

Article

Usability Evaluation of an Adaptive Serious Game Prototype Based on Affective Feedback

Lampros Karavidas , Hippokratis Apostolidis  and Thrasyvoulos Tsiatsos *

Department of Informatics, Aristotle University of Thessaloniki, GR-54124 Thessaloniki, Greece

* Correspondence: tsiatsos@csd.auth.gr; Tel.: +30-231-0998990

Abstract: Difficulty in video games is an essential factor for a game to be considered engaging and is directly linked to losing in a game. However, for the user to not feel bored or frustrated, it is necessary for the difficulty of the game to be balanced and ideally tailored to the user. This paper presents the design and development of a serious game that adjusts its difficulty based on the user's bio signals, so that it is demanding enough to match his/her skills, in order to enter the flow state. The serious game is accompanied by a server that uses machine learning algorithms to analyze the user's bio signals and classify them into different affective states. These states are later used to adjust the difficulty of the serious game in real-time, without interfering with the user's game experience. Finally, a heuristic evaluation was conducted in order to measure its usability and highlight the good practices and to draw attention to some elements of the game that should be changed in a future version.

Keywords: dynamic difficulty adjustment; affective computing; heuristic evaluation



Citation: Karavidas, L.; Apostolidis, H.; Tsiatsos, T. Usability Evaluation of an Adaptive Serious Game Prototype Based on Affective Feedback. *Information* **2022**, *13*, 425. <https://doi.org/10.3390/info13090425>

Academic Editors: Claude Frasson, Petros Lameris, Gabriele Gianini, Sylvester Arnab and Panagiotis Petridis

Received: 14 July 2022

Accepted: 5 September 2022

Published: 8 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

Technology-assisted education has been a clear trend over the past decade. As a result, serious games, video games that also have other goals apart from pure entertainment, are used more and more frequently. Serious games have been found to be effective in education [1]. Their success has been, at times, linked to the fact that students tend to be more engaged in learning when serious games are involved [2].

Engagement has, at times, been linked with notions such as flow and fun [3]. Engagement is defined as a player's loyalty to a game's activities and challenges. Deeply engaged users usually are so focused in the game that they do not perceive how fast time passes and are unaware of their surroundings while playing [4].

Beating the opponent, who either is another person or an artificial one, is also linked to the notion of engagement [5]. Particularly, when the player manages to win under difficult circumstances, the joy they experience is greater as they realize that they can outsmart their opponent.

Consequently, difficulty in video games is an essential factor for a game to be considered engaging and is directly linked to losing in a game. The difficulty level of a game can be affected by many factors; however, it is mostly directly linked to the user's inability to beat an opponent, who might be the video game or another player [6]. Losing in a game is perceived as the directly opposite state of winning. Players do not want to lose in a game as they feel inadequate but at the same time, a defeat leads the player to reconsider the strategy that he/she already used, making the actual game much more interesting [6]. Moreover, the joy the user experiences from playing a game is mostly affected by its difficulty [7]. In order for it to be as joyful as it can be, the difficulty should be balanced [6] in a way that the player should not feel bored, if the game is way too easy, or frustrated, when the game is too challenging. In other words, the goal is to achieve the flow state (Figure 1), as proposed by [8].

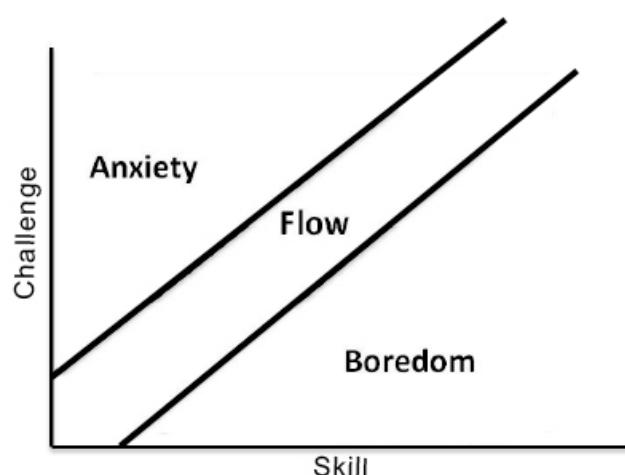


Figure 1. Flow state diagram.

To achieve this, the difficulty should be tailored to each player's skills, and it can be feasible using some dynamic difficulty adjustment methods for each player. The term dynamic difficulty adjustment requires the game to automatically detect the user's skills and instantly adjust to it in a way that is not easily noticeable. In addition, the game should keep monitoring the player's level throughout the game to ensure that it remains challenging for him/her [9]. In the past, there have been many methods that were used to achieve the dynamic difficulty adjustment. An unobtrusive way of realizing the user's state in a game is by monitoring his/her emotions through his/her bio signals [10]. Specifically, the player's electroencephalogram might be used to measure his/her excitement during the game and when it is lower than a certain boundary, a different difficulty level can be used in the game.

