information

MDPI

Article

Game On: Exploring the Potential for Soft Skill Development Through Video Games

Juan Bartolomé [1,*], Idoya del Río [1], Aritz Martínez [1], Andoni Aranguren [1], Ibai Laña [1] and Sergio Alloza [2]

[1] TECNALIA, Basque Research and Technology Alliance (BRTA), Parque Científico y Tecnológico de Bizkaia, Astondo Bidea, Edificio 700, E-48160 Derio, Bizkaia, Spain; idoya.delrio@tecnalia.com (I.d.R.); aritz.martinez@tecnalia.com (A.M.); andoni.aranguren@tecnalia.com (A.A.); ibai.lana@tecnalia.com (I.L.)
[2] Independent Researcher, 08192 Sabadell, Spain; sergio.psycogaming@gmail.com
* Correspondence: juan.bartolome@tecnalia.com

Abstract

Soft skills remain fundamental for employability and sustainable human development in an increasingly technology-driven society. These interpersonal and cognitive competencies—such as communication, adaptability, and critical thinking—represent uniquely human capabilities that current Artificial Intelligence (AI) systems cannot replicate. However, assessing and developing these skills consistently remains a challenge due to the lack of standardized evaluation frameworks. This study explores the potential of commercial video games as engaging environments for soft skills enhancement and introduces an AI-based assessment methodology to quantify such improvement. Using player data collected from the Steam platform, we designed and validated an AI model based on Gradient Boosting Regressor (GBR) to estimate participants' soft skill progression. The model achieved high predictive performance ($R^2 \approx 0.9$; MAE/RMSE ≈ 1), demonstrating strong alignment between gameplay behavior and soft skill improvement. The results highlight that video game-based data analysis can provide a reliable, non-intrusive alternative to traditional testing methods, reducing test-related anxiety while maintaining assessment validity. This approach supports the integration of video games into educational and professional training frameworks as a scalable and data-driven tool for soft skills development.

Keywords: artificial intelligence; assessment; soft skills; video games

check for updates

Academic Editors: Petros Lameras, Sylvester Arnab and Panagiotis Petridis

Received: 15 September 2025
Revised: 13 October 2025
Accepted: 16 October 2025
Published: 20 October 2025

Citation: Bartolomé, J.; del Río, I.; Martínez, A.; Aranguren, A.; Laña, I.; Alloza, S. Game On: Exploring the Potential for Soft Skill Development Through Video Games. *Information* **2025**, *16*, 918. https://doi.org/10.3390/info16100918

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

Soft skills are increasingly recognized as crucial for success in both academic and professional contexts [1–5]. These skills encompass personal attributes, interpersonal abilities, and social competencies that complement technical knowledge [6,7], and as global competition intensifies, employees must continuously upgrade both their hard and soft skills to remain competitive and adaptable [8]. They encompass a range of competencies, including communication, teamwork, problem solving, and emotional intelligence, which are essential for effective interaction and collaboration in the workplace and are essential for future career success [5,8–10].

Furthermore, improving soft skills is fundamental to advancing sustainability across education, employment, and organizational practices. These skills empower individuals and teams to effectively address complex sustainability challenges, making them essential for building a more sustainable future [11–13]. As society continues to grapple with

complex sustainability challenges, the human capabilities represented by soft skills will only grow in importance. By recognizing and intentionally developing these skills alongside technical knowledge, we can enhance our collective capacity to create a more sustainable future for all.

In today's rapidly evolving work and educational environments, the relevance of soft skills in enhancing employability is underscored by various studies. Seetha indicated that employers prioritize soft skills alongside technical competencies, as these skills are often indicative of an employee's ability to adapt, innovate, and thrive in dynamic work settings [14]. This dual emphasis on hard and soft skills reflects a broader understanding of employability, where technical knowledge alone is insufficient for long-term career success [15,16]. Basir et al. highlighted that graduates who equip themselves with soft skills significantly improve their employability prospects, aligning with findings from previous research that emphasizes the necessity of these skills for entering the job market and succeeding in self-employment ventures [17]. In this context, institutions are urged to integrate soft skills training into their programs to better prepare students for the demands of the labor market [18]. Consequently, educational institutions and job seekers are increasingly focusing on the cultivation and assessment of these skills to ensure a well-rounded and capable workforce [19].

However, the development of soft skills faces several significant gaps, primarily related to educational frameworks, assessment methods, and societal perceptions, which often lead to inconsistent development and undervaluation of these essential competencies [20]. The lack of standardized definitions or assessment frameworks means these skills are often developed through non-formal education channels, potentially limiting access and effectiveness [6,21]. The definition and scope of soft skills vary considerably across both academic literature and professional practice, highlighting the multidimensional nature of soft skills and their context-dependent interpretations [22]. Furthermore, many educational institutions lack structured programs that effectively integrate soft skills training into their curricula. For instance, Kutz and Stiltner emphasize the necessity of incorporating soft skill assessment within clinical education to enhance leadership and clinical practice [23]. Developing reliable measurement tools for soft skills is crucial, as these skills vary across disciplines and organizational contexts [24].

However, emerging opportunities, driven by advances in digital learning, gamification, and personalized coaching, have been implemented to enhance soft skills in response to the evolving demands of modern professional environments [25,26]. Examples include integrating traditional training with technological innovations [27], such as, for instance, gamification techniques [28], or experiential learning projects, such as the use of virtual reality and AI technologies like ChatGPT [29].

In this context and according to a report published by the European Commission, the number of European gamers has increased significantly over the COVID-19 pandemic, and now over half of the European population regularly plays video games [30]. This prevalent engagement offers a distinctive opportunity for educational interventions that capitalize on established behavioral patterns to enhance learning and employment outcomes. Notably, video games have been associated with improvements in a range of cognitive abilities and soft skills, further supporting their potential as effective educational tools [31–33].

This article shows how video games can be leveraged as an accessible educational tool to develop crucial soft skills, and describes the study carried out within the European project MEGASKILLS, with the aim of improving soft skills using commercial video games. With this objective, we selected four soft skills to train and evaluate (complex problem solving (CPS), critical thinking (CT), cognitive flexibility/adaptability (F/A), and time management (TM)), and several games from the Steamplatform. Steam, available at

https://store.steampowered.com/ (accessed on 30 May 2025), is a digital distribution platform developed by Valve, offering a vast library of video games, integrated community features, and cross-platform compatibility. For this purpose, we used the game data of the participants to design an Artificial Intelligence (AI)-based solution to obtain an indicator of the progress achieved at the soft skill level.

We sought to respond to the following research question: Is it possible to measure whether there has been an improvement in soft skills using players' gaming data?

To comprehensively address the complexities of soft skill development through gaming, our research further investigates the following:

- The specific in-game mechanics and player behavioral patterns that correlate with and contribute to the development of targeted soft skills.
- How player characteristics, such as prior gaming experience and engagement levels, influence the effectiveness of AI-driven soft skill enhancement and assessment within video game environments.

The key contributions of this paper are as follows:

- The development of a novel AI-based stealth assessment methodology that leverages commercial video game data for non-intrusive soft skill measurement addresses a significant gap in standardized evaluation methods. Note that stealth assessment refers to the integration of assessment mechanisms into the game in a discreet and fluid manner.
- The empirical validation of the AI model's (Gradient Boosting Regressor) high efficacy in predicting soft skill improvement, demonstrating a Coefficient of Determination (R^2) value of approximately 0.9 and low Root Mean Squared Error (RMSE) values near 1.
- The practical application of this methodology using real player data from the Steam platform showcases a scalable and engaging approach for integrating game-based soft skill enhancement into traditional training programs.

This research explores the development of a novel approach to soft skill enhancement, leveraging video game-based technological engagement to bridge the gap between education and employment. This paper begins by discussing, in Section 2, the background of the use of video games for soft skills development, the associated positive impact, and their potential as educational tools. In Section 3, we introduce stealth assessment, which is an evaluation approach that integrates assessment seamlessly into interactive and digital environments, such as educational games or simulations, allowing for continuous measurement of skills and competencies without interrupting the learning experience. Section 4 outlines the experimental methodology carried out, focused on gathering game data of the participants and designing an AI-based solution to calculate an indicator of the improvement acquired regarding soft skills while playing. Section 5 presents the results, followed by a conclusion and a discussion of future directions.

2. Video Games for Soft Skill Development

Numerous studies suggest that video games and gamification offer promising strategies for the development and assessment of soft skills [28]. These competencies are increasingly valued in the job market yet often underemphasized in formal education [34]. However, much of the existing literature highlights the potential of such approaches without adequately comparing them to established methods. As noted, "a comparison with existing state-of-the-art methods for soft skill assessment or training. Without benchmarking, it's difficult to judge the effectiveness of the proposed solution relative to existing literature." This observation underscores the importance of rigorous benchmarking to eval-

uate the real contribution of game-based interventions. In fact, Bezanilla et al. (2014) warn that serious games "may be excellent tools for supporting the development and assessment of generic competencies, but not as a stand-alone solution." [35]. Therefore, scholars advocate for the integration of video games within broader methodological frameworks, accompanied by comparative metrics that can validate their relative effectiveness.

While they are primarily designed for entertainment, research indicates that video games can also contribute to soft skill development, especially those with cooperative, strategic, or role-playing elements, which naturally encourage skills that are valuable in both personal and professional contexts, offering immersive environments that enhance teamwork, communication, and problem solving through real-time strategy coordination and decision-making [36–39]. Video games have emerged as innovative tools for fostering soft skills, including intrapersonal, interpersonal, personal social responsibility, and organizational sustainability [32,33]. These skills are crucial for future employees, and structured programs using video games have shown measurable improvements in soft skills among university and vocational students [40]. Indeed, Clark et al. in their meta-analyses provided additional evidence that digital game design is effective in fostering learning outcomes comparable to those obtained through conventional educational approaches [41].

Video games provide interactive environments where players often require them to assume leadership roles, negotiate conflicts, and collaboratively navigate complex scenarios, thereby mirroring the challenges found in real-world professional contexts [42–44]. In addition, digital game-based learning has been shown to significantly boost motivation and cognitive engagement [45]. Additionally, immersive gaming experiences offer a low-risk environment in which learners can experiment with decision-making and build resilience, skills that are increasingly critical in dynamic work environments [32].

Regarding the soft skills focused on the present investigation, recent research provides robust evidence that gaming environments accomplish the following:

- Foster a wide range of complex problem-solving skills: improving abilities in problem decomposition, systems thinking, and causal analysis [46]; enhancing spatial reasoning, sequential processing, and overall solution optimization [47]; enhancing strategic planning and analytical reasoning [48]; and improving specific cognitive areas like attention and reasoning [49].
- Enhance critical thinking skills: improving evidence evaluation and strategic decision-making [50] as well as developing stronger analytical reasoning and fostering sophisticated information synthesis [51].
- Improve adaptability, flexibility, and resilience skills: bolstering emotional intelligence and resilient thinking [52] as well as improving task-switching abilities, environmental adaptation, and adapting more efficiently to novel situations and exhibiting enhanced strategic versatility [53].
- Enhance time management skills: improving task prioritization, time estimation and project planning skills, and improving executive control skills [54].

Overall, these studies underscore the multifaceted impact of gaming on cognitive development, suggesting that interactive digital environments can serve as effective platforms for cultivating critical thinking, time management, adaptive reasoning, and strategic problem-solving skills.

3. Stealth Assessment in Video Games

Stealth assessment (SA) in the context of video games refers to the integration of assessment mechanisms into the gameplay in a way that is unobtrusive and seamless [55]. SA allows for the evaluation of players' skills, knowledge, and behaviors without interrupting the gaming experience or making the assessment explicitly apparent to the

player [56]. Furthermore, SA not only mitigates the anxiety associated with traditional testing but also promotes a more engaging learning environment conducive to skill acquisition and retention [55].

The evolution of stealth gameplay is closely tied to advancements in AI and game design [57]. However, while existing AI approaches partially address the weaknesses identified in the current assessment paradigm, such as, for example, the difficulty in the design and implementation of assessments by educators, there is still much to be done [58]. By focusing on the process rather than solely the outcome, educators can leverage such data to foster tailored learning experiences, underscoring the transformative power of video games in contemporary education.

