

# Towards Guidelines for Mental State Induction in First-Person Shooters

Tomás Alves

INESC-ID & Instituto Superior Técnico, INESC-ID & Instituto Superior Técnico, INESC-ID & Instituto Superior Técnico,  
University of Lisbon  
Lisboa, Portugal  
tomas.alves@tecnico.ulisboa.pt

Sandra Gama

University of Lisbon  
Lisboa, Portugal

sandra.gama@tecnico.ulisboa.pt

Francisco S. Melo

University of Lisbon  
Lisboa, Portugal  
fmelo@inesc-id.pt

**Abstract**—In this paper we develop design guidelines for mental state induction in first-person shooters (FPS). To illustrate our design guidelines, we develop four versions of a first-person shooter video game each inducing a specific mental state on the player following our guidelines. We also conduct an evaluation using the self-perceived mental state of the participants as a metric to compare the mental state induced with each version. Results show that we were able to induce frustration, suggesting that further research is needed to fully understand the relation of gameplay and mental state inducement on FPS games.

**Keywords**—video game, psychology, mental state induction, game design, first-person shooter

## I. INTRODUCTION

First-Person Shooter (FPS) is one of the most popular contemporary genre of digital games, given that its action gameplay leads players to an arousal and engagement state. Additionally, the present increase of competitiveness in the industry leads game designers to find ways to create more engaging and attractive games. One way to do so may rely on affective gaming [1], which is a sub-field of affective computing [2].

Affective gaming comprises two different approaches. Game designers can either assess the emotions of players and adapt the gameplay based on them (*e.g.*, [3]–[5]) or provoke chosen emotions into the players (*e.g.*, [6]). Several studies [7]–[9] address FPS games and assess how players felt while interacting with them, yet we found a lack of studies on how to induce mental states on the players.

Our research focuses on this shortage and aims to provide useful inputs about if and how it is possible to induce mental states in gamers. Therefore, we introduce several **design guidelines on how to induce a specific mental state on a player**. More precisely, we developed four different versions of a FPS each of them focusing on a specific mental state between anxiety, boredom, engagement and frustration. In order to study our research question, we run a user testing phase where participants play the four versions of our game in random order. We then use a questionnaire to assess how they felt in the end playing each version.

## II. DEVELOPED WORK

This section covers the design process of the base game and the rationale behind the game parameter adjustments to induce specific mental states.

### II-A. Base Game Design

The FPS game depicts a dark ambient toy zombie world, based on the Unity3D "Survival Shooter tutorial"<sup>1</sup>. The zombie genre gained popularity in recent years due to the development of themes like *The Walking Dead*<sup>2</sup>, *Resident Evil*<sup>3</sup> and *Left 4 Dead*<sup>4</sup>, which gives the player a feeling of full control given its familiarity and genre conventions [10]. Figures 1, 2 and 3 depict the three different enemies used in the tutorial: Zombunnies (low health points and high speed), Zombears (medium health points and medium speed) and Hellephants (high health points and low speed).



Fig. 1: Zombunny. Fig. 2: Zombear. Fig. 3: Hellephant.

The game object assets are designed with detail. According to Nacke and Lindley [11], this leads to greater immersion. Based on the work of Cowley *et al.* [10], we introduced several new features including the first-person camera, a rocket launcher, stamina and sprinting system, and a pick-up system with rewards of health and ammo that the player can catch scattered across the game world.

The first-person camera brings about an embodiment of the in-game avatar, which leads the player to feel like they are acting directly upon the virtual game world [12]. Thus, the player can fully identify with the game character represented

<sup>1</sup>[www.unity3d.com/learn/tutorials/projects/survival-shooter-tutorial](http://www.unity3d.com/learn/tutorials/projects/survival-shooter-tutorial)

<sup>2</sup>[www.thewalkingdead.com](http://www.thewalkingdead.com)

<sup>3</sup>[www.en.wikipedia.org/wiki/Resident\\_Evil](http://www.en.wikipedia.org/wiki/Resident_Evil)

<sup>4</sup>[www.en.wikipedia.org/wiki/Left\\_4\\_Dead](http://www.en.wikipedia.org/wiki/Left_4_Dead)

only by the weapons that reach into the game environment. The rocket launcher was a weapon addition to give the player another way to interact with the enemies. While the machine gun can only damage one enemy at a time, the rocket launcher can damage all enemies that are inside an area of effect when the missile explodes. This different gameplay mechanics is expected to give the player a greater sense of control in the game by allowing them to adopt different strategies. The stamina and sprinting system is expected to create an empathetic feeling between the player and the in-game avatar which leads to a full immersion in the task at hand [13]. Finally, the pick-up system provides immediate feedback on the player's inventory, promoting their engagement [14].