Our paper focuses on the use of affective computing to adjust the game's difficulty. Affective computing is the scientific field that utilizes emotion recognition techniques. It can be defined as "computing that relates to, arises from, or deliberately influences emotions" [11]. Due to the complexity of the human nature, it is quite challenging to describe and emulate an emotion. As a result, many scientific branches are required, such as psychology, physiology, sociology, mathematics, computer science, education and finally linguistics, to accomplish its goals [12].

The aim of this study was to present the computational system that was designed and developed to help students enhance their time-management skills by playing a game that adjusted its difficulty to their emotional state and test the usability of the serious game in question by a group of experts.

2. Related Work

Affective computing, used as a basis for dynamic difficulty adjustment by utilizing psychophysiological measurements, has been of great interest to many researchers in the past. Stein et al. [10] used electroencephalogram (EEG) signals to adjust a game's difficulty in real-time. The game that was used for this purpose was called Bootcamp and the tool to gather EEG measurements was Emotiv EPOC. Through this research, it was found that players reacted in a certain way, had certain measurements in other words, for some events and this piece of information was used to adjust the difficulty of the game. By using dynamic difficulty adjustment (DDA), the players answered in questionnaires that the game was more fun and the new EEG measurements had the same result and even though the DDA method was noticeable as the opponents received some powerups, the player said that they were not dissatisfied.

A few years earlier, Liu et al. [13] used some other bio-signals, such as heart-rate variability (HRV), galvanic skin response (GSR) and electromyography results (EMG), from several face muscles in order to determine one's emotional state. In the first stage, they

used various anagrams and played Pong to determine the usual anxiety level for each player. Later, they used DDA based on the player's anxiety while they played pong. The results showed that six out of nine people were less anxious while playing the DDA version of the game and seven participants improved their performance. They also suggested the use of other channels to recognize anxiety, such as body posture and eye movement.

A serious game that utilizes dynamic difficulty adjustment was later developed by Ninaus et al. [14], who used heart rate data to adapt the difficulty of the game to the user. The serious game "Emergency" was created to train emergency personnel by showcasing critical scenarios. The users, by playing the game, would learn how to react in certain occasions and prevent damage following the incident. The game consisted of three different scenarios, with a certain level of difficulty corresponding to each of them. The difficulty was a result of the tasks that the user should manage simultaneously. While playing the game, the heart rate data of each player were gathered to assess the player's level of arousal and provide an easier or harder scenario, to keep the player engaged. The adaptive version of the game was found to have a higher completion rate and the users described it as more challenging and fascinating.

Another serious game called "PERSON" was designed to create personalized therapies for maintaining or enhancing cognitive abilities that are being compromised by aging or disorders of the brain, as presented by Monaco et al. [15]. The initial profile with the baseline for every user was originally created from preliminary cognitive tests and through analysing EEG data, the learning level for every user was deduced, allowing the game difficulty to increase. Through technical evaluation, it was prominent that the volunteers showed great interest for the experiment and found the game enjoyable.

3. Method

This section presents the methodology followed in order to examine our research question.

3.1. Participants

Six experts were included in the experiment and filled out the questionnaire with the heuristic rules after playing the serious game. The mean age of the participants was 39.66 years (SD = 10.3). Before the activities, all participants signed a consent form.

3.2. Materials

The materials used to answer our research questions are presented in the paragraphs below.

3.2.1. System Architecture

The Affective Kitchen is an adaptive game that considers the user's affective state as a trigger to activate the corresponding adaptations. The user's real time affective state is detected utilizing the galvanic skin response (GSR) or electrodermal activity (EDA) biosignal.

The architecture of the real time affective state detection system is depicted in Figure 2. The user is connected to a bitalino GSR sensor (<https://bitalino.com/>, accessed on 10 July 2022) [16], with a sampling rate of 100 Hz and 10 bit resolution. The main parts of the bitalino kit include a microcontroller (MCU) for analog to digital conversion, an EDA sensor and a Bluetooth card for wireless communication.

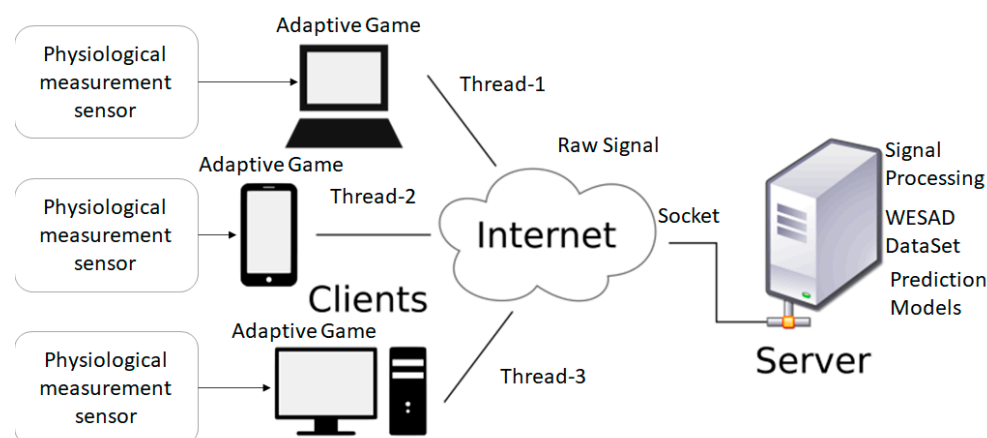


Figure 2. The affective recognition system architecture.

