A.S.A.; data curation, S.L.; writing—original draft preparation, S.L.; writing—review and editing, S.L. and A.S.A.; visualization, S.L.; supervision, A.S.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SG	Serious Game(s)
DT	Digital Twin(s)
ADT	Adaptive Digital Twin
DL	Data Lakes
VR	Virtual Reality
AR	Augmented Reality
XML	eXtensible Markup Language

Appendix A

```
<?xml version="1.0" encoding="UTF-8"?>
<game>
  <scenarios>
    <scenario id="1" name="Memory Challenge"
      description="Memory training using puzzles">
      <scenes>
        <scene id="1" name="Puzzle Solving"
          description="Solve a memory puzzle to proceed."
          complexity="medium">
          <handles>
            <handle id="1" type="click"
              action="match items"/>
            <handle id="2" type="listen"
              action="repeat the phonetics"/>
          </handles>
          <levels>
            <level id="1" type="visual"
              difficulty="easy" objects="5" moving="false"/>
            <level id="2" type="visual"
              difficulty="medium" objects="8" moving="true"/>
            <level id="3" type="visual"
              difficulty="hard" objects="10" moving="true" audio="true"/>
          </levels>
        </scene>
      </scenes>
    </scenario>
  </scenarios>
</game>
```

Figure A1. Description of the game environment in XML.

```

<game>
  <scenarios>
    <scenario id="1" name="Memory Challenge" description="Memory training using puzzles">
      <scenes>
        <scene id="1" name="Puzzle Solving"
          description="Solve a memory puzzle to proceed." complexity="medium">
          <handles>
            <handle id="1" type="click" action="match items"/>
            <handle id="2" type="listen" action="repeat the phonetics"/>
          </handles>
          <levels>
            <!-- Initial Level -->
            <level id="1" type="visual" difficulty="easy" objects="5" moving="false"/>

            <!-- Next Level, will be adapted dynamically by DT -->
            <level id="2" type="visual" difficulty="medium" objects="8" moving="true"/>

            <!-- Harder Level -->
            <level id="3" type="visual" difficulty="hard" objects="10"
              moving="true" audio="true"/>
          </levels>
        </scene>
      </scenes>
    </scenario>
  </scenarios>

  <!-- Digital Twin Information -->
  <dt>
    <dt_id>12345</dt_id>
    <current_state>tracking</current_state>
    <player_profile>
      <player_id>12345</player_id>
      <memory_skill_level>medium</memory_skill_level> <!-- Player's performance data -->
      <accuracy>0.85</accuracy> <!-- Percentage of correct actions -->
      <completion_time>4.5</completion_time> <!-- Task completion time -->
    </player_profile>

    <!-- Self-Adaptive Adjustment Rules -->
    <adjustment_rules>
      <!-- If player performs well in Level 1, move to Level 2 -->
      <rule>
        <!-- Example condition -->
        <condition>completion_time < 5.0 and accuracy > 0.80</condition>
        <action>adapt_level</action>
        <affected_level>2</affected_level>
        <new_type>listen</new_type> <!-- Change type to 'listen' -->
        <new_objects>10</new_objects> <!-- Increase objects to 10 -->
      </rule>

      <!-- If player struggles in Level 2, move to easier settings -->
      <rule>
        <condition>completion_time > 5.0 and accuracy < 0.60</condition>
        <action>decrease_level</action>
        <affected_level>2</affected_level>
        <new_type>visual</new_type> <!-- Change back to 'visual' -->
        <new_objects>8</new_objects> <!-- Adjust objects to 8 -->
      </rule>
    </adjustment_rules>
  </dt>
</game>