In the realm of stealth gameplay assessment, traditional metrics often fall short in capturing nuanced player behaviors that define successful stealth mechanics. Consequently, researchers are exploring advanced metrics to measure and analyze player performance, such as

- Player behavior and action logs: Data from players' actions, such as movement, interaction with objects, and decision-making, provide insights into skills development. These action logs can be analyzed to infer critical soft skills [59]. This information has been applied to profile player behavior and categorize players [60], to perform skill and performance analysis [61], and to visualize and cluster players with similar behaviors [62].
- Response time and decision speed: In games with time-sensitive challenges, response time is a critical metric. These metrics not only reflect the cognitive and perceptual benefits of gaming but also highlight potential areas for training and improvement: reaction times [63], sensorimotor decision-making capabilities [64], balance between speed and precision [65], and processing speed [66].
- Player interaction data: Interaction data, such as chat logs, cooperative task completion, and social dynamics, are important indicators of collaboration and communication skills [67]. Metrics such as interaction frequency, leadership role assumption, and conflict resolution strategies can be tracked.
- Success rates and achievement tracking: These metrics provide valuable insights into player behavior, game content consumption, and performance dynamics, which can inform game development, enhance player retention, and optimize gaming experiences across various genres [68]. Genres include the following: developing performance prediction models to analyze player actions, such as predicting hits or misses [69]; tracking individual performance to understand dynamics within ad hoc teams in team-based games [70]; reporting how cognitive skills are linked to gaming performance [71]; and analyzing player retention by modeling motivations, progression, and churn to predict dropout rates and improve retention strategies [72].
- Behavioral patterns: These are used to measure and enhance player performance by understanding their actions, strategies, and engagement levels [73,74]. This approach not only aids in improving game design and player satisfaction but also provides insights into broader social behaviors that reveal the development of soft skills, such as adaptability or leadership [75,76].

Apart from the previously listed metrics of interest for the objectives of this study, other advanced metrics have also been studied such as social network analysis, by providing insights into player interactions, engagement and retention, researchers can gain insights into players' collaboration, leadership, and influence [77]; similarity-based metrics, designed to differentiate novice from expert performance by comparing player actions to multiple expert solutions, or to rank players based on their competency levels [78]; and

audiovisual and feedback-based metrics, to analyze audiovisual streams to quantify player experiences and performance [79].

From the methodological point of view, several approaches have been employed to implement SA in video games [80]:

- In-game analytics: Games are often designed with built-in systems that automatically collect and analyze data from players' in-game behaviors. These analytics are used to enhance game design, improve player engagement, and tailor experiences to different player profiles [81,82]. The integration of machine learning and predictive models further enhances the ability to analyze and predict player actions, making in-game analytics an essential tool for developers and researchers [83].
- AI and machine learning models: provide tools for performance assessment [83], behavior prediction [84,85], and player engagement analysis [86].

Apart from the previously listed approaches of interest for the objectives of this study, other approaches have been employed such as observational assessment, where human observers track player behavior in multiplayer environments, paying attention to social interactions, leadership roles, and collaboration dynamic to supplement in-game data [87], in addition to player surveys and self-reports, by providing self-reflection from players about their perceived progress in developing soft skills. For instance, they were used to evaluate aspects like immersion, competence, and satisfaction [88], as well as longitudinal tracking, which, by leveraging spatio-temporal data and self-tracking techniques, players and analysts can gain deeper insights into gameplay, leading to improved strategies and skill development [89].

In conclusion, SA in video games is a viable and innovative approach for enhancing soft skills. It provides a dynamic, engaging, and effective means of assessing and developing competencies. While challenges remain in design and implementation, ongoing research and technological advancements continue to improve its efficacy and applicability. SA leverages in-game data to unobtrusively evaluate players' cognitive and interpersonal skills, providing a dynamic and adaptive means of fostering soft skills development. This approach allows for real-time assessment without disrupting the immersive experience, offering a more authentic and context-sensitive measure of competencies. Furthermore, by tailoring feedback and learning opportunities based on in-game behaviors, SA enables personalized skill development, reinforcing the potential of video games as powerful tools for education and professional growth.

4. Materials and Methods

4.1. Methodology

To test the hypotheses selected, we began selecting the soft skills for the study considering two different aspects: (1) the perceived importance in the academic and work contexts analyzed in the initial stages of the project, and (2) the hypothetical feasibility of measuring and training these soft skills, with both video games and standard tests (for this reason, important skills such as collaboration and communication have been left out of the scope of this study). The soft skills selected were CPS, CT, F/A, and TM.

Furthermore, since the measurement was carried out both outside and inside video games, both sources of data must be valid, objective, and accessible. Then, measurement vectors/elements (scales and subscales of standardized tests) were extracted in order to identify them in commercial video games. Eventually, 4 specific video games available on the Steam platform (Bellevue, WA, USA) were also chosen in that regard, associated with each soft skill, thus building a theoretical bridge between the soft skills standard measurement and how these specific video games could improve them, as an innovative justification element, in a pre–post phase method with a control group.

From October 2023 to January 2024, the recruitment of participants was carried out. The different entities contacted potential participants from all European universities and companies, and even from outside Europe, introducing them to the project, the objectives, and inviting them to participate. Once the participants were registered, the different groups were randomly generated in early February 2024, when the experimental playing period officially began.

During the registration, we asked the participants to introduce their Steam ID (a unique identifier used to identify a Steam account) and to make their profile in Steam public. This action was necessary since we designed an automatic task to gather the game data of the participants from the Steam platform and register it in our database. This automatic task was developed using the API (application programming interface) provided by Valve Corporation (Bellevue, WA, USA) and integrated into a web platform that was designed to be used during the piloting for participant registration and test data collection. Then, we carried out the piloting based on a pre-test post-test design with a control group. We created five groups of participants, with four of them associated with four separate soft skills and video games, and a last group, as a control, that only completed the pre-test and post-test without playing video games. During this period, the participants played video games at their own pace, with no timetables or minimum time restrictions per day. The piloting ended by the middle of April. Once the data was collected, the methodology we followed to design the AI-based model is as shown in Figure 1.

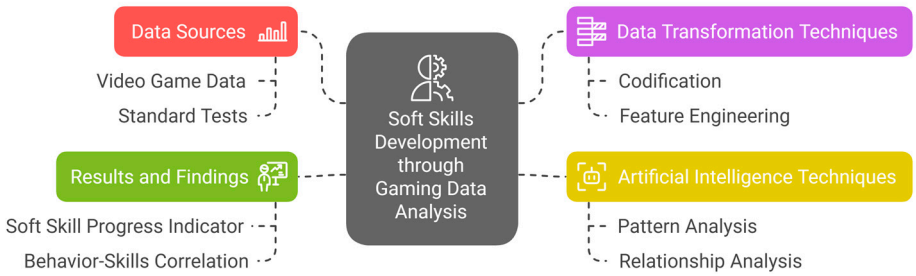

Figure 1. Methodology followed in the study.

Besides the data gathered during the experiments designed above, an achievement value database has been generated with expert knowledge. This database establishes a set of difficulty levels of each target soft skill for a given game. Then, the general difficulty of the soft skill is translated to a specific difficulty level per achievement and soft skill, which in turn allows for obtaining an index of the player's advancement on such skill, provided that the player has attained a given achievement. This mapping allows for relating the level of a player and the achievements the player has unlocked, and can be used to gauge the output of tests. A research flowchart showing the different steps followed in this study is shown in Figure 2.

It is important to note that, in the present study, statistical analyses have been conducted primarily on the relationships between gameplay behavior and the development of selected soft skills, specifically CPS, CT, F/A, and TM. At this stage, the analyses do not account for potential differences among participants based on demographic variables such as age, cultural background, or geographic region. These aspects remain outside the scope of the current statistical evaluation and will be addressed in future supplementary research aimed at identifying how such factors may influence soft skill acquisition through gaming. This clarification is intended to ensure transparency regarding the boundaries of the present findings and to prevent potential misinterpretations of the results.

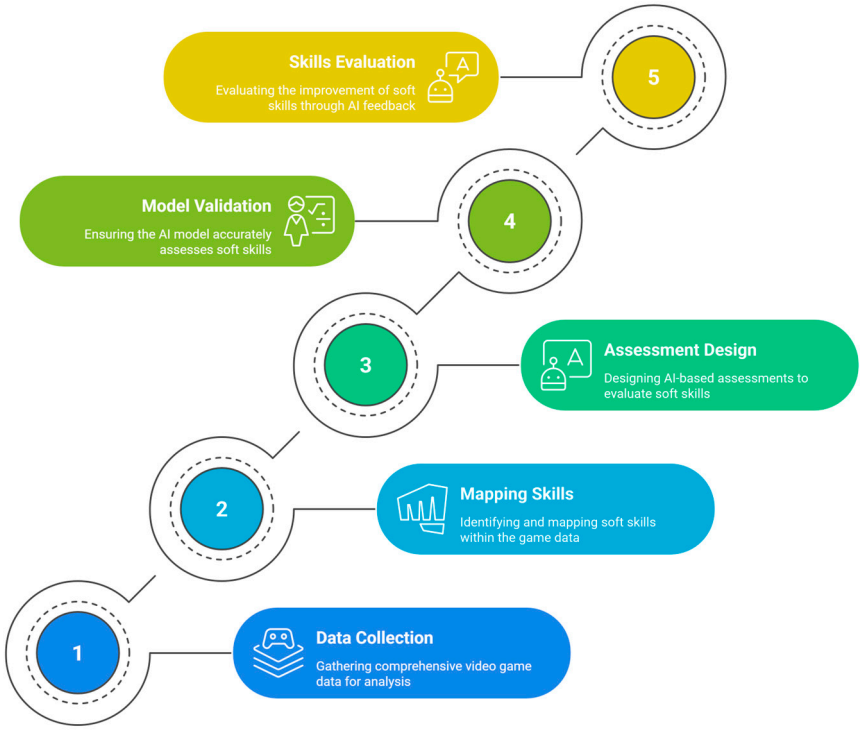

Figure 2. Research flowchart applied in the study.

4.2. Sample

In this phase, 437 email addresses from interested participants were gathered. However, only 120 participants completed all the steps required in the investigation for performing the comparison between the pre-test and the post-test. For the objectives of this study, we removed 2 participants from the control group, and, finally, 97 participants were considered. The details of the sample selected are shown in Table 1 and in the following paragraphs.

Table 1. Gender sample distribution.

Gender	Complex Problem Solving	Critical Thinking	Flexibility/Adaptability	Time Management	Control
Male	14	9	16	17	19
Female	11	5	5	9	10
Other	1	2	0	0	0
Prefer not	1	1	0	0	0
Total	27	17	21	26	29

The gender of the participants or the country of origin was not considered when selecting the groups. The distribution was random. Despite there being more males than other genders in each group, we can assume that equality is oscillating between 51% and 76% depending on the group, being that males make up 62.5% of the total sample. Most of the participants are considered young (around 80% of the sample are 35 years old or younger). They also present high educational levels, with a master's degree in first place (38.3%), followed by a bachelor's (30.8%) and doctorate or postdoc (15%). Regarding the video game experience, we can assume that most of the sample is somehow in touch with video games daily (49.1%) or at least weekly (78.3%), presenting a total of 14.2% of participants who play occasionally between months, and 7.5% of the sample that never plays video games.

As a summary, the final sample of 120 participants is made up mostly of young people, both men and women, with an university degree or higher education, most of them being students with previous experience in video games and having played a wide variety of video game genres.

4.3. Materials

To analyze the measurements obtained outside and inside the video games, we selected the following standardized tests, which were mapped in accordance with the video games selected.

4.3.1. Standardized Tests

To select the specific soft skill assessment tools, we carried out a short review of the current methodologies of soft skill assessments, identifying several standardized tests. Then, we analyzed them in depth to choose the best fit for the investigation, according to the feasibility of accessing, digitizing, and autocorrecting the tests, since their use had to be free and available online. We also considered other aspects, such as the age of the tests, the target group tested with, or the length of the tests, to avoid possible dropouts and biased answers due to boredom or frustration. The final standardized tests chosen were

- CPS: The Reflective Thinking Skill Scale for Problem Solving (RTSSPS) [90].
- CT: Two tests were selected, the Critical Thinking Assessment Scale Short Form [91] and the Test of Critical Thinking [92].
- F/A: Two tests were selected, the I-ADAPT-M [93] and the CAMBIOS test [94].
- TM: The Time Management Questionnaire [95].

4.3.2. Video Games

To select the video games, the main criterion followed was that they should have some kind of connection with each of the soft skills. First, we analyzed the indicators or measurement elements of the soft skills measurement tools, with the aim of being able to extract these elements and be able to identify them in the video games. In this way, a theoretical bridge is built between the measurement with standardized tests and the stimulation of skills through the use of already-existing commercial video games. Then, we conducted a search for video games that fit this criterion, thus generating a range of video game possibilities that presented from 1 to several soft skill indicators. Moreover, other aspects were also considered, such as price, inclusion of gender, existence of offensive language or violence, low technical requirements, and the duration of the video game itself, since the training period was set at a minimum of 15 h of play. Indeed, prior studies, e.g., [96], indicate that a minimum of 12–15 h of video game play is necessary to produce significant improvements in skill development.