When the signal acquisition process begins, the raw signals are transferred to the Affective Kitchen application. This application, as a client, joins to a server application thread and sends the raw signal via a TCP socket.

The server application is implemented in Python programming language and utilizes the open source libraries, “biosignalsnotebooks”, “sklearn”, “numpy”, “socket”, “threading” and “mysql”. At first, the raw signal is converted to uS, following the formula $((\text{RawValue}/1024) \times 3.3)/0.12$ [16]. Then, the data are normalized by applying the Min-MaxScalar object of the sklearn library. Then, we apply conventional 1st order Butterworth bandpass and conventional 2nd order Butterworth lowpass filtering. After this filtering process, the signal is successful in smoothing, but hand moving oscillations are still kept [17]. Thus, to remove the high-frequency noise components, we apply an algorithm called stationary wavelet transform (SWT), proposed by Chen et al. [18]. This algorithm is proposed in the bibliography, as a way to minimize the influence of motion artifacts in the EDA signal [16].

Then, we proceeded to the Gaussian mixture model, which can be described as follows: “One Gaussian component describes coefficients centered around zero, and the other describes those spread out at larger values. The Gaussian with smaller variance corresponds to the wavelet coefficients of Skin Conductance Level (SCL), while the Gaussian with larger variance corresponds to the wavelet coefficients of Skin Conductance Responses (SCRs)” [18].

In the phase of feature extraction of the collected signals, we break the continuous collected data into intervals with a sliding window with length (5 s). Then, the following features are extracted:

- Mean SCR, (F1);
- Max SCR, (F2);
- Min SCR, (F3);
- Range SCR, (F4);
- Skewness SCR, (F5);
- Kurtosis SCR (F6).

These features form vectors in the form $\langle F1, F2, F3, F4, F5, F6, \text{ClassValue} \rangle$. The Class-Value is the unknown user’s affective state. This vector will be supplied to a supervised learning algorithm to predict the user’s affective state.

At the beginning step of the server application execution, a classification model, implementing the KNN machine learning algorithm, is loaded.

3.2.2. The Classification Model

This model is trained utilizing a publicly available dataset [19], based on a combination between the “wearable stress and affect detection” (WESAD) [20] and the “SWELL

knowledge worker dataset" (SWELL-KW) [21,22]. The derived KNN model had a classification accuracy of 93% with 12 neighbors. The number of neighbors was provided after applying the grid search algorithm. The Python library sklearn includes GridSearchCV to find the best value of K neighbors from the range of values and to apply cross validation. We applied 10-fold cross validation. The derived accuracy is in line with other relevant research works. Bajpai and He [23] applied the KNN model with 90% accuracy, classifying the WESAD dataset and adjusting the number of nearest neighbors to the appropriate value. Moreover, Aqajari et al. [24] stated in their article that they could detect stress with an accuracy of 92%. They applied the KNN model, utilizing the WESAD dataset.

The derived confusion matrix, showing the true positive, false positive, true negative and false negative predicted values, is depicted in Figure 3.

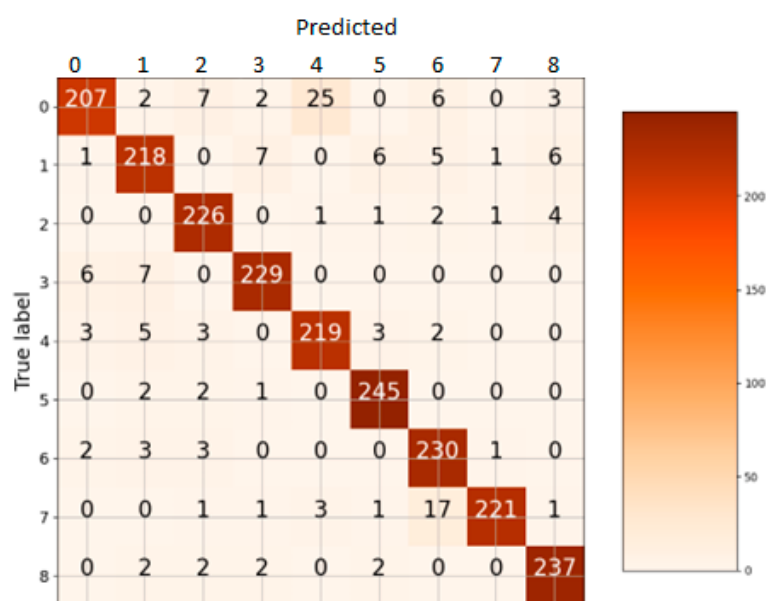


Figure 3. Confusion matrix.