```

Figure A2. Adaption rules in XML.

```

# Example: Player performance tracking and adaptation rules
# Sample player profile (from the Digital Twin)
player_profile = {
    "completion_time": 4.5,      # in minutes
    "accuracy": 0.85,           # percentage
    "level": 1,                 # current level
    "objects": 5,               # number of objects in the puzzle
    "type": "visual",           # game type (visual)
}

# Function to check adaptation conditions
def check_adaptation_rules(player_profile):
    # Rule 1: If completion time < 5 minutes and accuracy > 80%,
    # adapt to level 2 with 'listen' type and 10 objects
    if player_profile["completion_time"] < 5.0 \
        and player_profile["accuracy"] > 0.80:
        adapt_level(player_profile, 2, "listen", 10)

    # Rule 2: If completion time > 5 minutes and accuracy < 60%,
    # decrease difficulty and revert to visual type with 8 objects
    elif player_profile["completion_time"] > 5.0 \
        and player_profile["accuracy"] < 0.60:
        adapt_level(player_profile, 2, "visual", 8)

# Function to apply level adjustments
def adapt_level(player_profile, new_level, new_type, new_objects):
    player_profile["level"] = new_level
    player_profile["type"] = new_type
    player_profile["objects"] = new_objects

# Example call to check and apply adaptation rules
check_adaptation_rules(player_profile)

```

Figure A3. Pseudocode for tracking player performance and rules for adaption in real-time.



```
<?xml version="1.0" encoding="UTF-8"?>
<player_metadata>
  <player>
    <player_id>12345</player_id>
    <game_details>
      <game_id>6789</game_id>
      <scenario_id>1011</scenario_id>
      <scene_id>1213</scene_id>
    </game_details>
    <session>
      <timestamp>2024-12-07T14:30:00</timestamp>
    </session>
    <data_type>
      <structured>numerical</structured>
      <unstructured>voice</unstructured>
      <semi_structured>text</semi_structured>
    </data_type>
    <task>
      <task_type>puzzle</task_type>
      <difficulty>medium</difficulty>
    </task>
    <performance>
      <score>85</score>
      <time_taken>10:15</time_taken>
    </performance>
    <player_attributes>
      <age>25</age>
      <experience_level>intermediate</experience_level>
    </player_attributes>
  </player>
</player_metadata>
```

Figure A4. Semantic annotation and metadata in XML for describing game experience data.

References

1. Krath, J.; Schürmann, L.; von Korfflesch, H.F.O. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. *Comput. Hum. Behav.* **2021**, *125*, 106963. [CrossRef]
2. Ordaz, N.; Romero, D.; Gorecky, D.; Siller, H.R. Serious Games and Virtual Simulator for Automotive Manufacturing Education & Training. *Procedia-Procedia Comput. Sci.* **2015**, *75*, 267–274. [CrossRef]
3. Ogunsakin, R.; Mehandjiev, N.; Marin, C.A. Towards adaptive digital twins architecture. *Comput. Ind.* **2023**, *149*, 103920. [CrossRef]
4. IEEE P2948; Standard for Framework and Definitions for Cloud Gaming. IEEE Standards Association: Piscataway, NJ, USA, 2020. Available online: <https://standards.ieee.org/ieee/2948/10379> (accessed on 16 June 2025).
5. IEEE P3341; Recommended Practice for Mobile Game Experience Indicators and Evaluation Criteria. IEEE Standards Association: Piscataway, NJ, USA, 2023. Available online: <https://standards.ieee.org/ieee/3341/11184/> (accessed on 25 June 2025).
6. IMS Global Learning Consortium. Learning Tools Interoperability (LTI) Core Specification 1.3. Available online: <https://www.imsglobal.org/spec/lti/v1p3> (accessed on 25 June 2025).
7. IMS Global Learning Consortium. Content Packaging (CP) Specification. Available online: https://www.imsglobal.org/content/packaging/cpv1p1p3/imsdp_infv1p1p3.html (accessed on 25 June 2025).
8. Advanced Distributed Learning Initiative. SCORM Explained. Available online: <https://scorm.com/> (accessed on 25 June 2025).
9. Cuevas-ortuño, J. Serious Games or Challenge-based Learning—A comparative analysis of learning models in the teaching of lean manufacturing. In Proceedings of the 2020 IEEE Global Engineering Education Conference (EDUCON), Porto, Portugal, 27–30 April 2020; pp. 1542–1549.
10. Antunes, A.; Madeira, R.N.; Postolache, O. Digital Twin Framework for Personalized Serious Games-Based Therapy. In Proceedings of the 2024 International Symposium on Sensing and Instrumentation in 5G and IoT Era (ISSI), Lagoa, Portugal, 29–30 August 2024; Volume 1, pp. 1–6. [CrossRef]
11. Calvo-Morata, A.; Alonso-Fernández, C.; Freire, M.; Martínez-Ortiz, I.; Fernández-Manjón, B. Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review. *Comput. Educ.* **2020**, *157*, 103958. [CrossRef]