Specifically, we selected the Steam video game distribution platform developed by Valve Corporation in 2003 for several reasons. First, Steam provides an open API, which allows the gathering of player information, such as their performance in the video games played on Steam. In addition, Steam video games also present achievements as milestones to be achieved within each video game, which could be associated with the different soft skill levels. Finally, Steam presents the largest range of commercial video games in the personal computer (PC) market. Considering all the mentioned criteria, the final selection of video games was

- Train Valley 2: This video game fits perfectly in a complex problem-solving experiment context since the video game presents different types of problems and variables to consider. The user must learn to manage several train stations, from building the tracks and ways for the trains to circulate (having a limited budget for this task) and

managing the departures of said trains. The difficulty, in addition to increasing as the levels progress, lies in the anticipation and planning of the different train routes, having to either prevent problems or solve them in the event of an accident.

- Lightseekers: As a card video game, it meets many of the requirements to stimulate and train critical thinking. Players will have to plan and analyze the different possible compositions of cards or decks to find out which one has the best performance, depending on the environment or context of each game.
- Relic Hunters Zero: It is a very friendly and entertaining top-down shooting roguelike game. This video game genre is perfect in the flexibility experimentation context, since the rules of each game change once the player dies and resets the progress. Thus, it allows the use of different weapons and strategies that encourage paradigm and variable shifts, as a main flexibility indicator, and the parallel learning of each of the elements that may be present in the game.
- Minion Masters: It is a dual management game. On the one hand, the player must manage time, both macro and micro, during the games. This means that users must consider the countdown present in each game as well as the different deployment times, movement speed, and attack speed of each unit. On the other hand, outside of games, the player can manage the deck and the collection of units that he collects and improves. The gameplay mainly consists of attacking enemy turrets and bases while defending your own in 1vs1 battles, where all types of units with different characteristics are deployed to achieve this objective.

Table 2 shows the relationship between the soft skills trained, the scales and subscales of standardized tests used, and the video games selected.

Table 2. Soft skills selected, scales and subscales of standardized tests, and video games.

Soft Skill	Test	Video Games
CPS	Reflective Thinking Skill Scale for Problem Solving (RTSSPS) (onwards CPS)	Train Valley 2
CT	Critical Thinking Assessment Scale Short Form (onwards CTS) Test of Critical Thinking (onwards CTO)	Lightseekers
F/A	I-ADAPT-M (onwards F/AS) CAMBIOS (onwards F/AO)	Relic Hunters Zero: Remix
TM	Time Management Questionnaire (onwards TM)	Minion Masters

4.3.3. MEGASKILLS Platform

An online web platform was used and provided all the functionalities required in this phase, such as registration and identification on the platform, linking the registration with the Steam ID, integration of the standard tests (pre-test and post-test), a catalog of selected games, and details of the soft skills addressed. All the standard tests selected were digitized and integrated into the web platform. An example of one of the tests integrated on the platform is shown in Figure 3.

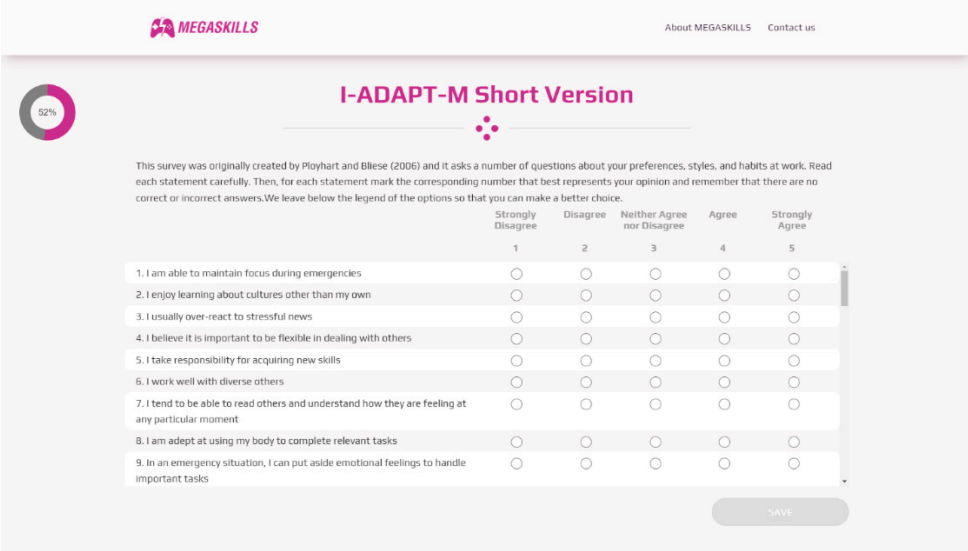

Figure 3. I-ADAPT-M Short Version [93].

In addition to registering on the MEGASKILLS platform, all participants had to register on Steam, available at https://store.steampowered.com/ (or use their account if they already had one). Furthermore, to play the selected games, it was necessary to have the Steam client installed on the participants' computers. Through the platform, participants were able to see where they were in the study, helping them to know what steps they needed to take to move forward in the study.

To collect the game data from the participants, we developed an automatic task that was integrated into the MEGASKILLS platform, which launched a series of queries once a day, making use of the Steam API. Through these queries, the platform's database was updated once a day. Table 3 presents the list of queries launched and the description of the information gathered.

Table 3. Description of the queries provided by the Steam API.

Query	Description
GetPlayerSummaries	Returns basic profile information. Some data associated with a Steam account may be hidden if the user has their profile visibility set to "Friends Only" or "Private". In that case, only public data will be returned.
GetFriendList	Returns the friend list of any Steam user, provided their Steam community profile visibility is set to "Public".
GetOwnedGames	Returns a list of games a player owns along with some playtime information.
GetRecentlyPlayedGames	Returns a list of games a player has played in the last two weeks.
GetPlayerAchievements	Returns a list of achievements for this user by the game identifier (app ID).
GetGlobalAchievement PercentagesForApp	Returns on global achievements overview of a specific game in percentages.

To streamline the pilot process on our platform, we introduced automated checks to confirm participants met key requirements, including Steam ID registration, public profile settings, minimum gameplay hours, and evaluation form completion.

4.3.4. Procedure

Between October 2023 and January 2024, participant recruitment was conducted via email. After providing informed consent, individuals were invited to register on a dedicated online platform designed for this study. This platform facilitated the collection of socio-demographic information and responses to soft skill assessments administered at both the pre-intervention and post-intervention phases. All the participants had to fill out the pre-intervention and post-intervention tests. In early February 2024, following registration, participants were randomly assigned to groups, marking the official start of the experimental phase. The groups were as follows: group1, Train Valley 2; group2: Lightseekers; group3, Relic Hunters Zero: Remix; group4: Minion Masters; and group5, control. During registration, participants completed baseline soft skills tests and received detailed instructions regarding their participation. Specifically, individuals in the experimental groups were required to play a minimum of 15 h on one of four designated video games during February and March 2024. In contrast, control group participants were instructed not to engage with any of these games to maintain the objectivity of the collected data. At the conclusion of the gaming period in late March 2024, a post-test phase was administered, and the collected data were subsequently analyzed.

The data gathered during the gaming phase were used to train an AI model that ultimately could operate in the place of the standardized tests designed to measure the level of a given soft skill. The main rationale establishes that at any given moment, an individual has a measurable level of a certain soft skill, developed through past life experience; gaming lies within this experience, and a period devoted to gaming can potentially increase the level of some of the soft skills, if they are part of the game mechanics. In principle, the soft skill level of experiment participants is measured before and after the gaming period, so the soft skill increment can be measured.

Thus, a machine learning model (M1) can be trained to receive an initial soft skill value and a set of variables that represent the playing time within a period and relate these two inputs to an output that reflects the increment in the soft skill value. However, if the model is intended to take the place of the tests, the variables related to the soft skill level prior to the gaming period would not be available to feed the input of the trained model, and, in fact, it would not be possible to estimate this initial level. For this reason, the model needs to be able to work without this information. To do so, two main changes are applied to the scheme above: on one hand, the initial soft skill value is allowed be 0 if unknown (normally no person would have a null soft skill, so 0 means unknown), and the model will also be trained with samples of unknown initial level; on the other hand, an additional piece of information, based on the features of the player is computed into an embedding, i.e., a way of converting different things like words or images into a list of numbers, where items with similar meanings are given similar numbers. This information is easy to obtain (provided that the player has a playing background) and gives an insight into the kind of soft skills that the player practices the most, so even if there is no soft skill information, the model has a profile of the individual and a measurement of the time played within a period.

In addition, in order to deal with the cases in which there is an unknown initial level of soft skill, a second model (M2) is trained to infer this value (which is known for the cases of the participants in the study). The second model receives the soft skill increment computed in the first model and the playtime within the gaming period and is trained to predict the initial level. Thus, this estimated initial level can be used as new input for the previous model to adjust the initial level to the one estimated by the second model.

The interactions of these two models are depicted in Figure 4, and their training involves using the data gathered during the experimental phase, for which the soft skill initial value and increment, as well as the playtime and player profiles, are known. In

addition, two ways of data augmentation are used: first, to train the model for the cases where no initial soft skill value is provided, new data analogous to existing ones but with this value initialized to 0 are created and added to the dataset; second, noise-based variations in actual data are included, in order to deal with their scarcity.

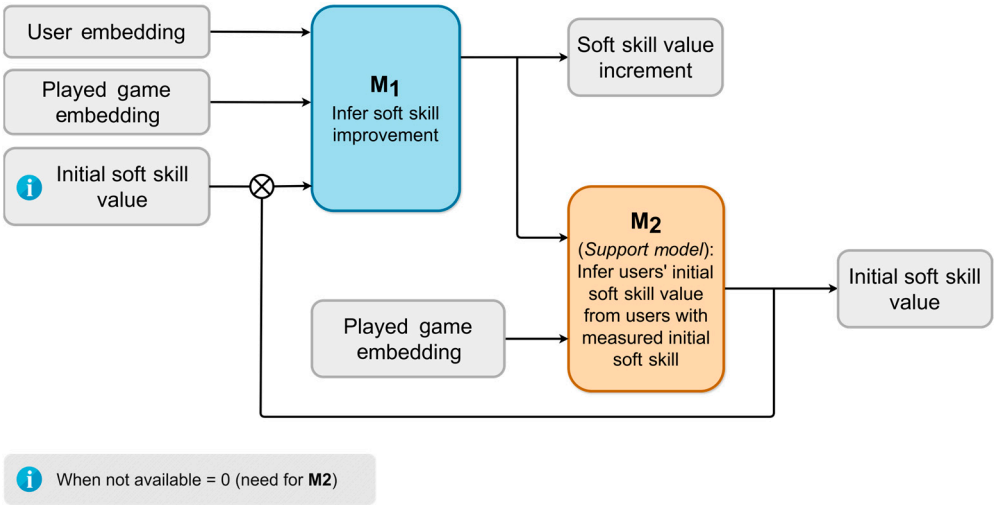

Figure 4. Overall AI methodology. Model 1 (M1) computes users' soft skill deltas (or increment in soft skill values) and improvement percentage. Model 2 (M2) is used as support for M1 to derive a baseline for users without initial soft skills.

The machine learning scheme presented above is trained for each target soft skill (e.g., there will be a model for TM increment estimation and a different one for CT). Once each model is trained, they are used for the estimation of the soft skill value increment for any new individual for whom playtime is measured, and a profile is created through the user embedding. These increments are therefore used to assess the impact of the gaming sessions on soft skill development.

The models described above rely on the definition of a user embedding $\overline{U}$ that describes the relation of the user's gaming profile with the soft skills. The user embedding is a 1-dimensional array with 4 elements that correspond to the values of soft skills attributed to a user, given his/her gaming profile. To generate this embedding, we have considered the user's previous experience playing games, this being the only available information that can be extracted from users who have not completed the soft skill evaluation tests. For each game that the user has played in his/her Steam profile, a game embedding $\overline{G}$ is used and weighted according to the time spent playing this game (at least 10 h of gameplay are required to compute the skills provided by a game). The values provided by all games in the user profile are aggregated to obtain each of the soft skill values that compose $\overline{U}$.

