Effectiveness of a Serious Game Design and Game Mechanic Factors for Attention and Executive Function Improvement in the Elderly: A Pretest-Posttest Study

Ying-Kit Wong 1,* , Chih-Fu Wu 2 and Yung-Hsiang Tu 2

1 Graduate Institute of Design Science, Tatung University, Taipei 104, Taiwan
2 Department of Industrial Design, Tatung University, Taipei 104, Taiwan; wcf@gm.ttu.edu.tw (C.-F.W.); tys@gm.ttu.edu.tw (Y.-H.T.)

* Correspondence: d10717004@ms.ttu.edu.tw

Abstract: Attention allows us to focus and process information from our environment, and executive function enables us to plan, work, and manage our daily lives. As individuals become older, both of these cognitive abilities decline. It is essential for the elderly to perform more cognitive exercises. Previous studies have shown that arithmetic calculations require attention span and that playing video games requires executive function. Therefore, we developed a serious game involving mental arithmetic calculations specifically for improving attention span and executive functions. Our objective was to analyze the effectiveness of the game and the efficacy of the game’s mechanic factors affecting attention span and executive function in the elderly. Forty elderly volunteers who are over 60 years of age were invited to join an eight-week cognitive training program through an elderly social welfare center. Four assessment tests were used in pre-test and post-test before and after the training period. D-CAT and SAT are used for screening attention span; TMT-A and TMT-B are used for screening executive function. They were instructed to play the game for at least 15 min per day, 5 days per week, for a total of 8 weeks. There were three independent variables (difficulty, pressure, and competition) with two parameters that could be selected. A paired-sample t-test showed the effective results by comparing the pre-test scores and post-test scores of the cognitive training. There were significant improvements in attention span and executive functions. The mixed repeated-measure ANOVA and MANCOVA results showed that two game mechanic factors (difficulty and pressure) had a significant effect and an interaction effect, but the other factor (competition) had a non-significant effect. In conclusion, the game showed a significant enhancement in both attention span and executive functions after training, and the difficulty factor and the pressure factor were shown to have an effect, but the competition factor was shown to have no effect.

Keywords: attention; executive function; cognitive training; mental calculation; serious games; game mechanic; game design; elderly

1. Introduction

1.1. Elderly

The global population is rapidly aging, which has become an important topic in all countries. The United Nations (UN) and the World Health Organization (WHO) define older persons as those aged 60 years or over [1,2]. The United Nation report “World Population Prospects 2019” data show that, by 2050, 1 in 6 individuals worldwide will be over the age of 65 (16%), while 1 in 11 (9%) were at this age in 2019. By 2050, 1 in 4 individuals in Europe and North America will be 65 years of age or older. In 2018, the number of individuals aged 65 or over globally surpassed the number of individuals under the age of 5. In addition, the population aged 80 or older is projected to rise from 143 million in 2019 to 426 million in 2050 [3]. Additionally, 80% of the elderly individuals will be living...
in low- and middle-income countries in 2050. Therefore, all countries need to ensure that their countries’ health and social systems are ready to support this trend [4].

Therefore, we must pay attention to all aspects of the elderly. Gerontological research has focused not only on the care of the elderly who are sick or have functional disorders but also on how elderly individuals become healthier and gain more satisfaction with their lives. Successful aging has therefore become an emerging research topic. In the previous literature on successful aging, there are two popular explanations: (a) Rowe and Kahn [5–7] define successful aging as the ability to maintain the following three main characteristics: avoiding disease and disability, high cognitive and physical function, and engagement with life. The intersection of these three points is the most successful aging condition for the elderly. (b) Another widespread view of successful aging is the “selection, optimization, compensation” model by Baltes and Baltes [8], in which successful aging can be achieved through a process of self-psychological adjustment and personal behavior changes via educational, motivational, or health-related activities. According to these two references, we can understand that activities related to cognitive function and individual health are important for the elderly in successful aging. Therefore, these studies indicate that increases and improvements in high cognitive function through health-related activities for the elderly are part of the main features of successful aging.

In the cognitive neuroscience of aging, there are three patterns of age-related changes in cognitive behavior [9]: life-long declines, late-life declines, and life-long stability. Life-long decline includes the mechanisms of cognitive information processing; late-life decline includes tasks involving knowledge; and life-long stability includes autobiographical memory, emotional processing, and automatic memory processes. Functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) can show the differential effects of aging on specific brain regions.

In 2002, Ball [10] researched the effects of three cognitive training interventions with older adults. The purpose was to determine whether cognitive training can improve the mental abilities and daily life functioning of independently living elderly adults. Memory training, reasoning training, and speed training comprise cognitive training. The results showed the effectiveness of the improvement in target cognitive abilities in elderly adult participants after cognitive training.

The aging population is rapidly increasing throughout the world, so the demand for various elderly products is increasing. According to Mintel’s Traditional Toys and Games US November 2016 report [11], the general toy product design, toy manufacturers, and the market are mainly targeted at children. It is rare to have toys or games for the elderly. In fact, toys and games for the elderly are frequently categorized as skill training and leisure entertainment [12].

1.2. Arithmetic and Mental Calculation

What is arithmetic? It is the mathematics of numbers organized by the human brain and human mental capacities [13]. The four basic arithmetic operations in mathematics are addition, subtraction, multiplication, and division.

There is a strong connection between arithmetic and the human brain [14]. Dehaene [15] in 2004 writes the following: “Recent studies in human neuroimaging, primate neurophysiology, and developmental neuropsychology indicate that the human ability for arithmetic has a tangible cerebral substrate. The human intraparietal sulcus is systematically activated in all number tasks and could host a central a modal representation of quantity. Areas of the precentral and inferior prefrontal cortex also activate when subjects engage in mental calculation”.

Mental calculation refers to an arithmetical calculating process without any support from other devices or tools (e.g., a calculator, a pen, or paper) [16]. It is a basic mathematical skill and relates to procedural fluency, which is one of the five main components of mathematical proficiency, which refers to the flexible, precise, and efficient ability to perform procedures appropriately. In arithmetic, procedural fluency refers to the automation of com-
putation, mental computation, and written computation [17]. Star [18], in 2005, mentioned that an in-depth understanding of mathematics is not possible without the acquisition of basic facts and procedural computational skills.

1.3. Serial Sevens

Serial Sevens is a well-known attention assessment. The process of the test is a mental arithmetic calculation [19]. It is a screening test for checking the attention span of individuals of different ages. The test is applied in popular global cognitive assessments and cognitive screening instruments (MMSE and MoCA) [20,21]. MMSE and MoCA are widely used in checking a person’s cognitive level at different ages. They are used for screening healthy elderly persons and those with mild cognitive impairment (MCI) and dementia [22–25].

Serial subtraction/addition tasks are serial processing calculations. They comprise a processing stage that is organized serially, where one stage always precedes or follows another, from the beginning to the end of a task—in this case, subtracting form or adding to a serial action repeatedly. These are mental calculation processes. For example, for the Serial Sevens test, participants need to subtract the number 7 from the number 100 serially. Previous studies have shown that both serial subtraction tasks and serial addition tasks require an attention span [26–28].

Besides the Serial Sevens (7’s), previous studies also use Serial Threes (3’s) or Serial Thirteens (13’s) for attention screening. There are three types of serial calculation because their difficulty levels are different. For example, for the 3’s, participants need to subtract the number 3 from the number 100 serially. When subtracting or adding the whole number 3, the ones place digit changes every time, but the tens place digit changes every 3rd or 4th time (for example, $79 - 3 = 76; 76 - 3 = 73; 73 - 3 = 70$). This is easier than the 7’s and 13’s. For the 7’s, participants need to subtract the number 7 from the number 100 serially. When subtracting or adding the whole number 7, the ones place digit changes every time, but the tens place digit changes every 1st or 2nd time (for example, $86 - 7 = 79; 79 - 7 = 72; 72 - 7 = 65$), so it is more difficult than the 3’s but easier than the 13’s. For the 13’s, participants need to subtract the number 13 from the number 100 serially. When subtracting or adding the whole number 13, the ones place digit changes every time, and the tens place digit change every time as well (for example, $87 - 13 = 74; 74 - 13 = 61; 61 - 13 = 48$), which is more difficult than the 3’s and 7’s [29–31].

1.4. Attention

William James [32] in 1890 wrote the following: “Attention is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence”.

There are three components of human attention: alertness, selectivity, and processing capacity. These components are all strongly related to attention [33]. Previous studies have shown that, by doing simple arithmetic, the frontal, temporal, and parietal cortices in the brain are activated during each task [34]. Those areas of activity in the brain are related to attention control [35,36]. Some studies show that solving simple arithmetic calculations can improve attention span [37–39]. Serial subtraction tasks and serial addition tasks require arithmetic and mental calculations, so attention span can be improved by these tasks as well.