12. Damaševičius, R.; Maskeliūnas, R.; Blažauskas, T. Serious Games and Gamification in Healthcare: A Meta-Review. *Information* **2023**, *14*, 105. [\[CrossRef\]](#)
13. Kara, N. A systematic review of the use of serious games in science education. *Contemp. Educ. Technol.* **2021**, *13*, ep295. [\[CrossRef\]](#)
14. Klein Haneveld, L.; Kip, H.; Bouman, Y.H.A.; Weerdmeester, J.; Scholten, H.; Kelders, S.M. Exploring the added value of virtual reality biofeedback game DEEP in forensic psychiatric inpatient care—A qualitative study. *Front. Psychol.* **2023**, *14*, 1201485. [\[CrossRef\]](#)
15. Pallavicini, F.; Pepe, A.; Mantovani, F. Commercial off-the-shelf video games for reducing stress and anxiety: Systematic review. *JMIR Ment. Health* **2021**, *8*, e28150. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Aganov, S.; Nayshtetik, E.; Nagibin, V.; Lebed, Y. Pure purr virtual reality technology: Measuring heart rate variability and anxiety levels in healthy volunteers affected by moderate stress. *Arch. Med. Sci.* **2022**, *18*, 336–343. [\[CrossRef\]](#)
17. Almqbaali, M.; Ouhbi, S.; Serhani, M.A.; Amiri, L.; Jan, R.K.; Zaki, N.; Sharaf, A.; Al Helali, A.; Almheiri, E. A Biofeedback-Based Mobile App With Serious Games for Young Adults With Anxiety in the United Arab Emirates: Development and Usability Study. *JMIR Serious Games* **2022**, *10*, e36936. [\[CrossRef\]](#)
18. Saeedi, S.; Bouraghi, H.; Seifpanahi, M.S.; Ghazisaeei, M. Application of Digital Games for Speech Therapy in Children: A Systematic Review of Features and Challenges. *J. Healthc. Eng.* **2022**, *2022*, 4814945. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Kritzinger, W.; Karner, M.; Traar, G.; Henjes, J.; Sihn, W. Digital Twin in manufacturing: A categorical literature review and classification. *IFAC-Pap.* **2018**, *51*, 1016–1022. [\[CrossRef\]](#)
20. Rasheed, A.; San, O.; Kvamsdal, T. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access* **2020**, *8*, 21980–22012. [\[CrossRef\]](#)
21. Hribernik, K.; Cabri, G.; Mandreoli, F.; Mentzas, G. Autonomous, context-aware, adaptive Digital Twins—State of the art and roadmap. *Comput. Ind.* **2021**, *133*, 103508. [\[CrossRef\]](#)
22. Erkoyuncu, J.A.; del Amo, I.F.; Ariansyah, D.; Bulka, D.; Vrabič, R.; Roy, R. A design framework for adaptive digital twins. *CIRP Ann.* **2020**, *69*, 145–148. [\[CrossRef\]](#)
23. Bucchiarone, A. Gamification and Virtual Reality for Digital Twins Learning and Training: Architecture and Challenges. *Virtual Real. Intell. Hardw.* **2022**, *4*, 471–486. [\[CrossRef\]](#)
24. Hocine, N.; Gouaïch, A.; Cerri, S.A.; Mottet, D.; Froger, J.; Laffont, I. Adaptation in serious games for upper-limb rehabilitation: An approach to improve training outcomes. *User Model. User-Adapt. Interact.* **2015**, *25*, 65–98. [\[CrossRef\]](#)