This system requires a way to compute a game embedding $\overline{G}$ for each game. Analogously to the user embedding, $\overline{G}$ is a 1-dimensional array with 4 elements that represent the level of soft skills provided by a game. This value is represented with a float number between 0 and 1, and it is computed considering the main tags of the game, information that is publicly available. The computation of this value is based on the 10 most relevant community-labeled tags of each game and their relevance for each soft skill based on expert knowledge. For example, co-op games are considered to be relevant to assess the cognitive flexibility of a player as they require reactivity to the actions of your teammates; similar logic can be extended to the other soft skills. Hence, given that each game in Steam comes with a set of labels that developers and even the community have designated for that game, which are known to be somehow representative of the game mechanics, storytelling, art, or mood, among others, expert knowledge can be applied to compute the percentage of the game that

could be represented by each soft skill. The Steam Games Dataset (Steam Games Dataset available at: https://www.kaggle.com/datasets/fronkongames/steam-games-dataset (accessed on 15 October 2025)) data used to compute the preliminary soft skills per game provides information about more than 100 k games. Even if the dataset includes diverse information, from game descriptions to images, the most valuable data fragments for the soft skill computation task are the tags and the number of clicks the community has given to each of them. This can be considered a useful and representative way of measuring the relevance of each tag for a given game, which will be used later in this same section. For the 4 soft skills that have been tackled within this paper, the assignment of tags to soft skills is performed according to Figure 5.

```
SOFT_SKILL_TAGS = {
    "Time Management": [
        "Resource Management", "City Builder", "Time Management",
        "Automation", "Tycoon", "Colony Sim", "Incremental",
        "Farming Sim", "Time Attack", "Base Building",
        "Tower Defense", "Sandbox", "City Planning", "RTS",
        "Simulation", "Cooking"
    ],
    "Flexibility/Adaptability": [
        "Roguelike", "Rogue-lite", "Roguelite",
        "Procedural Generation", "Metroidvania",
        "Souls-like", "Survival", "Bullet Hell",
        "Action Roguelike", "Hack and Slash", "MOBA",
        "PvP", "FPS"
    ],
    "Complex Problem Solving": [
        "Puzzle", "Detective", "Escape Room",
        "Hacking", "Choices Matter", "Point & Click",
        "Logic", "RTS", "Tower Defense"
    ],
    "Critical Thinking": [
        "Strategy", "Tactical", "Turn-Based Strategy",
        "Grand Strategy", "Deckbuilding", "4X", "Chess",
        "Wargame", "Turn-Based Tactics", "RTS", "Card Game",
        "Choices Matter"
    ]
}
```

Figure 5. Assignment of Steam tags to soft skills.

These tags are a subset of the whole Steam tag set. In order to reduce the whole set of tags to those that are relevant to the target soft skills, the following procedure is performed:

- Remove all games that contain tags suggesting the game does not contribute to the development of soft skills, e.g., "SEXUAL CONTENT", "NUDITY", "MATURE", "NSFW".
- Remove all tags that belong to groups that do not contribute to the development of soft skills, e.g., "Uncategorized", "Franchises", "Hardware", "Tools".

After having a curated tag list, the games are labeled into an embedding following the next steps:

- If a game has more than 10 tags, the embedding will only consider the first 10. This is performed by a function F(T) which also removes and filters tags, defined hereafter as T, and returns a curated list of tags T′.
- If the total amount of clicks in a game is less than 200 clicks, it is penalized. This is conducted when the total amount of clicks is computed. Given the function $C(t)\ \forall\ t\ \in T'$, returning the number of clicks for a tag, the total click amount is computed as $K = \max\left(\frac{1}{|T'|}\sum_{t\in T'} C(t),\ 200\right)$. This simple consideration will identify games with

fewer than 200 clicks in their total amount of clicks, in contrast with those having fewer tags with many clicks.

- The tags of a game are weighted depending on their position on the list (which depends on the clicks). This weighting process is applied for each soft skill as $P(t) = \frac{C(t)}{K}$, where $t \in T'$ and K is the total amount of clicks as introduced previously.
- Finally, a value between 0 and 1 that represents the contribution of the game to a soft skill is computed, considering the number of clicks and the ranking of tags, and the soft skill. For that, a function that returns the weight of a tag (t) for a soft skill is defined as $W(t, ss) \mid t \in T'$ and $ss \in S$, further defined as $S = \{(u, L_u) \mid u \in U\}$, where U is a list of soft skills and L_u is the list of tags under that category, for this study of those described in Figure 4. Finally, the sum of all the values for each soft skill is computed as $H_{ss} = \sum_{ss \in S} \sum_{t \in T'} W(t, ss)$.

With these criteria, a preliminary labeling of games has been carried out considering the available data (a database of a selection of games generated by experts that attributes a level for each soft skill), and a deep learning model was trained using the whole tag set for each game (i.e., nor filtering nor weighting tags) and the soft skill levels as the predicted value. This model is now able to estimate soft skill levels for games given their associated tags, thus creating a game embedding $\overline{G}$ for each game, which can be used for the models presented above. As a means to elucidate how this method can capture experts' logic to assign soft skills to each game and even help them in future game selection decisions, as a value for each soft skill is computed. Table 4 compares the assignations made by the experts with those made by the AI model introduced in this section.

Table 4. Contrasting experts' decisions with an AI-based solution. Numbers in the "AI embedding" column are ordered from more representative to less representative.

Name	Tags (Top 12)	Expert Assigned Soft Skill	AI Embedding
Lightseekers	Free to Play, Strategy, Card Game	CT	CT-0.2 TM-0.0 CPS-0.0 F/A-0.0
Minion Masters	PvE, Trading Card Game, Stylized, Turn-Based Tactics, Free to Play, Multiplayer, Card Battler, Tactical RPG, PvP, Strategy, Deckbuilding, Card Game	TM	CT-0.5 TM-0.3 CPS-0.2 F/A-0.0
Relic Hunters Zero: Remix	Looter Shooter, Twin Stick Shooter, Action, Top-Down Shooter, Action Roguelike, Adventure Co-Op Multiplayer, Free to Play, Bullet Hell, Pixel Graphics, Local Co-Op	F/A	F/A-0.4 CT-0.0 TM-0.0 CPS-0.0
Train Valley 2	Quick-Time Events, Multiple Endings, Trains, Relaxing, Simulation, Strategy, Singleplayer, Management, Level Editor, Indie, Building, Puzzle	CPS	CT-0.2 TM-0.2 CPS-0.1 F/A-0.0

5. Results

The average time played by the participants was 18.1 h (SD = 7.59), and the average of achievements was 14.18 (SD = 8.74). Table 5 presents the average time played and the average achievements by the participants in each of the video games.

Table 5. Average time played and achievements obtained per game.

Video Game	Average Time Played (h) (SD)	Average nº Achievements (SD)
Lightseekers (n = 15)	16.63 (SD = 1.89)	10.93 (SD = 4.69)
Minion Masters (n = 27)	18.20 (SD = 5.15)	13.27 (SD = 6.41)
Relic Hunters Zero: Remix (n = 19)	15.6 (SD = 1.14)	7.19 (SD = 6.54)
Train Valley 2 (n = 26)	20.34 (SD = 12.10)	21.45 (SD = 8.45)

The average of the results obtained in the pre-intervention and post-intervention tests is shown in Table 6. The different ranges of scores established for each test can be consulted in Appendix A.

Table 6. Average of results obtained in the pre-intervention (PRE) and post-intervention (PRO) tests.

Group	Pre CPS	Pre CTO	Pre CTS	Pre F/AO	Pre F/AS	Pre TM	Post CPS	Post CTO	Post CTS	Post F/AO	Post F/AS	Post TM
Lightseekers (n = 15)	42.2 (SD = 5.25)	32 (SD = 8.41)	235.26 (SD = 67.03)	18.86 (SD = 6.71)	126.73 (SD = 11.32)	56.26 (SD = 8.48)	42.06 (SD = 6.85)	31.53 (SD = 8.36)	246.93 (SD = 49.02)	20.13 (SD = 6.78)	128.46 (SD = 11.07)	55.13 (SD = 7.16)
Minion Masters (n = 27)	41.29 (SD = 4.85)	32.07 (SD = 4.67)	247.81 (SD = 41.41)	20.29 (SD = 5.80)	127.18 (SD = 8.52)	56.66 (SD = 8.41)	42.81 (SD = 5.09)	33.03(SD = 6.09)	256.88(SD = 39.77)	21.88(SD = 4.61)	131.81 (SD = 11.37)	57.81 (SD = 9.81)
Relic Hunters Zero: Remix (n = 19)	41.05 (SD = 5.32)	34.52 (SD = 3.53)	232.73 (SD = 44.54)	17.47 (SD = 7.11)	129 (SD = 12.45)	57.94 (SD = 7.74)	41.52 (SD = 7.23)	34.05 (SD = 3.59)	249.89 (SD = 50.01)	19.42 (SD = 6.51)	131.15 (SD = 10.95)	58 (SD = 8.47)
Train Valley 2 (n = 26)	41.65 (SD = 5.48)	32.34 (SD = 7.00)	245.42 (SD = 48.89)	18.30 (SD = 5.94)	129.84 (SD = 11.10)	61 (SD = 6.82)	43 (SD = 4.40)	33.88 (SD = 5.53)	256.07 (SD = 45.50)	19.44 (SD = 6.80)	133.30 (SD = 9.51)	61.42 (SD = 7.34)
Control Group (n = 29)	42.17 (SD = 6.39)	32.31 (SD = 5.49)	252.48 (SD = 54.34	18.10 (SD = 6.69	129.89 (SD = 11.75	59.10 (SD = 8.67	42.13 (SD = 5.78	33.27 (SD = 5.60	252.20 (SD = 47.64	19.24 (SD = 6.18	134.79 (SD = 11.06	58.82 (SD = 7.62

For each of the five groups, we performed a statistical validation to support that the observed improvements are significant. So, we conducted non-parametric tests, specifically the Mann–Whitney test, to compare the average scores of each game group to the control for both post scores and Change (post−pre) with Bootstrap Confidence Intervals (CIs) for median differences and attempted mixed-effects models (some show convergence warnings, likely due to small sample sizes and separation). We fitted mixed-effects models for each test to assess the impact of pre-scores and game type (see Table 7). For each comparison between a game group and a control group, we defined H_0 (null hypothesis), i.e., the distributions of improvement scores are identical between the game group and control group (no difference in central tendency), and H_1 (alternative hypothesis), i.e., the distributions differ in location (there is a difference in central tendency). We also considered the following assumptions: independence (observations within and between groups are independent), ordinal data (improvement scores can be ranked), and similar distribution shapes. Groups have similar distribution shapes, which allows interpretation in terms of medians. Each row corresponds to a specific soft skill test (see Table 2) assessed for each game. The U Statistic indicates the test value; the p-value shows the statistical significance ($p < 0.05$ typically considered significant); and the effect size indicates the magnitude and direction of the observed difference (positive values suggest improvement, negative values suggest decline).

Table 7. Average of results obtained in the Mann–Whitney test (ML = Median Diff CI low; MH = Median Diff CI high).

Test	Game Group	*p* Post	U Post	R Post	Post ML	Post MH	U Change	*p* Change	R Change	Change ML	Change MH	Effect Size
CPS	Lightseekers	0.7380	263	0.0497	−3	6	243.5	0.9377	−0.0128	−6	4	0.0161
CPS	Minion Masters	0.4650	495	0.0949	−3	4	514.5	0.3067	0.1323	−2	4	−0.1548
CPS	Relic Hunters Zero: Remix	0.8044	300	−0.0355	−4	4	331.0	0.7457	0.0461	−5	5	−0.0558
CPS	Train Valley 2	0.3086	496	0.1331	−2	5	487.0	0.3788	0.1152	−2.5	4	−0.1351
CTO	Lightseekers	0.8062	236	−0.0369	−4	5	200.0	0.2941	−0.1524	−5	2	0.1919
CTO	Minion Masters	0.9169	453	0.0143	−3	3	519.5	0.2718	0.1419	−1	3	−0.1661
CTO	Relic Hunters Zero: Remix	0.6749	336	0.0592	−3	4	260.0	0.3119	−0.1409	−3	2	0.1706
CTO	Train Valley 2	0.5502	468.5	0.0785	−2	4	462.5	0.6129	0.0665	−2	3	−0.0780
CTS	Lightseekers	0.8674	239.5	−0.0256	−48	50	295.0	0.2957	0.1524	−16	52	−0.1919
CTS	Minion Masters	0.5131	490	0.0853	−19	38	538.5	0.1692	0.1783	−11	37	−0.2087
CTS	Relic Hunters Zero: Remix	0.9772	315.5	0.0052	−37	44	416.5	0.0513	0.2714	3	42	−0.3285
CTS	Train Valley 2	0.5720	466.5	0.0745	−18.5	45	545.5	0.0764	0.2315	−1	42	−0.2715
F/AO	Lightseekers	0.3548	289.5	0.1348	−2	7	259.5	0.7967	0.0385	−2	3	−0.0484
F/AO	Minion Masters	0.0283	593	0.2829	−1	8	501.5	0.4074	0.1074	−2	3	−0.1257
F/AO	Relic Hunters Zero: Remix	0.7029	334	0.0540	−2	6	349.5	0.4970	0.0948	−2	3	−0.1148
F/AO	Train Valley 2	0.4449	461.5	0.1010	−5	7	434.0	0.7400	0.0443	−1	2	−0.0521
F/AS	Lightseekers	0.0885	170.5	−0.2472	−13	0	212.0	0.4356	−0.1139	−8	4	0.1434
F/AS	Minion Masters	0.1956	358	−0.1678	−10	2	509.5	0.3447	0.1227	−1	7	−0.1436
F/AS	Relic Hunters Zero: Remix	0.2820	256.5	−0.1502	−11	3	291.0	0.6754	−0.0592	−7	9	0.0717
F/AS	Train Valley 2	0.5769	392	−0.0735	−9	5	430.0	0.9939	0.0019	−4.5	6	−0.0023
TM	Lightseekers	0.2003	183.5	−0.1880	−8	1.5	223.5	0.7142	−0.0549	−6	4	0.0687
TM	Minion Masters	0.8789	421.5	−0.0207	−7	6	537.0	0.1106	0.2079	−1	5	−0.2430
TM	Relic Hunters Zero: Remix	0.9844	302.5	−0.0040	−6	6	289.5	0.7843	−0.0395	−4	4	0.0476
TM	Train Valley 2	0.1249	514.5	0.2022	−0.5	7.5	448.0	0.6213	0.0656	−2	4	−0.0769