1.5. Executive Function

Executive function is a mental process that enables us to plan, work, and manage our daily lives. It is a combination of cognitive skills. The foundational components are inhibition, working memory, and shifting [40]. Such functions allow individuals to perform goal-directed problem solving, and this depends upon neural networks involving the prefrontal cortex in the brain [41]. As individuals become older, the capacity for executive function declines [42], but it can be enhanced by doing cognitive training exercises [43,44].
Previous studies have shown that smartphone games or video games can improve executive function, so they could be used as cognitive training tools to enhance executive function [45–49].

1.6. Game Design

In the book “Rules of Play, Game Design Fundamentals”, Tekinbas [50] mentions that the principle game design schemas are in three areas: rules, play, and culture. Rules are the indispensable logical and mathematical structures of a game; play refers to the player’s interaction with the game and other players; and culture refers to the larger cultural context within the game.

Another widely used theory is the MDA [51], i.e., mechanics, dynamics, and aesthetics. Mechanics refers to the components of the game, i.e., the rules of the game; dynamics refers to the run-time behavior of the game, which represents the system of the game; and aesthetics refers to the interaction with the game system, representing “fun” or “gameplay”.

One study describes the game design elements in five parts: the game interface design pattern, the game design pattern and mechanics, the game design principles and heuristics, the game models, and the game design methods [52].

These previous studies share similar conceptions and representations of game design. We summarize them into three main parts: game mechanics, which represents the rules, methods, and patterns of the game; gameplay, which represents the interaction with the game; and game culture, which represents the principle and system of the game.

Game mechanics is important in game design. It can be designed and controlled in the game. It includes the rules, methods, and patterns [53,54]. There are three important game mechanic factors in game design that can affect the performance of a player: difficulty, pressure, and competition.

The difficulty of a video game affects player performance, and this is based on the flow theory by the psychologist Mihály Csíkszentmihályi [55] from 1975. In a flow state, a person fully focuses and enjoys the process of the activity. This can be controlled by setting the difficulty level so that it is moderate, and that can lead to a high level of concentration during the task [56]. Therefore, difficulty is an essential game mechanic factor. Nakamura [57] mentioned, in 2014, the following: “Consciousness is the complex system that has evolved in humans for selecting information from this profusion, processing it, and storing it. Information appears in consciousness through the selective investment of attention”. Therefore, attention is related to the flow state. It is affected by setting the difficulty level. A standard straight line graph showed in Figure 1 can be used to describe the flow theory. The vertical axis is the challenges (from low to high), and the horizontal axis is the skills (from low to high). Lines with positive slope upwards. The middle part between two lines (Area 2) represents a moderate level of challenges and skills, which means Flow State. The left side (Area 1) represents high challenges and low skills, which means Anxiety. The right side (Area 3) represents low challenges and high skills, which means Boredom [58].

Pressure is a major mechanic factor that affects the performance of a player. The relationship between pressure and performance is shown in the Yerkes–Dodson law, which is a theory posited by Robert Yerkes and John Dillingham Dodson [59] in 1908. The Yerkes–Dodson bell curve is the most popular curve to describe pressure and performance. A standard bell curve graph showed in Figure 2 can be used to describe the Yerkes–Dodson law, the vertical axis is the performance (from weak to strong), and the horizontal axis is the pressure (from low to high). A normal distribution with a rounded peak tapering away at each end. The peak of the curve (Area 2) represents optimal pressure and optimal performance. The left side of curve (Area 1) represents low pressure and low performance, which means increasing attention and interest with low pressure. The right side of curve (Area 3) represents high pressure and low performance, which means impaired performance because of strong pressure [60]. Time pressure or a time limit is commonly used in game design. It puts pressure on the player to finish a certain task or
reach a goal in a limited amount of time, and it is related to performance [61–63]. Therefore, suitable time pressure in the game can enhance performance and thus increase attention span and executive function.

![Figure 1. A standard straight line graph, lines with positive slope upwards.](image1)

![Figure 2. A standard bell curve graph, a normal distribution with a rounded peak tapering away at each end.](image2)

Competition is another main mechanic game factor that affects the performance of a player. Competition indicates the presence of a competitor or competitors that compete with the player. Previous studies have shown that competition can increase physical effort and motivation [64], so there is a relationship between competition and performance. Performance can increase cognitive function and enhance attention [65].

1.7. Serious Game

A serious game is a game designed for a specific purpose, not only for pure entertainment. It was first introduced by Clark [66] in the 1970s. Nowadays, because of the advance of science and technology in video games and smartphones, serious games have become more popular, especially in training and educational areas. They can be used for cognitive training, and previous studies have shown that there are different kinds of cognitive function training that employ serious games [67–70].
Most individuals have a smartphone for their daily communication or entertainment [71], but, by installing serious games on a smartphone, it can also be used as a training tool. It is convenient for players to do cognitive training at a place of their choosing. Previous studies have shown that smartphone applications can serve as effective cognitive training tools [72,73].

Therefore, we especially developed a serious game involving mental arithmetic calculations for cognitive training purposes. Most of the previous research used a variety of tasks or exercises in cognitive training to enhance multiple cognitive domains [74–76]. This serious game design is unique, and there has not been any research done only focusing on computerized serial calculation tasks, in particular for attention span and executive function training purposes.

Finally, after gathering all those literature reviews and the important key points, it is necessary to design serious games for the elderly to maintain or improve their cognitive ability. Our aim was to develop an effective serious game. Additionally, we sought to discover the game mechanic factors that can affect and enhance the attention span and executive function in the elderly.

2. Materials and Methods

This study was approved by the Ethics Committee of National Taiwan University and conducted in accordance with the good practice procedures of the Declaration of Helsinki.

2.1. Cleverbrain Game Design

We developed a game called Cleverbrain, which is a mental arithmetic calculation game. The game design is based on the features of a cognitive assessment, i.e., the Serial Sevens. Serial Sevens is a widely used screening test for attention span; it has also been used as a cognitive screening tool. During the test, participants need to mentally calculate questions, i.e., the subtraction of 7 from 100. After providing the correct answer, participants need to continue to subtract seven from the previous answer in a serial fashion. Serial threes and serial thirteens are also screening tests for attention span with the same rule but with different numbers. We also applied them in this game.

Cleverbrain is a serious game designed for smartphones. Its appearance is similar to a card game. There is a set of 10 cards numbered 0 to 9 located at the bottom area of the screen. Players use their fingers to control the movement of the objects (cards) on the touch screen. They need to move the correct answers (cards) into the designated area (dotted-line area) and then release their finger to complete the answering process. After that, the internal system can detect that a card is being occupied and then automatically distribute another card to the bottom area.

This is similar to Serial Sevens. Players need to mentally calculate and answer questions serially. Each card contains a one-digit number only, and each answer is formed by a two-digit number from 0 to 99. Players can only move one card at a time, so players need to complete a pair of cards for each answer. At the beginning of each game set, the internal system will firstly deal a pair of random cards, which represent the starting numbers, and players can then start the game from this number.

The goal of the game is to complete six pairs of cards in each set, and there are two sets for each round. The first set is a serial subtraction task, and the second set is a serial addition task. Players need to provide the correct answer in the targeted area (dotted-line area) by moving the cards from the bottom area. For the serial subtraction process, the designated area (dotted-line area) is shown on the left side of the cards. For the serial addition process, the designated area (dotted-line area) is shown on the right side of the cards. For each play, if the answer is correct, the next target area (dotted-line area) will pop up for players to place the next answer; if the answer is wrong, the selected pair of cards will be destroyed, and the player needs to redo it. The internal system inside the smartphone continuously checks the answer during the game.
The game user interface and game process pictures are shown in Figure 3. The top left corner shows the player’s name and avatar, which can be set in the game menu, and a coin icon with numbers shows the instant scores of the game. The top center area shows the round number of the game. The top right corner shows the time limit (in seconds) for each set, and a setting button icon is used for other functional controls, such as turning music on/off and pausing the game. The middle area is the gaming area. After the game is started, a pair of random cards is dealt in the middle; the targeted area (dotted-line area) then shows cards on the left side for the serial subtraction task, and the targeted area shows cards on the right side for the serial addition task. The bottom area shows a set of set of 10 cards numbered 0 to 9. Players select the bottom card and move it to the target area (dotted-line area) to answer (Figure 3).

![Figure 3.](image_url)

**Figure 3.** *Cleverbrain* game design—the game user interface and game process. (A) For each round, there are 2 sets of mental arithmetic questions. The 1st set is the serial subtraction task, and the 2nd set is the serial addition task. Before the game starts, the screen shows the game rules and details in the middle (for example, this set is Minus 7, a serial subtraction task). (B) After the game is started, a pair of random cards is distributed in the middle area by the internal system. This is the starting number. Players follow this number to play. They select the card in the bottom area and move it to the (left side) targeted area (dotted-line area). (C) Players keep answering serially (left side) until six pairs of cards have been played. (D) After completing the 1st set, the 2nd set then starts afterward. The screen shows the game details in the middle (for example, this set is Add 7, the serial addition task). (E) After the game is started, a pair of random cards is distributed to the middle area by the internal system. This is the starting number. Players follow this number to play. They select the card in the bottom area and move it to the (right side) targeted area (dotted-line area). (F) Players keep answering serially (right side) until six pairs of cards have been played.
We designed a feedback and reward system in this game, and the outcome is based on the player’s performance. After each round is finished, the internal system calculates the scores, the accuracy percentage, and the total game time. This feedback allows players to monitor and track their own performance. In terms of reward, if players complete the game with high accuracy, then prizes (trophies and stars) will appear on the screen after the game is completed. Higher accuracy leads to more prizes, and lower accuracy leads to fewer prizes. Moreover, the history of all rewards (total trophies) and feedback (total scores) is recorded and shown in the top center area of the screen. The outlook and layout of the feedback are shown in Figure 4. The feedback and the reward system are designed to motivate and encourage players, thus creating a sense of joy, success, and satisfaction after a game.