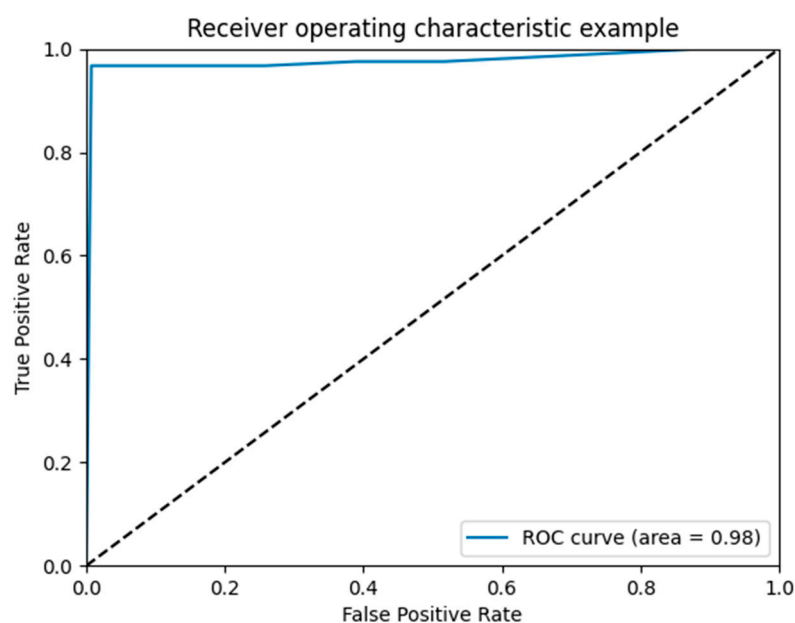


Figure 4. ROC curve (area under the ROC curve: 0.976666).

The ROC curve (receiver operating characteristic curve) is presented in Figure 4. The ROC curve shows the probability curve and the area under the curve (AUC) is a criterion of the separability of the applied model. Thus, the higher the AUC, the better the model is at predicting the classes.

When the server application is loaded, the KNN model is also loaded. Then, during the execution of the server application, every feature vector is formed and is utilized by the KNN prediction model to predict the ClassValue. The derived ClassValue corresponds to an affective state and a level of intensity according to the dataset used to train the prediction model. The dataset used [18] consists of three affective states. Every affective state is presented by three levels. The total of the classes of the affective states is nine and they are described below.

- Baseline
 - Low, class: 0—close to boredom;
 - Medium, class: 1—also close to boredom;
 - High, class: 2—close to neutral.
- Amusement
 - Low, class: 3—flow;
 - Medium, class: 4—flow;
 - High, class: 5—flow.
- Anxiety
 - Low, class: 6—low anxiety;
 - Medium, class: 7—anxiety;
 - High, class: 8—anxiety.

The classification model metrics are shown in Table 1.

Table 1. KNN classification model metrics.

Class	Precision	Recall	F1-Score
0	0.95	0.82	0.88
1	0.91	0.89	0.90
2	0.93	0.96	0.94
3	0.95	0.95	0.95
4	0.88	0.93	0.91
5	0.95	0.98	0.96
6	0.88	0.96	0.92
7	0.99	0.90	0.94
8	0.94	0.97	0.96
K Neighbors accuracy score: 0.9291266575217193			

3.2.3. The Serious Game

The Affective Kitchen is designed to be a time-management, point and click, two-dimensional serious game that is available as a computer and a smartphone application. The game was developed using the Unity game engine. By playing the game, the user is meant to learn how to perform well and not panic if (s)he has to complete several tasks in a certain amount of time and under pressure. By improving their time-management skills the students will perform better in multiple choice tests (MCT's). MCT's have been a valuable assessment tool the over last few decades due to the great number of benefits they come with, such as the fact that they are impartial, are supported by virtual learning environments and it is possible to provide the students with instant feedback, turning them into a great way for professors to support their students' learning [25]. However, Hlasny [26] has found that students tend to use the available time ineffectively while taking an MCT, completing the test before the time runs out. These students gravitate towards having lower scores compared to the others.

Once the user opens the application, (s)he must register with a username and a password and then use his/her credentials to enter the game. Once (s)he is logged in, the user is sent to the Main Menu screen, where (s)he is able to “Play” the game, open the “Settings” menu to change the volume and enable or disable the accompanying sensor (detailed description in Section 3.2.1), thus enabling or disabling the adaptivity component. By pressing the “Play” button, the player can start playing the game, as depicted in Figure 5.