We found no statistically significant differences, i.e., none of the game groups showed statistically significant improvements compared to the control group (all *p*-values > 0.05). However, some cases were closed to significance, such as Relic Hunters Zero: Remix for CT SUB ($p = 0.051$) or Train Valley 2 for CT SUB ($p = 0.076$). Furthermore, most effect sizes were small, ranging from −0.33 to +0.19, indicating minimal practical differences between groups. Concerning the sample sizes, the groups had reasonable sample sizes (control: 32–33 and game groups: 15–27 participants). However, the results showed that the differences are not statistically significant at the conventional $\alpha = 0.05$ level.

While participants averaged 18.1 h, skill gains might require more extended exposure. So, we conducted an analysis of dose–response effects (skill gain vs. total playtime) to provide deeper insight. So, we calculated skill gains for each participant across multiple tests. We performed Spearman and Kendall correlation analyses on skill gains versus hours played, and generated scatter plots with linear regression lines for overall and test-specific data (see Figures 6 and 7). If available, we added LOWESS trend lines to visualize relationships by test (see Figure 8).

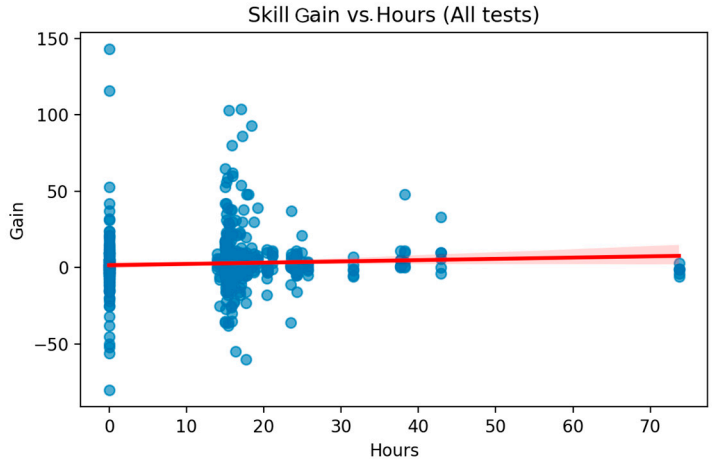

Figure 6. Skill gain vs. hours played (all tests).

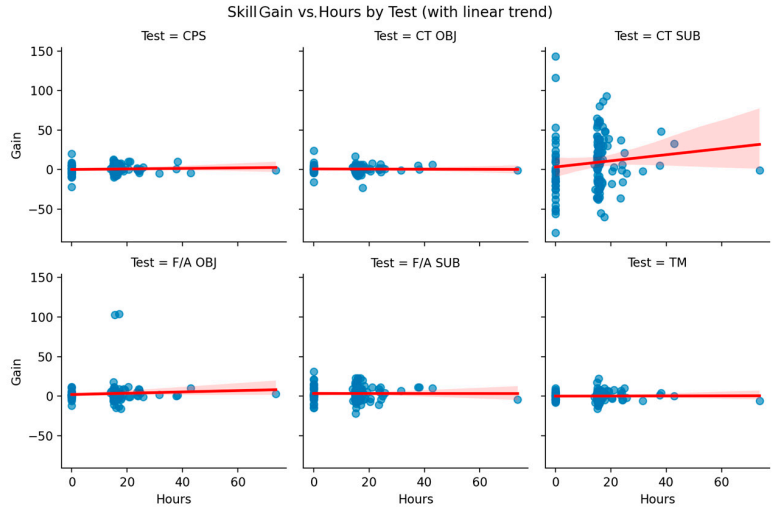

Figure 7. Skill gain vs. hours played by test.

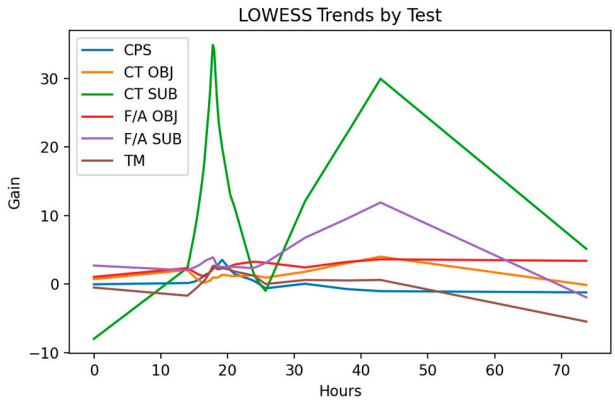

Figure 8. LOWESS trends by test.

In summary, the overall dose–response is weak. There is a small positive association across all tests combined, with borderline significance. Individual tests are mostly non-significant, except a hint for CT SUB. Furthermore, visuals show shallow trends and substantial spread. For most skill measures, extra hours did not reliably translate into bigger improvements. The one partial exception is CT SUB, where there is a slight hint that more hours might help, but even there, the relationship is modest and uncertain.

About efficiency these results imply several things: low marginal return, i.e., each additional hour buys very little average improvement, especially beyond the first dozen-or-so hours; diminishing predictability, i.e., the figures show that even if there is a small average effect, it is not consistent (some participants improve with more time, many do not); and opportunity cost, i.e., if the goal is to maximize skill gains per time invested, "more hours" is not an efficient strategy by itself.

For the design of the AI-based solution, we only chose the participants who played video games, not including participants in the control group. Participants who did not play for at least 10 h or who did not fill in the pre-test or post-test surveys were also excluded. Finally, 70 participants were considered to meet all the requirements.

The performance of AI models has been evaluated under different configurations to obtain a precise knowledge of the advantages and disadvantages of each model and the generalization capabilities. These evaluations have been accomplished by testing three models (Random Forest (RF), Gradient Boosting Regressor (GBR), Linear Regression (LR)) with five different arbitrary seeds that provide a description of the sensitivity of the initialization of each model and their generalization capabilities. The reason behind these seeds, instead of applying cross-validation or other techniques, is the low availability of data.

Derived from the scarcity of input data (approx. 30 users per game), a data augmentation technique is used to improve the model's stability and performance. This data augmentation technique consists of subtracting the "pre_test" values (i.e., Test-based user scores before playing) from the "post_test" and "pre_test" values, making the augmented samples start at zero but preserving the same delta (i.e., post_test-pre_test) value. This method also assumes a linear scale improvement when users play a video game, even if a linear scale might not faithfully represent a typical human learning curve; the scarcity and distribution of the data allow for generating more diverse samples through the data augmentation proposed in this section. Nevertheless, a thorough study of the learning curves for each soft skill will provide more accurate results in the future, once more data is available. However, even if the data augmentation technique employed evidence that increasing the number of samples will improve models' performance, there are some remarkable drawbacks, such as the data redundancy and the bias introduced to the models, that should be highlighted and carefully considered during the analysis of the results.

After the data augmentation has been applied to the dataset to meet the required minimum data, the models have then been trained with a train-test split of 20%. The performance metrics used for posterior analysis and validation are

- Coefficient of Determination (R^2): Indicates the proportion of variance in the observed data that is explained by the model, reflecting how good it fits. $1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2}$ where y_i are the observed values, $\hat{y}_i$ are the predicted values, $\overline{y}$ is the mean of the observed values, and n is the number of samples.
- Root Mean Squared Error (RMSE): Quantifies the standard deviation of the prediction errors, giving more weight to large deviations and highlighting models with big individual errors. $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

- Mean Absolute Error (MAE): Measures the average magnitude of the prediction errors, providing guidance on how close the predictions are to the observed values. $\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$

The configuration for each model included 100 estimators for both RF and GBR, which were selected to avoid overfitting, as more than 100 estimators could result in more estimators than data samples, which will be highly probable given the limited amount of data available, making overfitting a significant concern. The results of this performance evaluation can be observed in Figure 9.

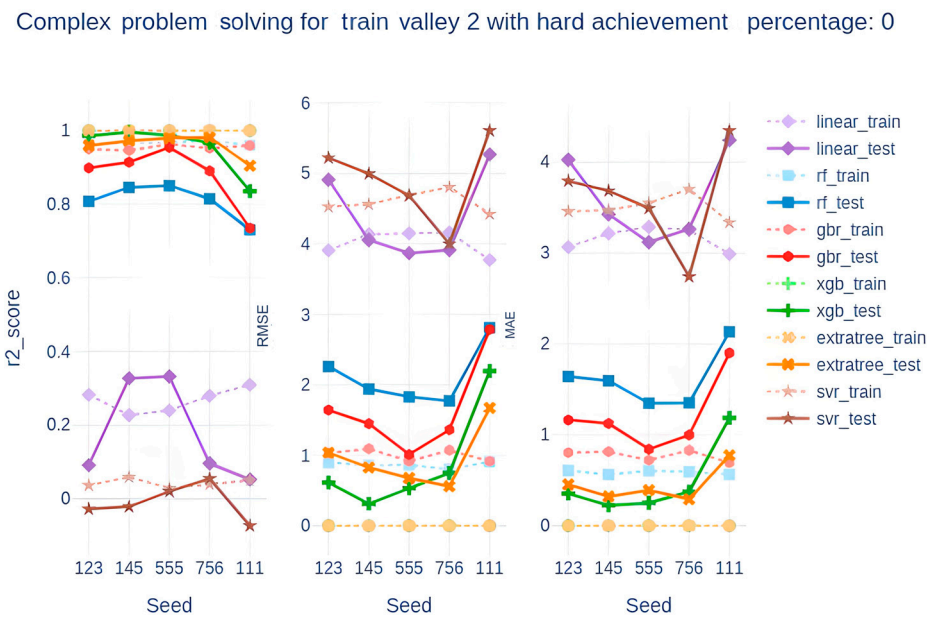

Figure 9. Performance of the model that predicts the improvement of soft skills. Evaluation includes the R^2 score, RMSE, and MAE metrics for train, test, and different seeds. The regression models are Linear Regressor (linear), Random Forest Regressor (rf), Gradient Boosting Regressor (gbr), Extreme Gradient Boosting (xgb), Extra Tree Regressor (extratree), and Support Vector Regressor (svr).

Figure 9 visualizes the performances of the three models in terms of R^2, MAE, and RMSE. R^2 reflects the ability to capture the improvement of soft skills given the user profile, and, ideally, should reach near 1. Meanwhile, both MAE and RMSE indicate the deviation of prediction from the real improvement in the same magnitude, as the data MAE is an indicator of the performance "on average" and RMSE, demonstrating the cases where the differences between predicted and real values are further apart. These, on the other hand, in the best case, are closer to 0.

What can be observed in Figure 9 is that the Linear Regressor model has poor performance with low R^2 values and high MAE and RMSE. This might suggest that the data that is being predicted is hard to grasp with simple linearity, and a more complex method might be required. That hypothesis contrasts greatly with the evidence of the performance of RF, GBR, XGB, and the supervised machine learning method for assembly ExtraTree, as all of them show high R^2 values and lower MAEs and RMSEs.