![Figure 4. Cleverbrain game design—feedback and reward. After finishing the game, the feedback and rewards of all rounds are shown.](image)

2.2. Game Mechanic Factor Design

In Cleverbrain, there are three independent variables: difficulty, pressure, and competition, and these are the game mechanic factors. Players need to follow the instruction to select the settings of the game (based on the corresponding group number on Table 1).

<table>
<thead>
<tr>
<th>Game Mechanic Factors (Independent Variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
</tr>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>Group 2</td>
</tr>
<tr>
<td>Group 3</td>
</tr>
<tr>
<td>Group 4</td>
</tr>
<tr>
<td>Group 5</td>
</tr>
<tr>
<td>Group 6</td>
</tr>
<tr>
<td>Group 7</td>
</tr>
<tr>
<td>Group 8</td>
</tr>
</tbody>
</table>

1. Difficulty: Previous studies have shown that difficulty can improve a player’s performance, and there are three types of serial calculation assessment: Serial 3’s, Serial 7’s, and Serial 13’s. We modified and applied them to this game. Their difficulty level is based on the numbers and the calculations involved. Serial 3’s is the easiest, Serial 7’s is of medium difficulty, and Serial 13’s is the most difficult. In this game, players can choose a suitable difficulty level or choose the default setting (easy level).
2. Pressure Previous studies have shown that pressure can enhance a player’s performance, and time pressure (a time limit) is an important factor in game design. Thus, a suitable time pressure (time limit) is important. In this game, players can choose a suitable time limit (in seconds) or choose the default setting (unlimited time).

3. Competition Previous studies have shown that competition can help individuals to improve performance. Therefore, having competition (a multiplayer mode) and setting the number of competitors should enhance the player’s performance. In this game, players can choose multiplayer mode (competition) and the number of competitors (1–3), or choose the default setting (single-player mode), which means there are no competitors.

2.3. Experimental Design

Forty elderly volunteers were invited to join an eight-week cognitive training program through an elderly social welfare center, which is a local community center especially for old people in Hong Kong. All volunteers live in the same common location, they can gather, meet, and interact with one another in the center. Additionally, there are a lot of group and community work activities in the center. The staff in this center invited the elderly who lived in the local community to join this program. They were assigned to play the game for at least 15 min per day, 5 days per week, for a total of 8 weeks. Every participant needed an MMSE score of at least 25 and had to be above 60 years old for enrollment in this study.

Participants needed to do a pretest (before training) and a posttest (after 8 weeks of training). There were four assessments: two for testing the attention span and two for testing executive function.

Before the training started, there was a preparation process. We supported all participants in installing the software on their smartphone or opening the website version on their smartphone. We taught them how to play the game and showed them the rules, the methods, and how to control movement and operation. Additionally, we selected the game settings beforehand for each participant and subsequently allowed them to familiarize themselves with it. After the preparation process was finished, participants could begin training.

2.3.1. Game Mechanic Factors

The game mechanic factors (difficulty, pressure, and competition) had two possible settings that could be selected (with or without). Forty participants were randomly divided into eight groups. Each group had five participants, and the settings of each group are shown in Table 1.

1. Difficulty

There are two difficulty settings that can be selected: “with” or “without”. The setting is selected before training. If the game is “with” a difficulty level setting, a moderate difficulty level (not too easy or not too hard) is set for the player, which should enhance performance. Therefore, before training, the difficulty level was set based on each participant’s evaluation of their skill and their acceptability of the difficulty level. The difficulty level can be adjusted or gradually increase during the period of training because the participant’s skill might increase during training. Thus, what is considered a suitable difficulty level can be adjusted over time. The literature review shows that players can improve attention span and executive function by adjusting the difficulty level [77–79]. If the game is “without” a difficulty level setting, a default setting (the easy level) is selected and fixed during the training period.

2. Pressure

There are two pressure settings that can be selected: “with” and “without”. The setting is selected before training. If the game is “with” a pressure setting, there is a time limit in each game. A moderate time limit (not too little or not too much) was selected for each participant, to enhance performance. Therefore, before training, the time limit (in seconds) was selected based on each participant’s evaluation of their skill and their acceptability of time pressure. The time limit (in seconds) can be adjusted or gradually decrease during the training because the participant’s skill
might increase. Thus, what is considered a suitable time pressure can be adjusted over time. The literature review shows that players can improve attention span and executive function by setting the time limit [46,80,81]. If it is “without” pressure, a default setting (unlimited time) is selected. Participants can spend an unlimited amount of time to complete the questions in each game, and they are fixed during the training period.

3. Competition There are two competition settings that can be selected: “with” and “without”. The setting is selected before training. If the game is “with” competition, there is a competitor in each game, which is the multiplayer mode. Therefore, a multiplayer mode was set, and a suitable number of competitors, from 1 to 3, could be selected. Because each participant’s skill might increase during the training, this can be adjusted or gradually increased during the training to fit what is considered a suitable number of competitors. Thus, what is considered a suitable number of competitors can be adjusted over time. We used a computer player (NPC, non-player character) as the competitor in this game. This character is controlled by the computer and acts as another player in multiplayer mode. The player is created by computer programming and attempts to mimic a human player. The literature review shows that players can improve attention span and executive function by interacting in the game [82,83]. If the game is “without” competition, a default setting (no competitor) is selected. It is then in single-player mode, and it is fixed during the training period.

2.3.2. Software and Game Engine

We used JavaScript as the programming language and used the Construct 3 game engine for development. JavaScript is a high-level programming language and can run in Construct 3. Construct 3 is a user-friendly game engine that allows users to export the game to a range of platforms, such as the web (HTML5), Android, iOS (via Cordova), and desktop apps (via Windows/macOS wrappers or NW.js). Thus, installation and application are convenient for users. We also provided a website version for users who prefer to play the game through their smartphone web browser.

During the eight weeks of cognitive training, we contacted each participant to check their game status and see whether they needed support.

2.3.3. Cognitive Screening Assessment Test

A cognitive screening assessment test (MMSE) [20] was used to screen the participants before training. Only participants who scored 25 or higher could join the experiment. The MMSE, Mini–Mental State Examination, is a widely used assessment for total cognitive screening. It is a 30-point questionnaire that measures cognitive impairment. A score of 25 or higher is normal. A score of 24 or lower is abnormal, which indicates that the person might have cognitive impairment.

The testing category includes orientation to time (5 points), orientation to place (5 points), registration (3 points), attention and calculation (5 points), recall (3 points), language (2 points), repetition (1 point), and complex commands (6 points).

2.3.4. Attention Assessment Test

The two attention assessment test (D-CAT and SAT) was used in this experiment to screen attention span.

D-CAT [84], a Cancellation Test of Digits, is an attention assessment. There are three 10 × 13 lists of number, each with 130 numbers, for a total of 390 numbers. Numbers include 0 to 9. For the first number list, subjects need to find one preset number (5) in the list; for the second number list, subjects need to find two preset numbers (2 and 6) in the list; for the third number list, subjects need to find three preset numbers (1, 4, and 9) in the list. Subjects need to look at all numbers in the list and recognize whether they are correct or not. The primary measure of this test is the correct answers. Therefore, the maximum score is 390.
SAT [85,86], the sustained attention task, is another attention assessment. There are 29 numbers, including numbers from 1 to 9. There are 11 instances of the number 1. SAT target detection uses tapping. When the subject hears the number 1, they have to tap on the table to respond. If it is not number 1, they should not respond. The primary measure of this test is the correct tapping number. Therefore, the maximum score is 11.

2.3.5. Executive Function Assessment Test

Two executive function assessments (TMT-A and TMT-B) were used in this experiment. They are well-known tests for screening executive function.

TMT-A [87–89], Trail Making Test A, is an executive function test. There are 25 numbers on a paper on which the number positions are irregularly placed. The task requires subjects to connect numbers from 1 to 25 in a sequence, and examiners need to measure the time it takes to complete all 25 numbers starting from the beginning (the number 1). The primary measure of this test is the time of completion. There is no time limit for this test.