Since the purpose of our work is to investigate the inducement of a mental state in players as they interact with the game, we created four different versions thereof by varying some of the game parameters, *e.g.*, the speed, health and spawn time of the enemies, as presented in the next section.

### II-B. Inducement Versions Design

The different versions were designed taking into account existing studies about flow [15] during gameplay [11], [16] as well as informal feedback from users. Csikszentmihalyi addresses the feeling of deep engagement as state of flow. The flow state (an equilibrium between skills and challenge) is compared against anxiety (challenge exceeds skills) and boredom (skills exceed challenge) conditions. Apathy was reported when challenges and skills were too low at the start or when a task had to be repeated frequently [11].

Figures 4a, 4c and 4b show sketches based on informal contact with users regarding how users felt while play testing differences in the game parameters values of enemies' speed, health and spawn time stats. Regarding the relationship between speed and health stats (Figure 4a), if all characters are too slow and easy to kill, the game has a low difficulty and the player becomes bored, because they can perform the task without a considerable challenge. As the value of the two parameters increases, the user becomes increasingly engaged and eventually anxious and frustrated. Finally, if the characters are too slow but have excessive health, they are easy to kill but doing so takes too much time, eventually inducing apathy in the player.

Figure 4b depicts the relationship between the spawn time of the enemies and their health. A balance between the values of the two parameters leads the user to engagement, since there is the challenge of fighting more enemies, but with an amount of health that allows the player to inflict critical damage before being overran. If enemies have a short spawn time, the player becomes anxious because they cannot kill the enemies as they spawn, so they become numerous. Coupled with an excess health, enemies surround the player in all fronts without the player being able to kill an enemy before another one spawns. Thus, the player becomes frustrated given the excessive difficulty. Finally, if the spawn time is high, players become bored due to the short number of enemies to fight against. Again, if they also have excess health, it leads players

to an apathetic state because they take too much time killing a low number of enemies that take too much time to kill without presenting a threat.

Finally, the relationship between the spawn time and the enemies speed is illustrated in Figure 4c. Testers said that, independently of the speed, for high values of spawn time, they would be bored and, in the limit, apathetic because there was a short number of enemies to fight against. As the value of the speed parameters increases and the spawn time decreases, the user becomes decreasingly engaged and eventually anxious and frustrated due to the increasing number of enemies that can overran the player in a shorter time. As mentioned, the plots are not validated, which means that there was no user testing phase to assess if the limits were well defined.

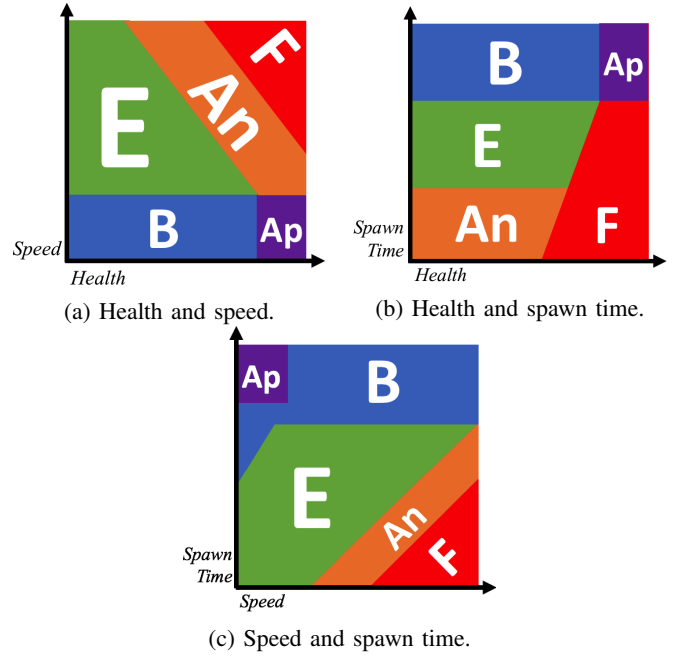


Fig. 4: Sketches of the mental states induced by the game as a function of the enemies' parameters. Labels: An - Anxiety, Ap - Apathy, B - Boredom, E - Engagement and F - Frustration.

Each version was fine-tuned to induce in the player a specific mental state, such as anxiety, boredom, engagement and frustration. The choice of which mental states to induce was based on the works of Csikszentmihalyi [17], Gilleade and Dix [18], and Poels *et al.* [19]. Players will also be able to play a sandbox version of the game and develop their initial skills. This takes away the need to create a game version focusing on apathy, since participants will not have skills too low for the task.