Figure 9 also visualizes the deviation of results given the initialization of different seeds, which can help to analyze the generalization capabilities of the model and the dependency on the data. Focusing on the best models (i.e., RF, GBR, XGB, and ExtraTree), RF and GBR, a deviation of around 0.1 to 0.2 in R^2 values can be observed, which prompts the idea of a high dependency on initialization. These deviations of R^2 values also reinforce the idea of a bigger training dataset, as more data can narrow down the deviation of R^2 given random initialization. With the aim of guaranteeing statistical significance of

the models conforming to this experimentation, a new set of experiments is conducted, now considering 60 seeds, which allows us to assess the stability and performance of the models. Table 8 shows the results of the conducted Wilcoxon statistical test for each pair of algorithms.

Table 8. Wilcoxon metric applied to each algorithm pair. Each cell is divided into four columns, one per soft skill (i.e., CT = critical thinking, FA = flexibility/adaptability, PS = complex problem solving, and TM = time management). Green dots represent $p < 0.005$ (significant differences), red crossesrepresent $p > 0.005$ (no significant differences).

	RF				GBR				XGB				EXTRATREE				SVR			
	CT	FA	PS	TM	CT	FA	PS	TM	CT	FA	PS	TM	CT	FA	PS	TM	CT	FA	PS	TM
LINEAR	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
RF	-	-	-	-	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●	●
GBR	-	-	-	-	-	-	-	-	●	●	●	●	×	×	×	×	●	●	●	●

As can be concluded from the analysis of the table, almost every pair of algorithms checks the null hypothesis, which makes them independent. However, ExtraTree and GBR models did not pass the test, which also showed very similar performances in terms of R^2. To help figure out how stable these models are, Table 9 presents the mean R^2 and standard deviation for each algorithm and soft skill.

Table 9. Mean R^2 and standard deviation for the 60-seed experiment (i.e., CT = critical thinking, FA = flexibility/adaptability, PS = complex problem solving, and TM = time management).

	LINEAR	RF	GBR	XGB	EXTRATREE	SVR
CT	0.055 ± 0.12	0.842 ± 0.07	0.914 ± 0.06	0.937 ± 0.08	0.905 ± 0.09	0.002 ± 0.06
FA	0.045 ± 0.13	0.640 ± 0.12	0.778 ± 0.11	0.847 ± 0.14	0.761 ± 0.14	-0.036 ± 0.04
PS	0.116 ± 0.17	0.770 ± 0.12	0.826 ± 0.14	0.911 ± 0.14	0.901 ± 0.09	-0.002 ± 0.04
TM	-0.066 ± 0.18	0.710 ± 0.14	0.747 ± 0.14	0.851 ± 0.15	0.851 ± 0.13	0.008 ± 0.06

Table 9 reinforces the previous conclusions: Linear and SVR models fail to capture the relationships in the data, whereas the other models do. Moreover, the small differences among them and the high R^2 values indicate low data variability—caused by data augmentation—and highlight the need for additional data to improve the generalization of the developed models.

Overall, Figure 8 and Tables 8 and 9 help to illustrate the complexity of the problem and the importance of the data, and, comparing the different models, either of the best models has resulted in acceptable performances, with EXTRATREE and XGB being the best-scoring models; R^2 is around 0.9, and both best MAE and best RMSE are around 1. Mean values for each metric are provided in Appendix B.

Finally, as an illustrative example, we present how the designed model was applied to two users who engaged with MEGASKILLS after its implementation. The different ranges of scores established for each test can be consulted in Appendix A.

One of these users was a beginner who had never played on the Steam platform, meaning no prior gameplay data was available. After registering on MEGASKILLS, this user played only one of the available games. Table 10 presents the model's output across several weeks and highlights the improvement in the soft skills of FA and time management, TM.

Table 10. Application of the model to a "novel" user, i.e., a user without previous data on Steam.

	Nº Games	Game Playtime (min)	Total Playtime (min)	Nº Achievements	CPS	CT	FA	TM
Start	1	0	0	0	0	0	0	0
Week1	1	726	726	6	44.08	32.02	21.89	57.59
Week2	1	982	982	9	44.08	32.02	23.34	59.04

The second case corresponds to a user with a prior Steam account who had logged over 1.500 h of gameplay (Steam level = 11) before engaging with the selected MEGASKILLS game. For this user, we started with baseline data, allowing the model to estimate their initial level of soft skills. Table 11 presents the model's output across several weeks, showing how the user's levels in FA and TM evolved and improved over time.

Table 11. Application of the model to an "experienced" user, i.e., a user with previous data on Steam.

	Nº Games	Game Playtime (min)	Total Playtime (min)	Nº Achievements	CPS	CT	FA	TM
Start	9	0	108,784	1528	40.40	31.48	21.43	55.20
Week1	10	605	109,289	1537	40.40	31.48	26.10	59.87
Week2	10	850	109,634	1538	40.40	31.48	28.81	62.57
Week3	10	921	109,705	1538	40.40	31.48	28.89	62.71

6. Discussion

Traditional approaches to soft skill assessment typically rely on standardized psychometric tests, self-report questionnaires, or observational evaluations conducted by educators or supervisors. While these methods provide structured and validated instruments, they often suffer from subjectivity, test-related anxiety, and limited ecological validity, as they capture performance in artificial or static contexts rather than in dynamic, behavior-rich environments. In contrast, the proposed AI-based assessment leverages real-time gameplay data to infer soft skill progression continuously and unobtrusively. The core research question, regarding the feasibility of measuring soft skill improvement through player gaming data, is affirmed by our findings. Compared to conventional tests, our AI-based solution, particularly the GBR model, demonstrated high efficacy ($R^2 \approx 0.9$, MAE/RMSE ≈ 1) in predicting soft skill increments from gaming data, validating this non-intrusive assessment approach. Moreover, the game-based approach offers a more engaging and scalable alternative, reducing participant fatigue and enabling longitudinal monitoring of skill development. Although traditional instruments remain useful for benchmarking and validation, the AI-driven video game assessment provides a complementary method that enhances measurement precision and learner motivation while maintaining methodological rigor.

Given that over half of the European population regularly plays video games, this medium offers a scalable, engaging, and non-intrusive alternative to traditional methods, effectively mitigating test-related anxiety. Their immersive environments mirror real-world professional challenges, fostering skills like problem solving and teamwork, with prior studies indicating measurable improvements after a minimum of 12–15 h of play. Our AI-based solution further validates this approach by effectively measuring these soft skill increments.

Our study reveals both promising directions and important nuances in the use of video games for soft skills development. Although our results partially corroborate earlier findings on the beneficial impact of video game engagement, they also highlight

the inherent challenges in accurately measuring and attributing these effects to specific gaming interventions.

The study aimed to explore the potential of commercial video games to enhance specific soft skills (CPS, CT, F/A, and TM) and to determine if an AI-based solution could measure improvement using player gaming data. Participants in the experimental groups played designated Steam games for a minimum of 15 h. The average time played by participants was 18.1 h, aligning with previous studies suggesting that 12–15 h may be necessary for significant improvement [95].

Despite observed trends toward improvement in mean pre- and post-intervention test scores, statistical analysis revealed no significant differences between the gaming groups and the control group (all p-values > 0.05), with most effect sizes being small.

In addition, we found that adding a few more hours is unlikely to move the needle much for most skills. Even larger increases in hours may yield only small average gains, and results vary a lot between people. If the goal is to boost skills meaningfully, simply telling participants to "play more" probably is not enough. More targeted practice (specific tasks/modes aligned to the skill), structured progression or feedback, or longer programs coupled with quality of practice might be necessary, not just quantity.

After internal discussions among the members of the MEGASKILLS project consortium, we concluded that the reasons for these results also could be caused by sample size limitations, and what we thought was the most critical issue, that the scales and subscales of standardized tests used required a lot of time and effort from the participants, who showed signs of fatigue, affecting the results, especially in the POST tests. We should think of ways of reducing the effort required from participants. Apart from the possible improvements identified in terms of study design, this highlights the complexity of measuring these skills with traditional methods and the potential need for further research to evaluate the impact of video games on soft skill development.

Crucially, this research developed and evaluated an AI-based approach using stealth assessment to measure soft skill improvement from game data. The core research question explored the feasibility of measuring soft skill improvement through players' gaming data. The results from training machine learning models (specifically RF and GBR) demonstrated a promising ability to predict the delta or increment in soft skill values based on user profiles (derived from their overall gaming history) and playtime. The GBR model showed particularly strong performance with an R^2 value around 0.9, along with low MAE and RMSE values, indicating it can effectively capture the relationship between gaming activity and soft skill improvement, even if this relationship is non-linear, as suggested by the poor performance of the LR model. However, it is worth highlighting once more that the scarcity of data, combined with data augmentation techniques, may cause overfitting, an issue that would be resolved when more real data becomes available.

This capability to predict skill improvement from game data directly supports the study's hypothesis regarding the measurability of soft skill development through video games. It offers a novel method for stealth assessment, which allows for the unobtrusive evaluation of skills integrated seamlessly into the gaming experience. This approach addresses a significant gap in soft skills assessment, which often suffers from a lack of standardized evaluation methods and inconsistent development. By leveraging game data like player behavior logs, response times, interaction data, success rates, achievements, and behavioral patterns, coupled with AI/machine learning models trained on this data, we can move towards more dynamic and context-sensitive measures of competency development.

The implications of these findings are broad, reaching across education, employment, and sustainability. Integrating video games with AI-powered stealth assessment into traditional training programs can offer a more engaging and effective method for enhancing

soft skills. Given that a significant portion of the European population plays video games regularly, this approach is highly scalable and can capitalize on established behavioral patterns. Strengthening soft skills is crucial for improving employability in a rapidly changing work environment where these skills are increasingly prioritized by employers alongside technical competencies. Furthermore, enhancing human capabilities like problem solving, critical thinking, and adaptability through such innovative methods is fundamental for addressing complex societal challenges and navigating sustainable development centered on the human being. Stealth assessment, by allowing continuous measurement without interrupting the learning or gaming experience, is conducive to skill acquisition and retention in engaging environments.

However, several limitations should be considered when interpreting these results. The AI model training was constrained by the scarcity of input data (approximately 30 users per game), necessitating data augmentation, which introduces potential drawbacks like redundancy and bias. The observed dependency of the AI model performance on initialization seeds also underscores the need for larger datasets to ensure robustness and generalization capabilities. Furthermore, while the application of the model to two illustrative user examples (novice and experienced) demonstrated its functionality, these limited cases are not intended to establish the overall generalization of the findings, which, as previously noted, is reliant on acquiring significantly larger datasets for future research. The challenge of precisely quantifying soft skill development remains, echoing the assessment challenges noted in our literature review. While the AI model can predict improvement, validating the scale and significance of this improvement against real-world applications requires further study. A significant limitation is that participants were not in a controlled environment and continued to engage in other activities where soft skills could be developed. Therefore, definitively attributing the observed soft skill improvements solely to video game play during the intervention period presents a challenge.

Future research should address the limitations identified in this study by

- Implementing longer intervention periods to better understand the temporal aspects of skill development and allow for potentially greater skill gains.
- Collecting significantly larger datasets to improve the robustness and generalization of the AI models used for stealth assessment.
- Conducting more detailed analyses of specific game mechanics within the selected or other commercial games and their relationship to the development of targeted soft skills.
- Developing more sophisticated measures and validation methods for assessing the transfer of soft skills developed in gaming environments to real-world professional and personal contexts.
- Investigating the role of player engagement, motivation, and individual differences in skill development through game-based learning.
- Further refining the AI methodology, including exploring different architectures or input features for the user and game embeddings to better capture nuanced gaming behaviors and their links to soft skills.

In conclusion, while our study supports the potential of commercial video games as tools for soft skills development and demonstrates the feasibility of using AI-based stealth assessment on game data to measure this improvement, it also emphasizes the need for a more nuanced understanding of how this development occurs and the conditions necessary to optimize it. The AI approach shows promise as a scalable, non-obtrusive method for soft skill assessment, paving the way for innovative educational and training interventions that leverage the widespread popularity of video games.

7. Conclusions

Our study yields crucial insights into the intricate relationship between commercial video games and soft skill development. Our findings could not confirm the positive impact of video games on soft skill enhancement, finding no statistically significant differences among the soft skills of gamers and non-gamers. Even more, the presented AI models showed limited reliability. The underlying mechanisms and contextual conditions are complex and require further research on how this development occurs and the conditions necessary to optimize it.

Our research successfully developed and evaluated an AI-based approach using stealth assessment to measure soft skill improvement directly from player gaming data. The results, particularly from the GBR model, demonstrated a promising ability to predict the increment in soft skill values. This capability supports the feasibility of measuring soft skill development through video games and presents the AI-based stealth assessment as a scalable, non-obtrusive, and dynamic method for skill evaluation integrated into the gaming experience.