TMT-B [89,90], Trail Making Test B, is another executive function test. The original English version connects the number and alphabetic character alternately in sequential order, such as 1, A, 2, B, 3, C, 4, D . . . , for a total of 12 numbers and 12 English letters on a paper on which the positions are irregularly placed. However, all the participants in this study are native Chinese speakers. Previous studies have shown that the language abilities of a person can affect the result of this test. Using the Chinese version of Trail Making Test B for native Chinese speakers is more suitable and will produce more accurate results. Therefore, we used the Chinese version, where Chinese characters representing sequential numbers replaced the English alphabetic characters, connecting the numbers and Chinese characters alternately in sequential order, for a total of 12 numbers and 12 Chinese characters. The primary measure of this test is the time it takes to complete it. There is no time limit for this test.

2.4. Statistical Methods

Firstly, a paired sample \( t \)-test was used to analyze the comparison of the pre-test score and the post-test score. There are four assessments. By doing this analysis, we can understand the effectiveness of the game design in the 8-week training period for elderly participants. The result shows whether there are significant differences between pre-test scores and post-test scores.

Secondly, a mixed repeated-measure ANOVA was used to analyze the pre-test score and the post-test score with three independent variables (difficulty, pressure, and competition) with two possible settings (with and without). By doing this analysis, we can understand whether there is a significant effect or an interaction effect of these three independent variables.

Thirdly, graphs were plotted to show the relationship between variables and the interaction effect of variables, to help one visualize the variation of effect and show how the interaction affects the participants’ performance.

Additionally, we conducted multivariate analyses of covariance (MANCOVA) for the post-test scores in four assessment tests (attention span and executive function). The post-test scores in all measures were the dependent variable, and the game mechanic factors (difficulty, pressure, and competition) were the independent variables. The pre-test scores were the covariate, which can show the effect of the independent variables on the dependent variables without unwanted interference.

Finally, we used test–retest reliability to measure the consistency of four assessment tests (D-CAT, SAT, TMT-A, and TMT-B) of pre-test and post-test scores in different groups. There were eight groups of participants with different game settings in the serious game (Table 1).
3. Results

There were 40 participants enrolled in this study, including 24 females and 16 males. The mean age (year) was 65.18, and the standard deviation was 2.07; the mean of education completion (year) was 7.78, and the standard deviation was 1.85; and the mean of MMSE score was 26.83, and the standard deviation was 1.06 (Table 2).

Table 2. Characteristics of the participants in the training of Cleverbrain serious game.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (year)</td>
<td>65.18</td>
<td>2.07</td>
<td>0.33</td>
</tr>
<tr>
<td>Education (year)</td>
<td>7.78</td>
<td>1.85</td>
<td>0.29</td>
</tr>
<tr>
<td>MMSE (score)</td>
<td>26.83</td>
<td>1.06</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Note: SD, standard deviation; SE, standard error.

3.1. Attention Span

3.1.1. D-CAT Assessment Test Results

The pre-test scores and post-test scores showed that the mean of the pre-test was 373.63, the standard deviation was 2.35, and the standard error was 0.37; the mean of the post-test was 382.30, the standard deviation was 5.06, and the standard error was 0.80 (Table 3) (Figure 5).

Table 3. Paired samples statistics of D-CAT.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-CAT Pre-test</td>
<td>373.63</td>
<td>40</td>
<td>2.35</td>
<td>0.37</td>
</tr>
<tr>
<td>D-CAT Post-test</td>
<td>382.30</td>
<td>40</td>
<td>5.06</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Note: N, number of participants; SD, standard deviation; and SE, standard error.

Figure 5. Results of attention span assessment tests before and after training (with 95% CI for the mean). (A) D-CAT assessment test results. (B) SAT assessment test results.

The paired sample correlation showed that there was a significant difference ($p < 0.001$) (Table 4). The paired sample $t$-test result showed that there was a significant difference ($p < 0.001$) (Table 5). The mixed repeated-measure ANOVA results were significantly different in the pre/post-test, $F[1, 32] = 1649.438, p < 0.001$; difficulty, $F[1, 32] = 174.918, p < 0.001$; pressure, $F[1, 32] = 168.781, p < 0.001$; and the interaction effect of difficulty and pressure, $F[1, 32] = 25.329, p < 0.001$ (Table 6).

Table 4. Paired sample correlations of D-CAT.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Correlation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-CAT Pre-/Post-test</td>
<td>40</td>
<td>0.512</td>
<td>0.001 ***</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.001$. 
Table 5. Paired sample test of D-CAT.

<table>
<thead>
<tr>
<th>Source</th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-Tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/Post-test</td>
<td>-8.68</td>
<td>4.35</td>
<td>0.69</td>
<td>-10.07</td>
<td>-7.28</td>
<td>-12.606</td>
<td>39</td>
<td>0.000 ***</td>
</tr>
</tbody>
</table>

Note: *** p < 0.001; SD, standard deviation; and SE, standard error.

Table 6. Tests of within-subject effects of D-CAT.

<table>
<thead>
<tr>
<th>Source (D-CAT)</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/Post-Test</td>
<td>1505.113</td>
<td>1</td>
<td>1505.113</td>
<td>1649.438</td>
<td>0.000 ***</td>
<td>0.981</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty</td>
<td>159.613</td>
<td>1</td>
<td>159.613</td>
<td>174.918</td>
<td>0.000 ***</td>
<td>0.845</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure</td>
<td>154.013</td>
<td>1</td>
<td>154.013</td>
<td>168.781</td>
<td>0.000 ***</td>
<td>0.841</td>
</tr>
<tr>
<td>Pre/Post-Test * Competition</td>
<td>2.113</td>
<td>1</td>
<td>2.113</td>
<td>2.315</td>
<td>0.138</td>
<td>0.067</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure</td>
<td>23.113</td>
<td>1</td>
<td>23.113</td>
<td>25.329</td>
<td>0.000 ***</td>
<td>0.442</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Competition</td>
<td>0.013</td>
<td>1</td>
<td>0.013</td>
<td>0.014</td>
<td>0.908</td>
<td>0.000</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure * Competition</td>
<td>1.013</td>
<td>1</td>
<td>1.013</td>
<td>1.110</td>
<td>0.300</td>
<td>0.034</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure * Competition</td>
<td>0.313</td>
<td>1</td>
<td>0.313</td>
<td>0.342</td>
<td>0.563</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Note: *** p < 0.001.

Table 7. Estimated marginal means of D-CAT- Interaction of difficulty * pressure * pre/post-pest.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pressure</th>
<th>Pre/Post-Test</th>
<th>Mean</th>
<th>SE</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>with</td>
<td>with</td>
<td>pre-test</td>
<td>373.70</td>
<td>0.80</td>
<td>372.07</td>
<td>375.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>post-test</td>
<td>386.90</td>
<td>0.90</td>
<td>385.06</td>
<td>388.74</td>
</tr>
<tr>
<td>with</td>
<td>without</td>
<td>pre-test</td>
<td>373.70</td>
<td>0.80</td>
<td>372.07</td>
<td>375.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>post-test</td>
<td>383.50</td>
<td>0.90</td>
<td>381.66</td>
<td>385.34</td>
</tr>
<tr>
<td>without</td>
<td>with</td>
<td>pre-test</td>
<td>373.80</td>
<td>0.80</td>
<td>372.17</td>
<td>375.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>post-test</td>
<td>383.50</td>
<td>0.90</td>
<td>381.66</td>
<td>385.34</td>
</tr>
<tr>
<td>without</td>
<td>without</td>
<td>pre-test</td>
<td>373.30</td>
<td>0.80</td>
<td>371.67</td>
<td>374.93</td>
</tr>
<tr>
<td></td>
<td></td>
<td>post-test</td>
<td>375.30</td>
<td>0.90</td>
<td>373.46</td>
<td>377.14</td>
</tr>
</tbody>
</table>

Note: SE, standard error.

Figure 6A shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the D-CAT. Figure 6B shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * pressure in the D-CAT.
Figure 6. The graph of two-factor interaction in D-CAT. Figure (A) shows the interaction of pre/post-test * difficulty; Figure (B) shows the interaction of pre/post-test * pressure. Note: pre/post-test (the horizontal axis); pre-test = 1 and post-test = 2.

The interaction of pre/post-test * difficulty with pressure condition (Figure 7A) shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the D-CAT. The interaction of pre/post-test * difficulty without the pressure condition (Figure 7B) shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the D-CAT.

Figure 7. The graph of three-factor interaction in D-CAT. Figure (A) shows the interaction of pre/post-test * difficulty with pressure condition; Figure (B) shows the interaction of pre/post-test * difficulty without pressure condition. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.

However, the forms of these two interaction graphs were different from each other (with and without pressure), indicating that the interaction of pre/post-test * difficulty had different effects on the dependent variable at different levels of the pressure factor, which showed that there was a three-factor interaction effect.

In conclusion, D-CAT results showed the significant improvement of attention span. The difficulty factor and pressure factor had a significant effect and an interaction effect.
3.1.2. SAT Assessment Test Results

The pre-test scores and post-test scores showed that the mean of the pre-test is 8.53, the standard deviation is 0.78, and the standard error is 0.12; the mean of the post-test is 9.68, the standard deviation is 0.99, and the standard error is 0.16 (Table 8) (Figure 5).