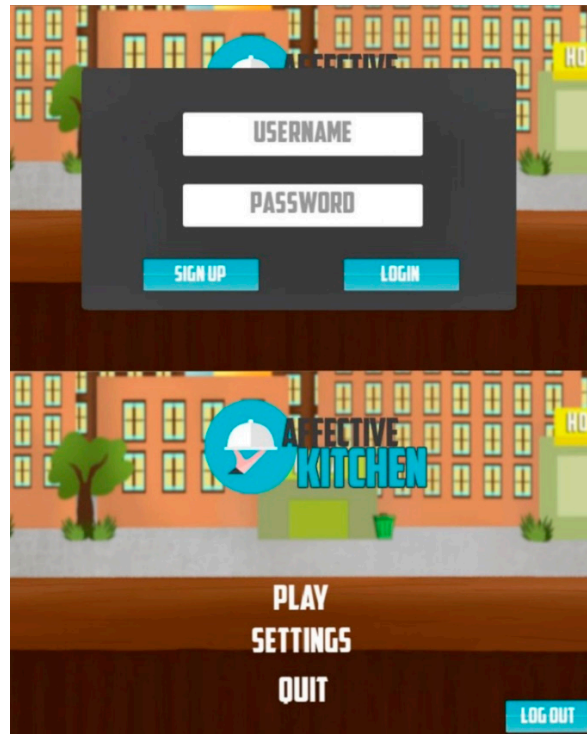


Figure 5. Login and Main Menu screen.



Figure 6. Screen of the gameplay.

On the screen of the main game, the user can find four different food items and an oven. Next to the oven, there is a book called “My Recipes” that the player can use to find the different combinations and the order that the food items should be put in the oven, to produce a specific outcome. Once the user places all the items in the oven, (s)he has to wait for the dish to be produced. Every few seconds, a new customer is spawned with a desire for a specific dish and a bar below him/her that indicates the time left until (s)he leaves,

as seen in Figure 6. If the customer leaves without being served in a specific amount of time, one of the red hearts on the top left of the screen vanishes, until there is none left and the game is over for the user. For every customer that is served, the user earns a certain amount of coins, while for every lost customer, some coins are deducted.

In the case that the user has chosen to play the game with the adaptivity component, in order to keep the user engaged and not feel either bored or anxious, based on the theoretical model of the flow state [8], a parameter of the game is modified according to his/her emotional state, as found in Section 3.2.1. The element that instantly affects the difficulty of the game and is the one that changes according to the player's emotions is the rate of decrease in the waiting time of the customer. In particular, if the integer that the game receives from the server is greater than six, this corresponds to an affective state that relates to anxiety, which might lead to the frustration of the player and him/her quitting the game. For this reason, to prevent him/her from leaving the game, the rate of decrease in the waiting time is reduced. When the integer that represents the affective state is lower than three, the value that affects the decrease in time increases, so the customers' duration on screen is smaller. An affective state below three indicates a bored player, so this is an attempt to challenge him/her. If the value is between six and three, then no changes are made.

When the user loses the game, then (s)he proceeds to the final screen where (s)he finds the amount of coins (s)he gathered and where (s)he stands among other players in the leaderboard, on the right of the screen.

3.3. Expert Evaluation Questionnaire

To find any usability issues and highlight the good practices of the user interface of the serious game, a heuristic evaluation was conducted [27]. In order to complete the evaluation, a questionnaire with ten questions was handed to the experts to fill out after examining the system based on the heuristic rules found in Table 2, as stated by Molich and Nielsen [28].

Table 2. Heuristic rules to examine the usability of the serious game.

Rule Number	Heuristic Rule
1	Simple and natural dialogue
2	Speak the user's language
3	Minimize user memory load
4	Be consistent
5	Provide feedback
6	Provide clearly marked exits
7	Provide shortcuts
8	Good error messages
9	Prevent errors
10	Provide help and documentation

3.4. Procedure

To evaluate the usability of the serious game, the experts (participants) started by playing the game. During all the evaluation activities, each expert (participant) was connected to one of the copies of the biofeedback device (Figure 7).

The game ran on a laptop and the users played the game with their hand of choice on a mouse, while their other hand was connected to the biofeedback device. The users were asked to play the game with the adaptivity component active and to play at least until they lose all their lives once. Then, in order to become familiarized with the game and find all the possible flaws, they played the game as much as they wanted to (Figure 8).

Once the play testing period was over, an online questionnaire with ten heuristic rules was handed to them and they were asked to rate the frequency that a rule was broken and the severity of the usability issue that affected the certain rule (Figure 8), as stated

by Nielsen [29]. The severity ratings of the usability problems, according to [29], can be interpreted as the following:

- 0 = not a usability problem;
- 1 = cosmetic problem, not necessary to be fixed if there is no extra time in the project;
- 2 = minor usability problem, fixing this should be given low priority;
- 3 = major usability problem, important to fix;
- 4 = usability catastrophe, must be fixed before product can be released.