The versions and their parameters' variations are briefly presented in Table I. In addition to the game parameters values of enemies' speed, health and spawn time stats, we decided to alter other variables in versions B and D, since we need to distinguish boredom from apathy and increase impatience in gamers to induce frustration in them [20]. Changes in version B aim to reinforce the induction of boredom and,

at the same time, prevent apathy. In particular, the player's speed is reduced, the pick-ups' spawn time is enlarged, the ammunition spawn probability is increased and the health pick-up probability is reduced. Such settings make the game less engaging, since the player cannot run fast, which is expected to induce boredom in the player, and avoid apathy, since the player has more ammunition to fight back. In version D, the player's shooting probability is reduced. The variation makes the game too hard, which is expected to disrupt the player's gameplay capabilities, thus inducing frustration in the player.

TABLE I: Design guidelines for developing a game version to induce a specific mental state. The table features the identifier of the game version, the mental state it is intended to induce and the range of game parameters values per version.

| Version | Mental State | Parameters Settings   |
|---------|--------------|---|
| A       | Anxiety      | Low enemy health<br>High enemy speed<br>Low enemy spawn time  |
| B       | Boredom      | High enemy health<br>Low enemy speed<br>High enemy spawn time<br>Low player speed<br>High pick-up spawn time<br>High ammunition pick-up probability<br>Low health pick-up probability |
| C       | Engagement   | Medium enemy health<br>Medium enemy speed<br>Medium enemy spawn time  |
| D       | Frustration  | Low enemy health<br>High enemy speed<br>Low enemy spawn time<br>Low shooting probability  |

### III. MATERIAL AND METHODS

As mentioned in Section I, we need to verify if the inducement versions provoke the target mental states. In order to do so, we measure the self-perceived assessment of the participant's mental state with a questionnaire. In order to verify if a version induced a target mental state on the user, we compare his self-assessment of mental states intra and inter-versions, *i.e.* we check if anxiety is the most predominant mental state induced by version A and if version A led the player to a higher degree of an anxiety state based on their responses on the items of the questionnaires related to anxiety. Therefore, we can create eight hypotheses:

**H1:** *Anxiety is the most predominant mental state induced by version A.*

**H2:** *Boredom is the most predominant mental state induced by version B.*

**H3:** *Engagement is the most predominant mental state induced by version C.*

**H4:** *Frustration is the most predominant mental state induced by version D.*

**H5:** *Version A leads to a higher anxiety state compared to B, C and D versions.*

**H6:** *Version B leads to a higher boredom state compared to A, C and D versions.*

**H7:** *Version C leads to a higher engagement state compared to A, B and D versions.*

**H8:** *Version D leads to a higher frustrated state compared to A, B and C versions.*

In this section, we describe the population sample used in our study, material we used in the tests and the procedure of our testing phase.

#### III-A. Participants

Subjects were recruited through standard convenience sampling procedures including direct contact and through word of mouth. Subjects included anyone interested in participating if they were at least 18 years old. Each participant was asked to sign a consent form. There were no potential risks and no anticipated benefits to individual participants.

During the testing period, 31 tests were conducted. Due to technical problems, two tests were discarded. Finally, a total of 29 fully completed responses were used for the analysis. The participants (21 males, 8 females) were between 19 and 27 years old ( $M = 21.87$ ;  $SD = 1.60$ ). Only six participants reported no video game playing time. The other participants play at least once a day (10.34%), at least once a week (34.48%) or at least once a month (34.48%). Regarding FPS games, 11 (38%) participants reported that they do not frequently play that video game type.

#### III-B. Apparatus

Since the game was played in a laptop, players interacted with the game through mouse, keyboard and headphones. Regarding questionnaires, we provided a consent form, a demographic questionnaire and a questionnaire to assess the mental state of the player. The mental state assessment questionnaire contains ten questions on a 10-point Likert scale. It consists of the following items in English:

- I'm getting tired;
- I feel like I cannot stop playing;
- I play without thinking how to play;
- I lose track of time;
- I get wound up;
- I felt anxious;
- I felt bored;
- I felt frustrated;
- I liked to play the game;
- I would like to play the game again.

The first five items are taken from the Game Engagement Questionnaire (GEQ) [21], with four of them regarding elements of flow and one concerning presence, since all this items are related to engagement. The first element in the original GEQ version was "I cannot tell I am getting tired", but, given a mistake in the translation from English to Portuguese, it was used as "I realize I am getting tired". There is no difference, except that one is the negation of the other, therefore the response in the item must be processed accordingly. It relates to the flow dimension concerning the loss of self-consciousness, the second question regards the autotelic dimension, the third one has to do with the merging of action and awareness and

the fourth point addresses the balance between the challenge of the task and the individual's skills. The element concerning presence can be representative of the flow dimension regarding the time transformation in which the sense of time is crooked. The three next questions consist of the self-assessment of anxiety, boredom and frustration. This helps to prevent the user from trying to guess what was special about the version they played and not focus on the task at hand. Finally, the order of the adapted GEQ items was randomized between the four versions.