Table 8. Paired samples statistics of SAT.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Pre-test</td>
<td>8.53</td>
<td>40</td>
<td>0.78</td>
<td>0.12</td>
</tr>
<tr>
<td>SAT Post-test</td>
<td>9.68</td>
<td>40</td>
<td>0.99</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Note: N, number of participants; SD, standard deviation; and SE, standard error.

The paired sample correlation showed that there was a significant difference ($p < 0.05$) (Table 9). The paired sample $t$-test result showed that there was a significant difference ($p < 0.001$) (Table 10). The mixed repeated-measure ANOVA results were significantly different in the pre-post-test, $F[1, 32] = 128.242, p < 0.001$; difficulty, $F[1, 32] = 15.515, p < 0.001$; pressure, $F[1, 32] = 40.97, p < 0.001$; and the interaction effect of difficulty and pressure, $F[1, 32] = 8.727, p < 0.01$ (Table 11).

Table 9. Paired sample correlations of SAT.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Correlation</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Pre/Post-test</td>
<td>40</td>
<td>0.322</td>
<td>0.043 *</td>
</tr>
</tbody>
</table>

Note: * $p < 0.05$.

Table 10. Paired sample test of SAT.

<table>
<thead>
<tr>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-Tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAT Pre/Post-test</td>
<td>−1.15</td>
<td>1.05</td>
<td>0.17</td>
<td>−1.49</td>
<td>−0.81</td>
<td>6.919</td>
<td>39</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.001$; SD, standard deviation; SE, standard error.

Table 11. Tests of within-subject effects of SAT.

<table>
<thead>
<tr>
<th>Source (SAT)</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig. (2-Tailed)</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/Post-Test</td>
<td>26.450</td>
<td>1</td>
<td>26.450</td>
<td>128.242</td>
<td>0.000 ***</td>
<td>0.800</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty</td>
<td>3.200</td>
<td>1</td>
<td>3.200</td>
<td>15.515</td>
<td>0.000 ***</td>
<td>0.327</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure</td>
<td>8.450</td>
<td>1</td>
<td>8.450</td>
<td>40.970</td>
<td>0.000 ***</td>
<td>0.561</td>
</tr>
<tr>
<td>Pre/Post-Test * Competition</td>
<td>0.800</td>
<td>1</td>
<td>0.800</td>
<td>3.879</td>
<td>0.058</td>
<td>0.108</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure</td>
<td>1.800</td>
<td>1</td>
<td>1.800</td>
<td>8.727</td>
<td>0.006 **</td>
<td>0.214</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Competition</td>
<td>0.050</td>
<td>1</td>
<td>0.050</td>
<td>0.242</td>
<td>0.626</td>
<td>0.008</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure * Competition</td>
<td>0.200</td>
<td>1</td>
<td>0.200</td>
<td>0.970</td>
<td>0.332</td>
<td>0.029</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure * Competition</td>
<td>0.450</td>
<td>1</td>
<td>0.450</td>
<td>2.182</td>
<td>0.149</td>
<td>0.064</td>
</tr>
</tbody>
</table>

Note: ** $p < 0.01$; *** $p < 0.001$.

Table 12 shows the mean of scores and other details of three independent variables (difficulty, pressure, and pre/post-test) with 2 levels in SAT.
Table 12. Estimated marginal means of SAT- Interaction of difficulty * pressure * pre/post-test.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pressure</th>
<th>Pre/Post-Test</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>pre-test</td>
<td>8.70</td>
<td>0.26</td>
<td>8.17</td>
<td>9.23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>10.60</td>
<td>0.19</td>
<td>10.21</td>
<td>10.99</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>pre-test</td>
<td>8.60</td>
<td>0.26</td>
<td>8.07</td>
<td>9.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>9.80</td>
<td>0.19</td>
<td>9.41</td>
<td>10.19</td>
<td></td>
</tr>
<tr>
<td>with</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>pre-test</td>
<td>8.20</td>
<td>0.26</td>
<td>7.67</td>
<td>8.73</td>
<td></td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>9.90</td>
<td>0.19</td>
<td>9.51</td>
<td>10.29</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>pre-test</td>
<td>8.60</td>
<td>0.26</td>
<td>8.07</td>
<td>9.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>post-test</td>
<td>8.40</td>
<td>0.19</td>
<td>8.01</td>
<td>8.79</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8A shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the SAT. In Figure 8B shows two lines that intersect each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * pressure in the SAT.

![Figure 8A](image1.png) ![Figure 8B](image2.png)

**Figure 8.** The graph of two-factors interaction in SAT. Figure (A) shows the interaction of pre/post-test * difficulty; Figure (B) shows the interaction of pre/post-test * pressure. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.

The interaction of pre/post-test * difficulty with pressure condition (Figure 9A) shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the SAT. The interaction of pre/post-test * difficulty without the pressure condition (Figure 9B) shows two lines intersecting each other, and they are not parallel due to the different direction. So, there was an antagonistic interaction of pre/post-test * difficulty in the SAT.

However, the forms of these two interaction graphs were different from each other (with and without pressure), indicating that the interaction of pre/post-test * difficulty had different effects on the dependent variable at different levels of the pressure factor, which showed that there was a three-factor interaction effect.

In conclusion, SAT results showed a significant improvement in attention span. The difficulty factor and pressure factor had a significant effect and an interaction effect.
3.2. Executive Function

3.2.1. TMT-A Assessment Test Results

The pre-test scores and post-test scores showed that the mean of the pre-test was 61.15, the standard deviation was 2.25, and the standard error was 0.36; the mean of the post-test was 56.53, the standard deviation was 3.77, and the standard error was 0.60 (Table 13) (Figure 10).

Table 13. Paired samples statistics of TMT-A.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>N</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-A Pre-test</td>
<td>61.15</td>
<td>40</td>
<td>2.25</td>
<td>0.36</td>
</tr>
<tr>
<td>TMT-A Post-test</td>
<td>56.53</td>
<td>40</td>
<td>3.77</td>
<td>0.60</td>
</tr>
</tbody>
</table>

Note: N, number of participants; SD, standard deviation; and SE, standard error.

The paired sample correlation showed that there was a significant difference \( (p < 0.05) \) (Table 14). The paired sample \( t \)-test result showed that there was a significant difference \( (p < 0.001) \) (Table 15). The mixed repeated-measure ANOVA results were significantly different in pre/post-test, \( F[1, 32] = 215.252, p < 0.001\); difficulty, \( F[1, 32] = 77.491, p < 0.001\); pressure, \( F[1, 32] = 4.585, p < 0.05\); the interaction effect of difficulty and pressure, \( F[1, 32] = 7.704, p < 0.01 \) (Table 16).
Table 14. Paired sample correlations of TMT-A.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Correlation</th>
<th>Sig.</th>
</tr>
</thead>
</table>
| TMT-A Pre/post-test      | 40 | 0.375       | 0.017 *

Note: * $p < 0.05$.

Table 15. Paired sample test of TMT-A.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>SE</th>
<th>Lower</th>
<th>Upper</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-Tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TMT-A Pre/post-test</td>
<td>4.63</td>
<td>3.59</td>
<td>0.57</td>
<td>3.48</td>
<td>5.77</td>
<td>8.142</td>
<td>39</td>
<td>0.000 ***</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.001$; SD, standard deviation; and SE, standard error.

Table 16. Tests of within-subject effects of TMT-A.

<table>
<thead>
<tr>
<th>Source (TMT-A)</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig. (2-Tailed)</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre/Post-Test</td>
<td>427.813</td>
<td>1</td>
<td>427.813</td>
<td>215.252</td>
<td>0.000 ***</td>
<td>0.871</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty</td>
<td>154.013</td>
<td>1</td>
<td>154.013</td>
<td>77.491</td>
<td>0.000 ***</td>
<td>0.708</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure</td>
<td>9.113</td>
<td>1</td>
<td>9.113</td>
<td>4.585</td>
<td>0.040 *</td>
<td>0.125</td>
</tr>
<tr>
<td>Pre/Post-Test * Competition</td>
<td>0.012</td>
<td>1</td>
<td>0.012</td>
<td>0.006</td>
<td>0.937</td>
<td>0.000</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure</td>
<td>15.313</td>
<td>1</td>
<td>15.313</td>
<td>7.704</td>
<td>0.009 **</td>
<td>0.194</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Competition</td>
<td>3.613</td>
<td>1</td>
<td>3.613</td>
<td>1.818</td>
<td>0.187</td>
<td>0.054</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure * Competition</td>
<td>1.513</td>
<td>1</td>
<td>1.513</td>
<td>0.761</td>
<td>0.390</td>
<td>0.023</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure * Competition</td>
<td>4.513</td>
<td>1</td>
<td>4.513</td>
<td>2.270</td>
<td>0.142</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Note: * $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$.

Table 17 showed the mean of scores and other details of three independent variables (difficulty, pressure, pre/post-test) with 2 levels in TMT-A.