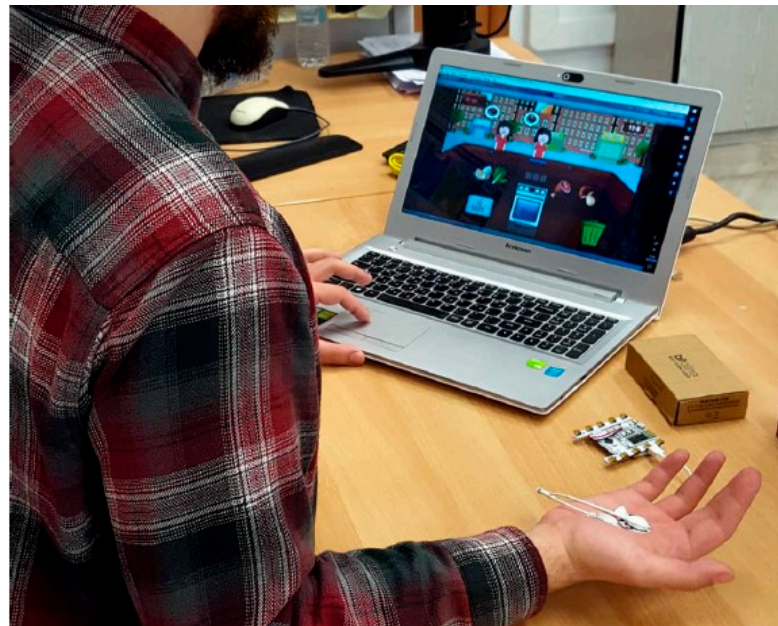


Figure 7. User playing the game while using the biofeedback device.

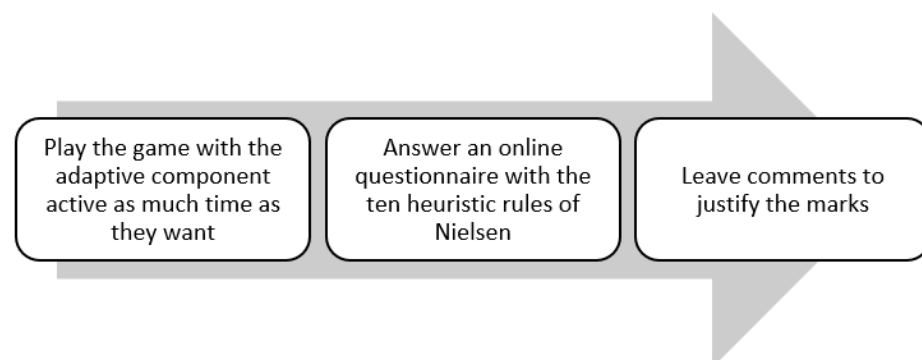


Figure 8. Description of the experiment procedure.

The scale of the frequency of occurrence of the specific usability issue was also 0 to 4.

Then, to help the developers create an updated version of the game and better understand the marks of every rule, a comment section was available, explaining the marks that were given.

3.5. Data Analysis

Six experts participated in the evaluation, as this is the optimal number of evaluators to spot as many flaws as possible, according to Nielsen [27]. All the experts were not involved in the design or the development of the software and had no prior exposure to it, so as to be objective. All of them were computer scientists, with a master's degree in Interactive

Technologies and have used the heuristic method numerous times. The following table (Table 3) contains some demographic information of the experts.

Table 3. Evaluators' demographic information.

Sex	Number	Percentage (%)
Female	0	0
Male	6	100
Age	Number	Percentage (%)
25–34	3	50.0
35–44	2	33.3
>45	1	16.6
Education level	Number	Percentage (%)
Master's degree	5	83.3
PhD	1	16.6

Each evaluator had to fill in the severity of the usability issue and the frequency that it occurred in the system, which are written as “S” and “F”, respectively, in the following table (Table 4).

Table 4. Heuristic evaluation results.

No.	Heuristic Rule	Ev.1		Ev.2		Ev.3		Ev.4		Ev.5		Ev.6	
		S	F	S	F	S	F	S	F	S	F	S	F
1	Simple and natural dialogue	1	0	3	0	3	0	3	0	2	0	1	0
2	Speak the user's language	1	0	3	0	2	0	3	0	1	0	1	0
3	Minimize user memory load	1	0	2	3	3	0	4	0	2	0	1	0
4	Be consistent	1	0	2	0	3	0	4	0	2	0	1	0
5	Provide feedback	1	0	3	1	3	1	3	2	2	0	2	0
6	Provide clearly marked exits	1	0	3	1	3	0	2	1	2	2	2	1
7	Provide shortcuts	1	0	2	2	3	0	2	0	2	0	1	0
8	Good error messages	1	0	3	0	3	0	2	0	1	0	1	0
9	Prevent errors	1	1	3	0	3	0	3	0	1	0	1	0
10	Provide help and documentation	1	1	3	2	3	0	4	2	3	2	3	1

These marks are accompanied by some comments to highlight the good and bad practices of the game. For the fifth heuristic rule regarding the feedback for every action, the third evaluator highlighted that he needed a symbol above the customer's head to indicate that he was served with the wrong dish instead of just reducing the coins. Moreover, the sixth evaluator commented on the fact that the bars representing the time left for every customer did not change their color when they were closer to zero, to grab the player's attention. Regarding the clearly marked exits rule, the fourth evaluator needed a reset button to restart the whole game without needing to exit and start over. The fifth evaluator, for the same rule, commented on the lack of the quit button in the final screen with the leaderboards. Moreover, for the ninth heuristic rule, the first evaluator said that there should exist an undo button on the oven to prevent the user from errors. Finally, almost all the experts stated that they needed a tutorial in the beginning.