Author Contributions: Conceptualization, J.B. and I.L.; Methodology, J.B., I.d.R. and I.L.; Software, J.B., I.d.R., A.M., A.A. and I.L.; Validation, J.B., I.L. and S.A.; Investigation, J.B., I.d.R., I.L. and S.A.; Resources, J.B. and I.d.R.; Data curation, J.B., I.L., A.M., A.A. and S.A.; Writing—original draft, J.B.; Writing—review and editing, I.L.; Visualization, A.M., A.A. and I.L.; Supervision, I.L. and S.A.; Project administration, J.B. All authors equally contributed to writing and reviewing this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union, MEGASKILLS—HORIZON-CL2-2022-TRANSFORMATIONS-01, grant number 101094275.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Advisor of the MEGASKILLS External Advisory Board, affiliated with the University of Strathclyde. The ethical and legal frameworks for the MEGASKILLS project were formally approved on 7 December 2023.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to privacy reasons.

Acknowledgments: The research team thanks the individuals who generously shared their time, experience, and materials for the purposes of this project. During the preparation of this manuscript, the authors used Napkin for the design of some figures, and ChatGPT for the revision of the comprehension of some texts in English. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Authors Juan Bartolomé, Idoya del Río, Aritz Martínez, Andoni Aranguren and Ibai Laña were employed by the company TECNALIA. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PRE	Pre-intervention
POST	Post-intervention
Obj	Objective
Sub	Subjective

Appendix A

Table A1. Ranges of scores established for each test.

Soft Skill	Test	Very Low	Low	Moderate	High	Very High
CPS	Reflective Thinking Skill Scale for Problem Solving	14–27	28–41	42–55	56–63	64–70
CT	Critical Thinking Assessment (SUB)		60–159	160–259	260–360	
CT	Test of Critical Thinking (OBJ)		0–15	16–25	26–45	
F/A	I-ADAPT-M (SUB)		55–128	129–202	203–275	
F/A	CAMBIOS (OBJ)		0–11	12–18	19–27	
TM	Time Management Questionnaire		18–41	42–65	66–90	

Appendix B

Table A2. Critical thinking performance metrics—mean across seeds (Lightseekers).

Model	Train_R^2	Test_R^2	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE
extratree	1.0	0.9051	0.0	1.2716	0.0	0.6673
gbr	0.9806	0.9141	0.614	1.2145	0.4543	0.8483
linear	0.2097	0.0549	3.956	4.3111	2.6747	3.0102
rf	0.9735	0.8424	0.7079	1.7153	0.4392	1.1583
svr	0.0514	0.0018	4.3317	4.4261	2.9084	3.0472
xgb	1.0	0.9375	0.0014	0.9176	0.0009	0.4411

Table A3. Time management performance metrics—mean across seeds (Minion Masters).

Model	Train_R^2	Test_R^2	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE
extratree	1.0	0.8505	0.0	2.0174	0.0	0.9537
gbr	0.9464	0.7473	1.3323	2.7669	0.9861	1.9595
linear	0.1626	−0.0662	5.2848	5.8661	4.0393	4.5063
rf	0.9603	0.7098	1.144	2.9992	0.7653	2.0933
svr	0.065	0.0078	5.5839	5.6846	4.1561	4.3284
xgb	1.0	0.851	0.0019	1.9605	0.0012	0.9859

Table A4. Flexibility/adaptability performance metrics—mean across seeds (Relic Hunters Zero Remix).

Model	Train_R^2	Test_R^2	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE
extratree	1.0	0.761	0.0003	4.0102	0.0	2.2805
gbr	0.9499	0.7783	1.9313	3.9277	1.4679	2.911
linear	0.2412	0.0454	7.5258	8.4144	6.1603	6.8591
rf	0.9499	0.6401	1.9279	5.1109	1.4221	3.8847
svr	0.008	−0.0362	8.607	8.8023	6.7411	6.9645
xgb	1.0	0.8474	0.004	3.0861	0.0026	2.0641

Table A5. Complex problem-solving performance metrics—mean across seeds (Train Valley 2).

Model	Train_R^2	Test_R^2	Train_RMSE	Test_RMSE	Train_MAE	Test_MAE
extratree	1.0	0.9015	0.0	1.3417	0.0	0.6102
gbr	0.9582	0.8262	0.9646	1.8605	0.7309	1.2738
linear	0.2777	0.1158	4.0275	4.44	3.1751	3.5527
rf	0.9689	0.7702	0.8302	2.2213	0.5761	1.5885
svr	0.037	−0.0016	4.6521	4.7545	3.4996	3.643
xgb	1.0	0.9108	0.0016	1.0959	0.001	0.5137

References

1. Cinque, M. "Lost in translation". Soft skills development in European countries. *Tuning J. High. Educ.* **2016**, *3*, 389–427. [CrossRef]
2. Kubátová, J.; Müller, M.; Kosina, D.; Kročil, O.; Slavíčková, P. *Soft Skills for the 21st Century: Defining a Framework for Navigating Human-Centered Development in an AI-Driven World*; Springer Nature: Berlin/Heidelberg, Germany, 2025.
3. Labobar, J.; Malatuny, Y.G. Soft Skill Development of Stakpn Sentani Christian Religious Education Graduates in the Workplace. In Proceedings of the 4th International Conference on Progressive Education 2022 (ICOPE 2022), Lampung, Indonesia, 15–16 October 2022; Atlantis Press: Dordrecht, The Netherlands, 2022; pp. 226–236. [CrossRef]
4. Sejzi, A.A.; Aris, B.; Yuh, C.P. Important soft skills for university students in 21th century. In Proceedings of the 4th International Graduate Conference on Engineering, Science, and Humanities (IGCESH 2013), Johor, Malaysia, 16–17 April 2013.
5. Schulz, B. The importance of soft skills: Education beyond academic knowledge. *J. Lang. Commun.* **2008**, 146–154.
6. Marin-Zapata, S.I.; Román-Calderón, J.P.; Robledo-Ardila, C.; Jaramillo-Serna, M.A. Soft skills, do we know what we are talking about? *Rev. Manag. Sci.* **2022**, *16*, 969–1000. [CrossRef]
7. Vasanthakumari, S. Soft skills and its application in work place. *World J. Adv. Res. Rev.* **2019**, *3*, 066–072. [CrossRef]
8. Tripathy, M. Relevance of Soft Skills in Career Success. *MIER J. Educ. Stud. Trends Pract.* **2020**, *10*, 91–102. [CrossRef]
9. Fletcher, S.; Thornton, K.R. The Top 10 Soft Skills in Business Today Compared to 2012. *Bus. Prof. Commun. Q.* **2023**, *86*, 411–426. [CrossRef]
10. Robles, M.M. Executive perceptions of the top 10 soft skills needed in today's workplace. *Bus. Commun. Q.* **2012**, *75*, 453–465. [CrossRef]
11. Abina, A.; Batkovič, T.; Cestnik, B.; Kikaj, A.; Kovačič Lukman, R.; Kurbus, M.; Zidanšek, A. Decision support concept for improvement of sustainability-related competences. *Sustainability* **2022**, *14*, 8539. [CrossRef]
12. Farao, C.; Bernuzzi, C.; Ronchetti, C. The crucial role of green soft skills and leadership for sustainability: A case study of an Italian small and medium enterprise operating in the food sector. *Sustainability* **2023**, *15*, 15841. [CrossRef]
13. Sujová, E.; Čierna, H.; Simanová, Ľ.; Gejdoš, P.; Štefková, J. Soft skills integration into business processes based on the requirements of employers—Approach for sustainable education. *Sustainability* **2021**, *13*, 13807. [CrossRef]
14. Seetha, N. Are soft skills important in the workplace? A preliminary investigation in Malaysia. *Int. J. Acad. Res. Bus. Soc. Sci.* **2014**, *4*, 44. [CrossRef] [PubMed]
15. Daly, S.; McCann, C.; Phillips, K. Teaching soft skills in healthcare and higher education: A scoping review protocol. *Soc. Sci. Protoc.* **2022**, *5*, 1–8. [CrossRef]
16. Ntola, P.; Nevines, E.; Qwabe, L.Q.; Sabela, M.I. A survey of soft skills expectations: A view from work integrated learning students and the chemical industry. *J. Chem. Educ.* **2024**, *101*, 984–992. [CrossRef]
17. Basir, N.M.; Zubairi, Y.Z.; Jani, R.; Wahab, D.A. Soft skills and graduate employability: Evidence from Malaysian tracer study. *Pertanika J. Soc. Sci. Humanit.* **2022**, *30*, 1975–1986. [CrossRef]
18. Touloumakos, A.K. Expanded yet restricted: A mini review of the soft skills literature. *Front. Psychol.* **2020**, *11*, 2207. [CrossRef]

19. Bhati, H.; Khan, P.; The importance of soft skills in the workplace. Journal of Student Research. 2022. Available online: https://www.jsr.org/hs/index.php/path/article/view/2764 (accessed on 12 October 2025).
20. Tadimeti, V. E-Soft Skills Training: Challenges and Opportunities. The IUP Journal of Soft Skills, 8, 34. Tadimeti, Vasundara, E-Soft Skills Training: Challenges and Opportunities. *IUP J. Soft Ski.* **2014**, *VIII*, 34–44.
21. Gibb, S. Soft skills assessment: Theory development and the research agenda. *Int. J. Lifelong Educ.* **2014**, *33*, 455–471. [CrossRef]
22. Cukier, W.; Hodson, J.; Omar, A. *"Soft" Skills Are Hard. A Review of the Literature 2015*; Ryerson University: Toronto, ON, Canada, 2015.
23. Kutz, M.R.; Stiltner, S. Program Directors' Perception of the Importance of Soft Skills in Athletic Training. *Athl. Train. Educ. J.* **2021**, *16*, 53–58. [CrossRef]
24. AbuJbara, N.A.K.; Worley, J.A. Leading toward new horizons with soft skills. *Horiz. Int. J. Learn. Futures* **2018**, *26*, 247–259. [CrossRef]
25. Dondi, M.; Klier, J.; Panier, F.; Schubert, J. *Defining the Skills Citizens Will Need in the Future World of Work*; McKinsey & Company: New York, NY, USA, 2021; Volume 25, pp. 1–19.
26. Kyllonen, P.C. Soft skills for the workplace. *Change Mag. High. Learn.* **2013**, *45*, 16–23. [CrossRef]
27. Snape, P. Enduring Learning: Integrating C21st Soft Skills through Technology Education. *Des. Technol. Educ. Int. J.* **2017**, *22*, 48–59.
28. Altomari, L.; Altomari, N.; Iazzolino, G. Gamification and Soft Skills Assessment in the Development of a Serious Game: Design and Feasibility Pilot Study. *JMIR Serious Games* **2023**, *11*, e45436. [CrossRef]
29. Nihal, K.S.; Pallavi, L.; Raj, R.; Babu, C.M.; Mishra, B. Enhancing Soft Skill Development with ChatGPT and VR: An Exploratory Study. In Proceedings of the 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), Chennai, India, 1–2 November 2023; pp. 1–6. [CrossRef]
30. European Commission: Directorate-General for Communications Networks, Content and Technology. *Understanding the Value of a European Video Games Society—Final Report*; Publications Office of the European Union: Luxembourg, 2023; Available online: https://data.europa.eu/doi/10.2759/332575 (accessed on 12 October 2025).
31. Barr, M. Video games can develop graduate skills in higher education students: A randomised trial. *Comput. Educ.* **2017**, *113*, 86–97. [CrossRef]
32. De Freitas, S. Are games effective learning tools? A review of educational games. *J. Educ. Technol. Soc.* **2018**, *21*, 74–84.
33. Qian, M.; Clark, K.R. Game-based Learning and 21st century skills: A review of recent research. *Comput. Hum. Behav.* **2016**, *63*, 50–58. [CrossRef]
34. McGowan, N.; López-Serrano, A.; Burgos, D. Serious games and soft skills in higher education: A case study of the design of compete! *Electronics* **2023**, *12*, 1432. [CrossRef]
35. Bezanilla, M.J.; Arranz, S.; Rayón, A.; Rubio, I.; Menchaca, I.; Guenaga, M.; Aguilar, E. A proposal for generic competence assessment in a serious game. *J. New Approaches Educ. Res.* **2014**, *3*, 42–51. [CrossRef]
36. Connolly, T.M.; Boyle, E.A.; MacArthur, E.; Hainey, T.; Boyle, J.M. A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* **2012**, *59*, 661–686. [CrossRef]
37. Krath, J.; Schürmann, L.; Von Korflesch, H.F. 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]
38. Subhash, S.; Cudney, E.A. Gamified learning in higher education: A systematic review of the literature. *Comput. Hum. Behav.* **2018**, *87*, 192–206. [CrossRef]
39. Zhonggen, Y. A meta-analysis of use of serious games in education over a decade. *Int. J. Comput. Games Technol.* **2019**, *2019*, 4797032. [CrossRef]
40. Sutil-Martín, D.L.; Otamendi, F.J. Soft skills training program based on serious games. *Sustainability* **2021**, *13*, 8582. [CrossRef]
41. Clark, D.B.; Tanner-Smith, E.E.; Killingsworth, S.S. Digital games, design, and learning: A systematic review and meta-analysis. *Rev. Educ. Res.* **2016**, *86*, 79–122. [CrossRef]
42. Granic, I.; Lobel, A.; Engels, R.C. The benefits of playing video games. *Am. Psychol.* **2014**, *69*, 66. [CrossRef]
43. Plass, J.L.; Homer, B.D.; Kinzer, C.K. Foundations of game-based learning. *Educ. Psychol.* **2015**, *50*, 258–283. [CrossRef]
44. Squire, K. From content to context: Videogames as designed experience. *Educ. Res.* **2006**, *35*, 19–29. [CrossRef]
45. Ryan, R.M.; Deci, E.L. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* **2000**, *55*, 68. [CrossRef]
46. Eseryel, D.; Law, V.; Ifenthaler, D.; Ge, X.; Miller, R. An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *J. Educ. Technol. Soc.* **2014**, *17*, 42–53.
47. Shute, V.J.; Wang, L.; Greiff, S.; Zhao, W.; Moore, G. Measuring problem solving skills via stealth assessment in an engaging video game. *Comput. Hum. Behav.* **2016**, *63*, 106–117. [CrossRef]