Table 17. Estimated marginal means of TMT-A—Interaction of difficulty * pressure * pre/post-test.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pressure</th>
<th>Pre/Post-Test</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>with</td>
<td>with</td>
<td>pre-test</td>
<td>61.30</td>
<td>0.74</td>
<td>59.80</td>
<td>62.80</td>
</tr>
<tr>
<td>with</td>
<td>post-test</td>
<td>54.10</td>
<td>0.78</td>
<td></td>
<td>52.51</td>
<td>55.69</td>
</tr>
<tr>
<td>without</td>
<td>with</td>
<td>pre-test</td>
<td>61.80</td>
<td>0.74</td>
<td>60.30</td>
<td>63.30</td>
</tr>
<tr>
<td>without</td>
<td>post-test</td>
<td>54.20</td>
<td>0.78</td>
<td></td>
<td>52.61</td>
<td>55.79</td>
</tr>
<tr>
<td>without</td>
<td>with</td>
<td>pre-test</td>
<td>59.90</td>
<td>0.74</td>
<td>58.40</td>
<td>61.40</td>
</tr>
<tr>
<td>without</td>
<td>post-test</td>
<td>56.50</td>
<td>0.78</td>
<td></td>
<td>54.91</td>
<td>58.09</td>
</tr>
<tr>
<td>with</td>
<td>without</td>
<td>pre-test</td>
<td>61.60</td>
<td>0.74</td>
<td>60.10</td>
<td>63.10</td>
</tr>
<tr>
<td>with</td>
<td>post-test</td>
<td>61.30</td>
<td>0.78</td>
<td></td>
<td>59.71</td>
<td>62.89</td>
</tr>
</tbody>
</table>

Figure 11A shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was exponential an interaction of pre/post-test * difficulty in the TMT-A. Figure 11B shows two separate lines, and they are not parallel due to the different slopes. The blue line is steeper than...
the green line. So, there was an exponential interaction of pre/post-test * pressure in the TMT-A.

![Figure 11](image1.png)

**Figure 11.** The graph of two-factors interaction in TMT-A. Figure (A) shows the interaction of pre/post-test * difficulty; Figure (B) shows the interaction of pre/post-test * pressure. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.

The interaction of pre/post-test * difficulty with pressure condition (Figure 12A) shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the TMT-A. The interaction of pre/post-test * difficulty without pressure condition (Figure 12B) shows two lines intersecting each other and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the TMT-A.

![Figure 12](image2.png)

**Figure 12.** The graph of three-factor interaction in TMT-A. Figure (A) shows the interaction of pre/post-test * difficulty with pressure condition; Figure (B) shows the interaction of pre/post-test * difficulty without pressure condition. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.

However, the forms of these two interaction graphs were different from each other (with and without pressure), indicating that the interaction of pre/post-test * difficulty had different effects on the dependent variable at different levels of the pressure factor, which showed that there was a three-factor interaction effect.

In conclusion, the TMT-A results showed the significant enhancement of executive function. The difficulty factor and pressure factor had a significant effect and an interaction effect.
3.2.2. TMT-B Assessment Test Results

The pre-test scores and post-test scores showed that the mean of the pre-test is 69.43, the standard deviation is 1.65, and the standard error is 0.26; the mean of the post-test is 66.25, the standard deviation is 2.55, and the standard error is 0.40 (Table 18) (Figure 10).

<table>
<thead>
<tr>
<th>Table 18. Paired samples statistics of TMT-B.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>TMT-B Pre-test</td>
</tr>
<tr>
<td>TMT-B Post-test</td>
</tr>
</tbody>
</table>

Note: N, number of participants; SD, standard deviation; and SE, standard error.

The paired sample correlation showed that there was a significant difference ($p < 0.05$) (Table 19). The paired sample $t$-test result showed that there was a significant difference ($p < 0.001$) (Table 20). The mixed repeated-measure ANOVA results were significantly different in pre/post-test, $F[1, 32] = 192.012, p < 0.001$; difficulty, $F[1, 32] = 36.012, p < 0.001$; pressure, $F[1, 32] = 38.679, p < 0.001$; the interaction effect of difficulty and pressure, $F[1, 32] = 4.298, p < 0.05$ (Table 21).

<table>
<thead>
<tr>
<th>Table 19. Paired sample correlations of TMT-B.</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>TMT-B Pre/post-test</td>
</tr>
</tbody>
</table>

Note: * $p < 0.05$.

<table>
<thead>
<tr>
<th>Table 20. Paired sample test of TMT-B.</th>
</tr>
</thead>
<tbody>
<tr>
<td>95% Confidence Interval of the Difference</td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>TMT-B Pre/post-test</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.001$; SD, standard deviation; and SE, standard error.

<table>
<thead>
<tr>
<th>Table 21. Tests of within-subject effects of TMT-B.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source (TMT-B)</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Pre/Post-Test</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure</td>
</tr>
<tr>
<td>Pre/Post-Test * Competition</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Competition</td>
</tr>
<tr>
<td>Pre/Post-Test * Pressure * Competition</td>
</tr>
<tr>
<td>Pre/Post-Test * Difficulty * Pressure * Competition</td>
</tr>
</tbody>
</table>

Note: * $p < 0.05$; *** $p < 0.001$.

Table 22 shows the mean of scores and other details of three independent variables (difficulty, pressure, and pre/post-test) with two levels in TMT-B.
Table 22. Estimated marginal means of TMT-B—Interaction of difficulty * pressure * pre/post-test.

<table>
<thead>
<tr>
<th>Difficulty</th>
<th>Pressure</th>
<th>Pre/Post-Test</th>
<th>Mean</th>
<th>Std. Error</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>with</td>
<td>pre-test</td>
<td>70.00</td>
<td>0.54</td>
<td>68.89</td>
<td>71.11</td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>post-test</td>
<td>64.50</td>
<td>0.49</td>
<td>63.49</td>
<td>65.51</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>pre-test</td>
<td>68.90</td>
<td>0.54</td>
<td>67.79</td>
<td>70.01</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>post-test</td>
<td>65.30</td>
<td>0.49</td>
<td>64.29</td>
<td>66.31</td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>pre-test</td>
<td>69.20</td>
<td>0.54</td>
<td>68.09</td>
<td>70.31</td>
<td></td>
</tr>
<tr>
<td>with</td>
<td>post-test</td>
<td>65.50</td>
<td>0.49</td>
<td>64.49</td>
<td>66.51</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>pre-test</td>
<td>69.60</td>
<td>0.54</td>
<td>68.49</td>
<td>70.71</td>
<td></td>
</tr>
<tr>
<td>without</td>
<td>post-test</td>
<td>69.70</td>
<td>0.49</td>
<td>68.69</td>
<td>70.71</td>
<td></td>
</tr>
</tbody>
</table>

Figure 13A shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the TMT-B. Figure 13B shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * pressure in the TMT-B.

![Figure 13](image1.png)

Figure 13. The graph of two-factors interaction in TMT-B. Figure (A) shows the interaction of pre/post-test * difficulty; Figure (B) shows the interaction of pre/post-test * pressure. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.

The interaction of pre/post-test * difficulty with the pressure condition (Figure 14A) shows two lines intersecting each other, and they are not parallel due to the different slopes. The blue line is steeper than the green line. So, there was an exponential interaction of pre/post-test * difficulty in the TMT-B. The interaction of pre/post-test * difficulty without the pressure condition (Figure 14B) shows two separate lines, and they are not parallel due to the different direction. So, there was an antagonistic interaction of pre/post-test * difficulty in the TMT-B.

However, the forms of these two interaction graphs were different from each other (with and without pressure), indicating that the interaction of pre/post-test * difficulty had different effects on the dependent variable at different levels of the pressure factor, which showed that there was a three-factor interaction effect.

In conclusion, the TMT-B results showed the significant enhancement of executive function. The difficulty factor and pressure factor had a significant effect and an interaction effect.
3.3. MANCOVA Results

The results of the MANCOVA showed the significant difference of attention span and executive function.

A $2 \times 2 \times 2$ between-subjects multivariate analysis of covariance (MANCOVA) was performed on four dependent variables: the post-test scores of D-CAT and SAT (for attention span), and the post-test scores of TMT-A and TMT-B (for executive function), after controlling for pre-test scores. Independent variables are levels of difficulty (with and without), pressure (with and without), and competition (with and without).

Assumptions of normality, and the homogeneity of variance-covariance matrices, were evaluated. The Box’s M result showed 159.77, which indicates that the homogeneity of covariance matrices across groups is assumed ($F(70, 1401.94) = 1.25, p = 0.084$). The linearity and multicollinearity were satisfactory.

For multivariate tests, after controlling for pre-test scores, with the use of Wilks’ Lambda criterion, the combined dependent variables were significantly different in terms of levels of difficulty ($Wilk's \Lambda = 0.06, F(4, 25) = 106.95, p < 0.001$, partial $\eta^2 = 0.95$), and pressure ($Wilk's \Lambda = 0.07, F(4, 25) = 85.29, p < 0.001$, partial $\eta^2 = 0.93$), and not significant in terms of competition ($Wilk's \Lambda = 0.81, F(4, 25) = 1.44, p > 0.05$, partial $\eta^2 = 0.19$). Additionally, there was a significant interaction effect of difficulty and pressure ($Wilk's \Lambda = 0.22, F(4, 25) = 22.77, p < 0.001$, and partial $\eta^2 = 0.79$) (Table 23).