On the other hand, some good practices were also highlighted. For the second rule, the third evaluator stated that the game is easy to understand, as there are only a few words and most of the game is based on images. The fifth evaluator said that the game is easy to handle, and the use of graphics and buttons is consistent on every screen for the third and fourth heuristic rules, respectively.

4. Discussion

To evaluate the usability of the serious game with the adaptive component, we conducted a heuristic evaluation of it with some experts. According to the participants' answers to the questionnaire that was handed to them after playing the game, most of them did not notice an error of great severity happening too frequently.

To further understand the marks given, some comments were also provided by the evaluators. As it is prominent in the marks table, the last heuristic rule regarding help and documentation in the game had the most significant issues, as most of the evaluators needed some help at the start of the game. Moreover, most of the negative comments were gathered in the fifth rule, emphasizing the lack of feedback on certain actions when serving the wrong dish.

All in all, it is safe to state that no major usability issues were spotted and certainly not ones regarding the integration of the biofeedback device while playing the game. The fact that the adaptivity was not easily observable in the game is important for our component, as the dynamic difficulty adjustment should go unnoticed in order to be well designed and implemented according to Andrade [9].

5. Limitations and Future Work

The present study has some limitations that will have an impact on the progression of our research. Even though a heuristic evaluation took place and only a several usability issues, not directly related to the adaptivity component, were, it surely is not enough to test the effectiveness of the serious game or even draw final conclusions for the usability of the adaptivity component of the serious game.

To do so, a larger experiment has been planned with students to gather data on the adaptivity algorithms' effectiveness and their influence on the gameplay. These data will help us show the frequency that the difficulty is adjusted because of the physiological data and a histogram of the different values of the parameter that affects the difficulty of the game. We intend to test if the engagement is higher with or without the adaptivity component and whether the serious game helps the students manage their time better and obtain higher scores when taking a multiple-choice questions test. Moreover, to test the game's usability, meaning the usefulness, ease of use, ease of learning and satisfaction of the user while playing the adaptive version of the application, an experiment with the final users that are part of the game's target group should take place. This experiment will test the hypothesis that was formed by the results of this paper that the adaptivity component causes no usability issues.

Finally, apart from further developing the game according to the experts' comments, we intend to incorporate an eye tracker or more physiological data, such as heart rate variability, and combine them with the biofeedback device data that we currently have so as to be more confident about the user's affective state.

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

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to restrictions enforced by the process of Aristotle University Ethics and legislation committee.