48. Lie, A.; Stephen, A.; Supit, L.R.; Achmad, S.; Sutoyo, R. Using strategy video games to improve problem solving and communication skills: A systematic literature review. In Proceedings of the 2022 4th International Conference on Cybernetics and Intelligent System (ICORIS), Prapat, Indonesia, 8–9 October 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5. [CrossRef]
49. Baniqued, P.L.; Kranz, M.B.; Voss, M.W.; Lee, H.; Cosman, J.D.; Severson, J.; Kramer, A.F. Cognitive training with casual video games: Points to consider. *Front. Psychol.* **2014**, *4*, 1010. [CrossRef]
50. Barr, M. *Graduate Skills and Game-Based Learning: Using Video Games for Employability in Higher Education*; Springer Nature: Berlin/Heidelberg, Germany, 2019.
51. Mao, W.; Cui, Y.; Chiu, M.M.; Lei, H. Effects of game-based learning on students' critical thinking: A meta-analysis. *J. Educ. Comput. Res.* **2022**, *59*, 1682–1708. [CrossRef]
52. Pusey, M.; Wong, K.W.; Rappa, N.A. Resilience interventions using interactive technology: A scoping review. *Interact. Learn. Environ.* **2022**, *30*, 1940–1955. [CrossRef]
53. Glass, B.D.; Maddox, W.T.; Love, B.C. Real-time strategy game training: Emergence of a cognitive flexibility trait. *PLoS ONE* **2013**, *8*, e70350. [CrossRef]
54. Strobach, T.; Frensch, P.A.; Schubert, T. Video game practice optimizes executive control skills in dual-task and task switching situations. *Acta Psychol.* **2012**, *140*, 13–24. [CrossRef]
55. Shute, V.J. Stealth assessment in computer-based games to support learning. *Comput. Games Instr.* **2011**, *55*, 503–524.
56. Rahimi, S.; Shute, V.J. Stealth assessment: A theoretically grounded and psychometrically sound method to assess, support, and investigate learning in technology-rich environments. *Educ. Technol. Res. Dev.* **2024**, *72*, 2417–2441. [CrossRef]
57. Yannakakis, G.N.; Togelius, J. *Artificial Intelligence and Games*; Springer: New York, NY, USA, 2018; Volume 2, pp. 1502–2475.
58. Swiecki, Z.; Khosravi, H.; Chen, G.; Martinez-Maldonado, R.; Lodge, J.M.; Milligan, S.; Selwyn, N.; Gašević, D. Assessment in the age of artificial intelligence. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100075. [CrossRef]
59. Fragoso, L.; Stanley, K.G. Stable: Analyzing player movement similarity using text mining. In Proceedings of the 2021 IEEE Conference on Games (CoG), Copenhagen, Denmark, 17–20 August 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1–8. [CrossRef]
60. Melo, S.A.; Kohwalter, T.C.; Clua, E.; Paes, A.; Murta, L. Player behavior profiling through provenance graphs and representation learning. In Proceedings of the 15th International Conference on the Foundations of Digital Games 2020, Bugibba, Malta, 15–18 September 2020; pp. 1–11. [CrossRef]
61. Lee, H.; Lee, S.; Nallapati, R.; Uh, Y.; Lee, B. Characterizing and Quantifying Expert Input Behavior in League of Legends. In Proceedings of the CHI Conference on Human Factors in Computing Systems 2024, Honolulu, HI, USA, 11–16 May 2024; pp. 1–21. [CrossRef]
62. Thawonmas, R.; Iizuka, K. Visualization of online-game players based on their action behaviors. *Int. J. Comput. Games Technol.* **2008**, *2008*, 906931. [CrossRef]
63. Dye, M.W.; Green, C.S.; Bavelier, D. Increasing speed of processing with action video games. *Curr. Dir. Psychol. Sci.* **2009**, *18*, 321–326. [CrossRef] [PubMed]
64. Jordan, T.; Dhamala, M. Enhanced dorsal attention network to salience network interaction in video gamers during sensorimotor decision-making tasks. *Brain Connect.* **2023**, *13*, 97–106. [CrossRef]
65. McDermott, A.F.; Bavelier, D.; Green, C.S. Memory abilities in action video game players. *Comput. Hum. Behav.* **2014**, *34*, 69–78. [CrossRef]
66. Mack, D.J.; Wiesmann, H.; Ilg, U.J. Video game players show higher performance but no difference in speed of attention shifts. *Acta Psychol.* **2016**, *169*, 11–19. [CrossRef]
67. Steinkuehler, C.; Duncan, S. Scientific habits of mind in virtual worlds. *J. Sci. Educ. Technol.* **2008**, *17*, 530–543. [CrossRef]
68. Bailey, E.; Miyata, K. Improving video game project scope decisions with data: An analysis of achievements and game completion rates. *Entertain. Comput.* **2019**, *31*, 100299. [CrossRef]
69. Lopez-Gordo, M.A.; Kohlmorgen, N.; Morillas, C.; Pelayo, F. Performance prediction at single-action level to a first-person shooter video game. *Virtual Real.* **2021**, *25*, 681–693. [CrossRef]
70. Sapienza, A.; Zeng, Y.; Bessi, A.; Lerman, K.; Ferrara, E. Individual performance in team-based online games. *R. Soc. Open Sci.* **2018**, *5*, 180329. [CrossRef]
71. Cretenoud, A.F.; Barakat, A.; Milliet, A.; Choung, O.H.; Bertamini, M.; Constantin, C.; Herzog, M.H. How do visual skills relate to action video game performance? *J. Vis.* **2021**, *21*, 10. [CrossRef] [PubMed]
72. Karmakar, B.; Liu, P.; Mukherjee, G.; Che, H.; Dutta, S. Improved retention analysis in freemium role-playing games by jointly modelling players' motivation, progression and churn. *J. R. Stat. Soc. Ser. A Stat. Soc.* **2022**, *185*, 102–133. [CrossRef]
73. Lange, A.; Somov, A.; Stepanov, A.; Burnaev, E. Building a behavioral profile and assessing the skill of video game players. *IEEE Sens. J.* **2021**, *22*, 481–488. [CrossRef]
74. Zhao, S.; Xu, Y.; Luo, Z.; Tao, J.; Li, S.; Fan, C.; Pan, G. Player behavior modeling for enhancing role-playing game engagement. *IEEE Trans. Comput. Soc. Syst.* **2021**, *8*, 464–474. [CrossRef]

75. Hou, H.T. Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Comput. Hum. Behav.* **2015**, *48*, 424–435. [CrossRef]
76. Kang, J.; Liu, M.; Qu, W. Using gameplay data to examine learning behavior patterns in a serious game. *Comput. Hum. Behav.* **2017**, *72*, 757–770. [CrossRef]
77. Han, Y.J.; Moon, J.; Woo, J. Prediction of churning game users based on social activity and churn graph neural networks. *IEEE Access* **2024**, *12*, 101971–101984. [CrossRef]
78. Loh, C.S.; Sheng, Y. Measuring expert performance for serious games analytics: From data to insights. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 101–134. [CrossRef]
79. Marczak, R.; Schott, G.; Hanna, P. Postprocessing gameplay metrics for gameplay performance segmentation based on audiovisual analysis. *IEEE Trans. Comput. Intell. AI Games* **2014**, *7*, 279–291. [CrossRef]
80. Tlili, A.; Chang, M.; Moon, J.; Liu, Z.; Burgos, D.; Chen, N.S. A systematic literature review of empirical studies on learning analytics in educational games. *Int. J. Interact. Multimed. Artif. Intell.* **2021**, *7*, 250–261. [CrossRef]
81. Afonso, A.P.; Carmo, M.B.; Gonçalves, T.; Vieira, P. VisuaLeague: Player performance analysis using spatial-temporal data. *Multimed. Tools Appl.* **2019**, *78*, 33069–33090. [CrossRef]
82. El-Nasr, M.S.; Drachen, A.; Canossa, A. *Game Analytics*; Springer: Berlin/Heidelberg, Germany, 2016; p. 13.
83. Min, W.; Frankosky, M.H.; Mott, B.W.; Rowe, J.P.; Smith, A.; Wiebe, E.; Boyer, K.E.; Lester, J.C. DeepStealth: Game-based learning stealth assessment with deep neural networks. *IEEE Trans. Learn. Technol.* **2019**, *13*, 312–325. [CrossRef]
84. Smerdov, A.; Somov, A.; Burnaev, E.; Zhou, B.; Lukowicz, P. Detecting video game player burnout with the use of sensor data and machine learning. *IEEE Internet Things J.* **2021**, *8*, 16680–16691. [CrossRef]
85. de Almeida Rocha, D.; Duarte, J.C. Simulating human behaviour in games using machine learning. In Proceedings of the 2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames), Rio de Janeiro, Brazil, 28–31 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 163–172. [CrossRef]
86. Mustač, K.; Bačić, K.; Skorin-Kapov, L.; Sužnjević, M. Predicting player churn of a Free-to-Play mobile video game using supervised machine learning. *Appl. Sci.* **2022**, *12*, 2795. [CrossRef]
87. Gee, J.P. What video games have teach us about learning and literacy. *Comput. Entertain. (CIE)* **2003**, *1*, 20. [CrossRef]
88. Frommel, J.; Phillips, C.; Mandryk, R.L. Gathering self-report data in games through NPC dialogues: Effects on data quality, data quantity, player experience, and information intimacy. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Okohama, Japan, 8–13 May 2021; pp. 1–12. [CrossRef]
89. Stafford, T.; Dewar, M. Tracing the trajectory of skill learning with a very large sample of online game players. *Psychol. Sci.* **2014**, *25*, 511–518. [CrossRef]
90. Kizilkaya, G.; Askar, P. The development of a reflective thinking skill scale towards problem solving. *Egit. Ve Bilim* **2009**, *34*, 82.
91. Payan-Carreira, R.; Sacau-Fontenla, A.; Rebelo, H.; Sebastião, L.; Pnevmatikos, D. Development and Validation of a Critical Thinking Assessment-Scale Short Form. *Educ. Sci.* **2022**, *12*, 938. [CrossRef]
92. Bracken, B.A.; Bai, W.; Fithian, E.; Lamprecht, M.S.; Little, C.; Quek, C. *The Test of Critical Thinking*; Center for Gifted Education, College of William y Mary: Williamsburg, VA, USA, 2003.
93. Burke, C.S.; Pierce, L.G.; Salas, E. (Eds.) *Understanding Adaptability: A Prerequisite for Effective Performance Within Complex Environments*; JAI Press: Stamford, CT, USA, 2006.
94. Seisdedos, N. *CAMBIOS Test de Flexibilidad Cognitiva*; TEA Ediciones: Madrid, Spain, 2004.
95. Britton, B.K.; Tesser, A. Effects of time-management practices on college grades. *J. Educ. Psychol.* **1991**, *83*, 405–410. [CrossRef]
96. Oei, A.C.; Patterson, M.D. Enhancing cognition with video games: A multiple game training study. *PLoS ONE* **2013**, *8*, e58546. [CrossRef]

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.