Table 23. Multivariate tests (Wilks’ lambda).

<table>
<thead>
<tr>
<th>Effect</th>
<th>Value</th>
<th>F</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>0.06</td>
<td>106.95</td>
<td>4</td>
<td>25</td>
<td>0.000 ***</td>
<td>0.95</td>
</tr>
<tr>
<td>Pressure</td>
<td>0.07</td>
<td>85.29</td>
<td>4</td>
<td>25</td>
<td>0.000 ***</td>
<td>0.93</td>
</tr>
<tr>
<td>Competition</td>
<td>0.81</td>
<td>1.44</td>
<td>4</td>
<td>25</td>
<td>0.251</td>
<td>0.19</td>
</tr>
<tr>
<td>Difficulty * Pressure</td>
<td>0.22</td>
<td>22.77</td>
<td>4</td>
<td>25</td>
<td>0.000 ***</td>
<td>0.79</td>
</tr>
<tr>
<td>Difficulty * Competition</td>
<td>0.90</td>
<td>0.73</td>
<td>4</td>
<td>25</td>
<td>0.58</td>
<td>0.11</td>
</tr>
<tr>
<td>Pressure * Competition</td>
<td>0.834</td>
<td>1.24</td>
<td>4</td>
<td>25</td>
<td>0.32</td>
<td>0.17</td>
</tr>
<tr>
<td>Difficulty * Pressure * Competition</td>
<td>0.82</td>
<td>1.34</td>
<td>4</td>
<td>25</td>
<td>0.28</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Note: *** $p < 0.001$. 

Figure 14. The graph of three-factor interaction in TMT-B. Figure (A) shows the interaction of pre/post-test * difficulty with pressure condition; Figure (B) shows the interaction of pre/post-test * difficulty without pressure condition. Note: pre/post-test (the horizontal axis), pre-test = 1, and post-test = 2.
Levene’s test of equality of error variances showed that all four assessment post-test scores were not significant \( p > 0.05 \). A non-significant result indicates it had met the assumption of homogeneity of variance (Table 24).

Table 24. Levene’s test of equality of error variances.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty</td>
<td>DCAT Posttest</td>
<td>293.68</td>
<td>1</td>
<td>293.68</td>
<td>151.27</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>7.56</td>
<td>1</td>
<td>7.56</td>
<td>34.48</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>251.27</td>
<td>1</td>
<td>251.27</td>
<td>64.53</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>56.95</td>
<td>1</td>
<td>56.95</td>
<td>46.96</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Pressure</td>
<td>DCAT Posttest</td>
<td>267.97</td>
<td>1</td>
<td>267.97</td>
<td>138.03</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>13.31</td>
<td>1</td>
<td>13.31</td>
<td>60.66</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>27.62</td>
<td>1</td>
<td>27.62</td>
<td>7.09</td>
<td>0.013 *</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>77.25</td>
<td>1</td>
<td>77.25</td>
<td>63.70</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Competition</td>
<td>DCAT Posttest</td>
<td>3.12</td>
<td>1</td>
<td>3.12</td>
<td>1.61</td>
<td>0.215</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>0.68</td>
<td>1</td>
<td>0.68</td>
<td>3.12</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>0.02</td>
<td>1</td>
<td>0.02</td>
<td>0.00</td>
<td>0.945</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>2.02</td>
<td>1</td>
<td>2.02</td>
<td>1.67</td>
<td>0.207</td>
</tr>
<tr>
<td>Difficulty * Pressure</td>
<td>DCAT Posttest</td>
<td>45.57</td>
<td>1</td>
<td>45.57</td>
<td>23.47</td>
<td>0.000 ***</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>2.66</td>
<td>1</td>
<td>2.66</td>
<td>12.14</td>
<td>0.002 **</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>35.07</td>
<td>1</td>
<td>35.07</td>
<td>9.01</td>
<td>0.006 **</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>22.81</td>
<td>1</td>
<td>22.81</td>
<td>18.81</td>
<td>0.000 ***</td>
</tr>
<tr>
<td>Difficulty * Competition</td>
<td>DCAT Posttest</td>
<td>0.02</td>
<td>1</td>
<td>0.02</td>
<td>0.01</td>
<td>0.915</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>0.36</td>
<td>1</td>
<td>0.36</td>
<td>1.62</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>5.53</td>
<td>1</td>
<td>5.53</td>
<td>1.42</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>1.95</td>
<td>1</td>
<td>1.95</td>
<td>1.61</td>
<td>0.215</td>
</tr>
<tr>
<td>Pressure * Competition</td>
<td>DCAT Posttest</td>
<td>2.30</td>
<td>1</td>
<td>2.30</td>
<td>1.19</td>
<td>0.285</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>0.14</td>
<td>1</td>
<td>0.14</td>
<td>0.65</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>3.59</td>
<td>1</td>
<td>3.59</td>
<td>0.92</td>
<td>0.345</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>2.31</td>
<td>1</td>
<td>2.31</td>
<td>1.91</td>
<td>0.178</td>
</tr>
<tr>
<td>Difficulty * Pressure * Competition</td>
<td>DCAT Posttest</td>
<td>0.47</td>
<td>1</td>
<td>0.47</td>
<td>0.24</td>
<td>0.627</td>
</tr>
<tr>
<td></td>
<td>SAT Posttest</td>
<td>1.14</td>
<td>1</td>
<td>1.14</td>
<td>5.19</td>
<td>0.030 *</td>
</tr>
<tr>
<td></td>
<td>TMTA Posttest</td>
<td>9.81</td>
<td>1</td>
<td>9.81</td>
<td>2.52</td>
<td>0.124</td>
</tr>
<tr>
<td></td>
<td>TMTB Posttest</td>
<td>0.60</td>
<td>1</td>
<td>0.60</td>
<td>0.50</td>
<td>0.487</td>
</tr>
</tbody>
</table>

Note: * \( p < 0.05 \); ** \( p < 0.01 \); and *** \( p < 0.001 \).
3.4. Test–Retest Reliability Results

Results of test–retest reliability (intraclass correlation coefficient) in different groups are shown in Table 26. We use test–retest reliability to measure four assessment tests (D-CAT, SAT, TMT-A, and TMT-B) consistency of pre-test scores and post-test scores in different groups. There were eight groups of participants with different game settings in the serious game (Table 26).

Table 26. Test–retest reliability (intraclass correlation coefficient).

<table>
<thead>
<tr>
<th>Game Mechanic Factors</th>
<th>Assessment Tests</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D-CAT</td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>With</td>
</tr>
<tr>
<td>Group 2</td>
<td>With</td>
</tr>
<tr>
<td>Group 3</td>
<td>With</td>
</tr>
<tr>
<td>Group 4</td>
<td>With</td>
</tr>
<tr>
<td>Group 5</td>
<td>Without</td>
</tr>
<tr>
<td>Group 6</td>
<td>Without</td>
</tr>
<tr>
<td>Group 7</td>
<td>Without</td>
</tr>
<tr>
<td>Group 8</td>
<td>Without</td>
</tr>
</tbody>
</table>

D-CAT results showed a high correlation coefficient in groups 1–6 and 9 (good reliability, ≥0.8 and <1.0); a medium correlation coefficient in group 7 (acceptable reliability, ≥0.6 and <0.8). The SAT results showed a high correlation in groups 1–3, 5, 7, 8 (good reliability, ≥0.8 and <1.0); a medium correlation coefficient in group 6 (acceptable reliability, ≥0.6 and <0.8); and a low correlation coefficient in group 4 (poor reliability, ≥0.4 and <0.6). The TMT-A results showed a high correlation in groups 2 and 5–7 (good reliability, ≥0.8 and <1.0); a medium correlation coefficient in groups 1, 3, and 4 (acceptable reliability, ≥0.6 and <0.8); and a low correlation coefficient in group 8 (poor reliability, ≥0.4 and <0.6). The TMT-B results showed a high correlation in groups 3, 5, 6, and 8 (good reliability, ≥0.8 and <1.0) and a medium correlation coefficient in groups 1, 2, 4, and 7 (acceptable reliability, ≥0.6 and <0.8).

In conclusion, the total number of results was 32. Twenty-one results displayed good reliability (≥0.8 and <1.0); nine results displayed acceptable reliability (≥0.6 and <0.8), and two results showed poor reliability (≥0.4 and <0.6).

4. Discussion

Our objective of the study was to analyze the effectiveness of the game and the efficacy of game mechanic factors affecting attention span and executive function in the elderly.

Regarding attention span, from the data of the two attention assessments (D-CAT and SAT), we used a paired sample t-test to compare the pre-test and post-test results of the game training. This showed that there is a significant difference between D-CAT (p < 0.001) and SAT (p < 0.001), which indicates that the game is effective. With the advancement of technology, individuals can improve their attention span by training with the serious mental calculation game in this study. This also proves that this game design can successfully turn traditional mental calculation into a new game on smartphones and with attention-span-improvement outcomes. Moreover, many individuals have smartphones and use them daily; it is more convenient for them to undergo cognitive training with mobile technology devices.