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

References

1. Zhonggen, Y. A Meta-Analysis of Use of Serious Games in Education over a Decade. *Int. J. Comput. Games Technol.* **2019**, 2019, 1–8. [CrossRef]
2. Kim, B.; Park, H.; Baek, Y. Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. *Comput. Educ.* **2009**, 52, 800–810. [CrossRef]
3. Calleja, G. Digital Games and Escapism. *Games Cult.* **2010**, 5, 335–353. [CrossRef]
4. Oksanen, K.; Lainema, T.; Hämäläinen, R. Learning from Social Collaboration. In *Gamification in Education*; IGI Global: Hershey, PA, USA, 2018; pp. 500–524. [CrossRef]
5. Vorderer, P.; Hartmann, T.; Klimmt, C. Explaining the enjoyment of playing video games: The role of competition. In Proceedings of the ICEC '03 Proceedings of the Second International Conference on Entertainment Computing, Pittsburgh, PA, USA, 8–10 May 2003; pp. 1–9. [CrossRef]
6. Juul, J. *Fear of Failing? The Many Meanings of Difficulty in Video Games: The Video Game Theory Reader*; Routledge: New York, NY, USA, 2009; Volume 2, pp. 237–252.
7. Su, Y.-S.; Chiang, W.-L.; Lee, C.-T.J.; Chang, H.-C. The effect of flow experience on player loyalty in mobile game application. *Comput. Hum. Behav.* **2016**, 63, 240–248. [CrossRef]
8. Csikszentmihalyi, M.; Nakamura, J. Flow Theory and Research. In *The Oxford Handbook of Positive Psychology*; Oxford University Press: Oxford, UK, 2009; pp. 195–206.
9. Andrade, G.; Ramalho, G.; Santana, H.; Corruble, V. Automatic computer game balancing. In Proceedings of the International Conference on Autonomous Agents, New York, NY, USA, 25–29 July 2005; pp. 1229–1230. [CrossRef]
10. Stein, A.; Yotam, Y.; Puzis, R.; Shani, G.; Taieb-Maimon, M. EEG-triggered dynamic difficulty adjustment for multiplayer games. *Entertain. Comput.* **2018**, 25, 14–25. [CrossRef]
11. Picard, R.W. Affective Computing. In *M.I.T. Media Laboratory Perceptual Computing*; M.I.T. Media Laboratory Perceptual Computing Section Technical Report No. 321; MIT Media Laboratory: Cambridge, MA, USA, 1995; pp. 1–11.
12. Daily, S.B.; James, M.T.; Cherry, D.; Porter, J.J.; Darnell, S.S.; Isaac, J.; Roy, T. Affective Computing: Historical Foundations, Current Applications, and Future Trends. In *Emotions and Affect in Human Factors and Human-Computer Interaction*; Academic Press: Cambridge, MA, USA, 2017; pp. 213–231. [CrossRef]
13. Liu, Y.; Sourina, O. Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm. In *Transactions on Computational Science XXIII*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 199–223.
14. Ninaus, M.; Tsarava, K.; Moeller, K. A Pilot Study on the Feasibility of Dynamic Difficulty Adjustment in Game-Based Learning Using Heart-Rate. In *International Conference on Games and Learning Alliance*; Springer: Cham, Switzerland, 2019; pp. 117–128. [CrossRef]
15. Monaco, A.; Sforza, G.; Amoroso, N.; Antonacci, M.; Bellotti, R.; de Tommaso, M.; Di Bitonto, P.; Di Sciascio, E.; Diacono, D.; Gentile, E.; et al. The PERSON project: A serious brain-computer interface game for treatment in cognitive impairment. *Health Technol.* **2019**, 9, 123–133. [CrossRef]
16. Bitalino. Electrodermal Activity (EDA) User's Manual. 2020. Available online: <https://bitalino.com/storage/uploads/media/electrodermal-activity-eda-user-manual.pdf> (accessed on 4 July 2022).
17. Biosignalnotebook. EDA Signal Analysis—A Complete Tour. 2018. Available online: http://notebooks.pluxbiosignals.com/notebooks/Categories/Other/eda_analysis_rev.html (accessed on 20 December 2021).
18. Chen, W.; Jaques, N.; Taylor, S.; Sano, A.; Fedor, S.; Picard, R.W. Wavelet-Based Motion Artifact Removal for Electrodermal Activity. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; Volume 2015, pp. 6223–6226. [CrossRef]
19. Nkurikiyeyezu, K.; Yokokubo, A.; Lopez, G. The influence of person-specific biometrics in improving generic stress predictive models. *Sens. Mater.* **2019**, 32, 703–722.
20. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing wesad, a multimodal dataset for wearable stress and affect detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, Boulder, CO, USA, 16–20 October 2018; pp. 400–408.
21. Koldijk, S.; Neerincx, M.A.; Kraaij, W. Detecting Work Stress in Offices by Combining Unobtrusive Sensors. *IEEE Trans. Affect. Comput.* **2016**, 9, 227–239. [CrossRef]
22. Koldijk, S.; Sappelli, M.; Verberne, S.; Neerincx, M.; Kraaij, W. The SWELL Knowledge Work Dataset for Stress and User Modeling Research. In Proceedings of the 16th ACM International Conference on Multimodal Interaction (ICMI 2014), Istanbul, Turkey, 12–16 November 2014.
23. Bajpai, D.; He, L. Evaluating KNN Performance on WESAD Dataset. In Proceedings of the 12th International Conference on Computational Intelligence and Communication Networks (CICN), Bhimtal, India, 25–26 September 2020; pp. 60–62. [CrossRef]
24. Aqajari, S.A.A.H.; Kasaeyan Naeini, E.; Mehrabadi, M.A.; Labbaf, S.; Rahmani, A.M.; Dutt, N. GSR Analysis for Stress: Development and Validation of an Open Source Tool for Noisy Naturalistic GSR Data. *arXiv* **2020**, arXiv:2005.01834, preprint.
25. Douglas, M.; Wilson, J.; Ennis, S. Multiple-choice question tests: A convenient, flexible and effective learning tool? A case study. *Innov. Educ. Teach. Int.* **2012**, 49, 111–121. [CrossRef]
26. Hlasny, V. Students' Time-Allocation, Attitudes and Performance on Multiple-Choice Tests (January 4, 2014). Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2379024 (accessed on 10 July 2022).

-
27. Nielsen, J. *How to Conduct a Heuristic Evaluation*; Nielsen Norman Group: Fremont, CA, USA, 1994; Volume 1, p. 8.
 28. Nielsen, J.; Molich, R. Heuristic evaluation of user interfaces. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, New York, NY, USA, 1–5 April 1990; pp. 249–256.
 29. Nielsen, J. *Usability Engineering*; Academic Press: Boston, MA, USA, 1993.