25. Amengual Alcover, E.; Jaume-I-Capó, A.; Moyà-Alcover, B. PROGame: A process framework for serious game development for motor rehabilitation therapy. *PLoS ONE* **2018**, *13*, e0197383. [\[CrossRef\]](#)
26. Antunes, A.; Madeira, R.N. PLAY—Model-based Platform to Support Therapeutic Serious Games Design. *Procedia Comput. Sci.* **2021**, *198*, 211–218. [\[CrossRef\]](#)
27. Bexelius, A.; Carlberg, E.B.; Löwing, K. Quality of goal setting in pediatric rehabilitation—A SMART approach. *Child. Care. Health Dev.* **2018**, *44*, 850–856. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Pingos, M.; Andreou, A.S. A Smart Manufacturing Data Lake Metadata Framework for Process Mining. *Zenodom* **2023**. [\[CrossRef\]](#)
29. Pingos, M.; Andreou, A.S. Discovering Data Domains and Products in Data Meshes Using Semantic Blueprints. *Technologies* **2024**, *12*, 105. [\[CrossRef\]](#)
30. Pingos, M.; Andreou, A.S. A Data Lake Metadata Enrichment Mechanism via Semantic Blueprints. In Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering, Online, 25–26 April 2022; pp. 186–196. [\[CrossRef\]](#)
31. Mayer, I.; Bekebrede, G.; Harteveld, C.; Warmelink, H.; Zhou, Q.; Van Ruijven, T.; Lo, J.; Kortmann, R.; Wenzler, I. The research and evaluation of serious games: Toward a comprehensive methodology. *Br. J. Educ. Technol.* **2014**, *45*, 502–527. [\[CrossRef\]](#)
32. Pacheco-Velazquez, E.; Rabago-Mayer, L.; Bester, A.; Rodes-Paragarino, V. What do we Evaluate in Serious Games? A Systematic Review. In Proceedings of the 17th European Conference on Games Based Learning, Enschede, The Netherlands, 5–6 October 2023; pp. 482–489. [\[CrossRef\]](#)
33. Martinez, K.; Menéndez-Menéndez, M.I.; Bustillo, A. A New Measure for Serious Games Evaluation: Gaming Educational Balanced (GEB) Model. *Appl. Sci.* **2022**, *12*, 11757. [\[CrossRef\]](#)
34. Brooke, J. SUS: A “Quick and Dirty” Usability Scale. In *Usability Evaluation in Industry*; CRC Press: Boca Raton, FL, USA, 2020; pp. 207–212. [\[CrossRef\]](#)
35. Vlachogianni, P.; Tselios, N. Perceived Usability Evaluation of Educational Technology Using the Post-Study System Usability Questionnaire (PSSUQ): A Systematic Review. *Sustainability* **2023**, *15*, 12954. [\[CrossRef\]](#)
36. Norman, K.L. GEQ (Game Engagement/experience questionnaire): A review of two papers. *Interact. Comput.* **2013**, *25*, 278–283. [\[CrossRef\]](#)

37. Saeedi, S.; Ghazisaeedi, M.; Ramezanghorbani, N.; Seifpanahi, M.S.; Bouraghi, H. Design and evaluation of a serious video game to treat preschool children with speech sound disorders. *Sci. Rep.* **2024**, *14*, 17299. [[CrossRef](#)]
38. Peery, A. *A Brief History of A Brief History*; Indiana University Press: Bloomington, IN, USA, 2011; Volume 235, ISBN 9781118901731.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.