Regarding executive function, other studies have shown that traditional arithmetic calculation does not enhance executive function [91,92]. However, by using this serious smartphone game on mental calculation, participants can improve their attention span and executive function. According to the data of two executive function assessments (TMT-A and TMT-B), we used the paired sample t-test to compare the results of the pre-test and post-test scores of the game. This showed that there is a significant difference between TMT-A (p < 0.001) and TMT-B (p < 0.001). This might be because the process of the traditional
arithmetic calculation during training is straightforward. Subjects either verbally answer an examiner or write down the answer on paper. However, the game requires both the use of the brain and the movement of the hands and fingers. Players need to mentally calculate the question, choose cards, use their fingers to select them, and then execute commands by moving the correct answer into the target position. After that, players removed their finger from the screen for the final answering process. This combination of mental and physical operation activity on a smartphone is more complicated than a traditional arithmetic calculation. Therefore, this could be the reason that this game can improve both attention span and executive function.

Difficulty is an important game mechanic factor. The mixed repeated-measure ANOVA showed that there are significant differences between D-CAT ($p < 0.001$), SAT ($p < 0.001$), TMT-A ($p < 0.001$), and TMT-B ($p < 0.001$). In the theory of flow, a moderate difficulty level can increase performance, where a suitable level of difficulty means that a game is moderately difficult, which is an effective result in this experiment. The difficulty level might provide a challenge for individuals. They want to overcome a challenge and feel successful. This makes the game more attractive and makes players focus on it, thus improving attention span and executive function. On the other hand, if it is too easy, players might feel less challenged and become bored, which would cause players to focus less on the game and thus not improve their cognitive function at all. Additionally, if the game is too difficult, players might feel that the goal is unreachable, that there is a low chance of winning, and that there is a high chance of disappointment, which would lead to decreased concentration and thus no improvements in cognitive function [55,93,94]. Moreover, players can adjust or gradually increase the difficulty level during the game. Players might have increased their skill during the training period. They might feel that the original level is too easy, in which case players can gradually increase the difficulty level. Examples such as these might be what effectively leads to improvements in attention span and executive function.

Pressure is another important game mechanic factor. The mixed repeated-measure ANOVA showed that there are significant differences between D-CAT ($p < 0.001$), SAT ($p < 0.001$), TMT-A ($p < 0.005$), and TMT-B ($p < 0.001$). The Yerkes–Dodson law states that appropriate and suitable pressure can increase performance, thereby improving attention span and execution ability. When the time is limited in the game, players must complete the goal of each game within the time limit. A suitable amount of time pressure can make the player focus more on the game, which can improve their attention span and executive function. The Yerkes–Dodson bell curve describes the relationship between pressure and performance. We can see that the best performance is at the highest middle point of the bell curve, which is a moderate amount of pressure [60,95–97]. This might lead to suitable time pressure, which could enhance attention span and executive function. The time limit (in seconds) can be adjusted or gradually decreased during the game. The participant’s skill might have increased in the training period. They might feel that the original time limit is too long, in which case they can decrease the time limit. The time limit can be adjusted to best fit each participant during the training. These might be what effectively leads to improvements in attention span and executive function.

Competitive interaction is another important game mechanic factor. The mixed repeated-measure ANOVA showed that there are no significant differences between D-CAT ($p > 0.05$), SAT ($p > 0.05$), TMT-A ($p > 0.05$), and TMT-B ($p > 0.05$). Previous studies have shown that competition can increase performance for young people and adults, which includes the improvement of attention, physical effort, and motivation [65,98,99]. Not all digital games are suitable for the elderly. Therefore, a special design competitive game is necessary that can provide a positive experience and suitable for older players [100]. However, this experiment result shows that competition had a non-significant effect. There might be two reasons for this. Firstly, we used computer players (NPC, non-player character) as competitors. The competitor is controlled by a computer and is not human. Therefore, interactivity between humans might be different, and this might affect the effectiveness of.
the competition factor as human players might feel and act differently if they know that their competitors are computer players [101–103]. Another possible reason for this is the lack of communication with teammates or other competitors during the game [104,105]. In a competition, there is communication, for example, between teammates; teamwork and inspiration are necessary, and players can chat with other competitors during a game to exchange information. These two reasons might explain why the competition factor had a non-significant effect in game training.

There is an interaction effect between the difficulty factor and the pressure factor. The mixed repeated-measure ANOVA showed that there are significant differences between D-CAT \((p < 0.001)\), SAT \((p < 0.01)\), TMT-A \((p < 0.01)\), and TMT-B \((p < 0.05)\). This interaction effect indicates that the difficulty factor and the pressure factor are both essential and strongly related to the game. The reason for this might be that both factors can improve the attention span and executive function. Moreover, based on the theory of flow and the Yerkes–Dodson law, both factors have a similar feature, i.e., they both need to have a suitable amount of input to obtain the best performance. A moderate amount of input can lead to a more effective performance result. Therefore, the similarity of these two game mechanic factors might explain the interaction effect.

MANCOVA results were similar to the results of mixed repeated-measure ANOVA. For multivariate tests, after controlling for pre-test scores, with the use of Wilks’ Lambda criterion, the combined dependent variables were significantly different regarding the levels of difficulty and pressure, and the interaction effect of difficulty and pressure. Tests of between-subjects effects showed that the main effects of difficulty and pressure were significantly different, and the interaction effect of difficulty and pressure was significant different in all four assessment tests (D-CAT, SAT, TMT-A, and TMT-B). Additionally, there was one interaction effect of difficulty and pressure, and competition was significantly different in the SAT assessment test but no significant in any other assessment tests (D-CAT, TMT-A, or TMT-B).

Test–retest reliability measured the test consistency (the reliability) over time in this study. We measured the test results when repeating the same test on the same group of people in the pre-test and post-test. The results showed the consistency of pre-test scores and post-test scores of four assessment tests (D-CAT, SAT, TMT-A, and TMT-B) in eight different groups. Most of the results showed good reliability \((\geq 0.8 \text{ and } <1.0)\), and some results showed acceptable reliability \((\geq 0.6 \text{ and } <0.8)\).

5. Conclusions

Attention ability and executive function will gradually decline as individuals age. It is essential that the elderly do more cognitive exercise. This study concerns a serious game designed for attention span and executive function training in the elderly. We developed a mental arithmetic calculation game for use on a smartphone based on an attention assessment, Serial Sevens, and also by combining both serial subtraction tasks and serial addition tasks. The results show that participants improved attention span and executive function after training. The difficulty and pressure game mechanic factors showed a significant effect and an interaction effect.

This study makes some important contributions. First of all, the elderly are normally unfamiliar with ICT (information and communication technologies), and most do not play video games or smartphone games. This study shows that the elderly can be trained with a serious smartphone game and that it is suitable for them. Nowadays, with the advancement of technology, a smartphone is not only a communication tool but can also be a cognitive training tool for the elderly.

Furthermore, previous studies have shown that arithmetic calculation traditionally requires a long attention span and that attention can only be enhanced through training. We modified a new mental calculation game designed for smartphones to provide cognitive training. The results show that participants not only significantly improved attention span but also effectively improved executive function.
Finally, the difficulty and pressure game mechanic factors are the important elements in serious game design. The results show that a moderated level of both factors can obtain better performance. These are the valuable and crucial contributions of this study.

6. Future Recommendation

This research can act as a reference for game developers designing a game for cognitive training in the future. They can apply this design and the game mechanic factors to other games designed to increase cognitive function.

Analyzing other age groups is suitable for further research on the effectiveness of serious games involving mental arithmetic calculations, for children, teenagers, and adults. Because individuals of different ages have different levels of cognitive function abilities, by comparing the pre-test and post-test scores of training, we can determine whether there are significant effects on attention span and executive functions. Moreover, analyzing the effectiveness of game mechanic factors (difficulty, pressure, and competition) on groups of different ages is another suggestion for future research.

Author Contributions: Y.-K.W. provided conceptualization, designed the experiments, wrote the original draft and manuscript, and conducted the investigation and data analysis; C.-F.W. supervised the entire study; C.-F.W. and Y.-H.T. provided visualization and resources; and Y.-K.W., C.-F.W. and Y.-H.T. provided the methodology and project administration. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of National Taiwan University (protocol code: 202012ES066 and date of approval: 15 March 2021).

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

Acknowledgments: The authors would like to thank the research protocol of “A study on design specifications and human–machine interaction model of service robots for different tasks” from Tatung University, and to thank the editor and reviewers for their constructive suggestions, which improved this manuscript.

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

References


46. Gray, S.I.; Robertson, J.; Manches, A.; Rajendran, G. BrainQuest: The use of motivational design theories to create a cognitive training game supporting hot executive function. *Int. J. Hum.-Comput. Stud.* **2019**, *127*, 124–149. [CrossRef]