### III-C. Procedure

Upon arrival, the assistant explained the participants the purpose of the study and what they would be doing. The assistant asked the participant to fill a consent form and a questionnaire regarding their demographic information and gaming experience.

Before playing any version of the game, users were asked to rest for three minutes to reset their mental state. While the players were resting, they were looking at a set of images. We chose these images from the International Affective Picture System (IAPS) [22], as done in [23]–[25].

After that, users were given a sandbox version of the game to try the sensitivity and in-game interactions. The tester could play as much time as they wanted in order to develop the minimum skills to play the game in the different versions.

The participant then played each of the four versions for ten minutes where the order by which each player interacted with the four versions of the game was randomized.

The following procedure was used for each version the participant played:

- 1) The assistant asked the player if they was comfortable and ready to play;
- 2) The assistant started the game for the player;
- 3) If the player died, the game automatically restarted;
- 4) After ten minutes, the game closed;
- 5) The assistant asked the tester to fill the form addressing the version the latter played;
- 6) The player rested for three minutes looking at a picture in order to return to a neutral mental state.
- 7) After three minutes, the assistant repeated this procedure to the next version, if there was other version to play.

The decision regarding the game closing automatically after ten minutes seeks to avoid breaking immersion (by talking to the subject) before ten minutes of game have passed.

When they were done playing the four versions, they were invited to write free comments. Testers entered in a contest to determine which player achieved the highest score across all gaming sessions. The winner received a gift card worth 20EUR.

## IV. RESULTS

We can assess if the participant felt anxious, bored or frustrated because there is a explicit item that addresses those states. Yet, we cannot do the same for engagement. Since we can not say that a user was engaged just by addressing

individual GEQ elements, we decided to create the Flow Degree Scale (FDS).

This is a scale that uses the items from the Adapted GEQ questionnaire and helps to distinguish when the user felt engaged from anxious, bored and frustrated. The FDS gives a discrete Flow Degree number and it uses the first five items. This Flow Degree number is calculated by the sum of each points from the items "I felt like I could not stop playing", "I played without thinking how to play" and "I lost track of time" minus the points from the items "I was getting tired" and "I got wound up". The values vary between  $-17$  and  $28$ . We cannot compare the values from 1 to 10 from the anxiety, boredom and frustration items to the Flow Degree because of the different value intervals. With this in mind, when compared to the other mental states, Flow Degree was normalized using the min-max normalization.

We started by performing Shapiro-Wilk tests to our hypotheses. Table II shows the results. Shapiro-Wilk tests showed that results were significant for at least one data set of each hypotheses, except for **H7**. Afterwards, we verified if data for hypotheses with at least a significant result in Table II had a chi-square distribution (see Table III). All hypotheses had data with a chi-squared distribution, except for **H6**. With this in mind, we performed Wilcoxon tests to follow-up these findings of hypotheses **H1**, **H2**, **H3**, **H4**, **H5** and **H8**. For the first four hypotheses, the control data set was the mental state that we aimed to induce for that specific version. For the last three hypotheses, the control data set was the version that was designed to induce the specific mental state that we are addressing. We also applied a Bonferroni correction and so all effects are reported at a 0.0167 level of significance in Table IV.

Regarding **H1**, we can conclude that playing version A led to higher anxiety compared to boredom, and this effect was medium in size. Nevertheless, version A did not produce any substantial difference of anxiety relative to engagement and frustration. Therefore, this test is inconclusive for this hypothesis. In Figure 5, engagement was the predominant mental state with the highest median ( $Mdn = 6$ ), compared to the ones of anxiety ( $Mdn = 4$ ), boredom ( $Mdn = 3$ ) and engagement ( $Mdn = 5$ ). This suggests that users were more engaged playing version A compared to any other mental state. Regarding Interquartile Range (IQR), anxiety, boredom and frustration have large IQRs, which suggests that they have a similar variation. In contrast, engagement has a small IQR, which means that the values were close to the median.

Results for **H2** show that playing version B led to higher engagement compared to boredom, and this effect was large in size. Moreover, version B did not produce any substantial difference of boredom relative to anxiety and frustration. Thus, this test is inconclusive for this hypothesis. Figure 6 pictures the boxplots of the medians of the answers for the mental states. Engagement has the highest median ( $Mdn = 6.2$ ), compared to anxiety ( $Mdn = 4$ ), boredom ( $Mdn = 4$ ) and frustration ( $Mdn = 2$ ). This suggests that engagement was the predominant mental state while testers played version B.

TABLE II: Normality test results for all hypotheses. Significant results are in bold.

|           | Data set    | D(29) | p                 |
|-----------|-------------|-------|-------------------|
| <b>H1</b> | Anxiety     | 0.920 | <b>0.031</b>      |
|           | Boredom     | 0.888 | <b>0.005</b>      |
|           | Engagement  | 0.963 | 0.382             |
|           | Frustration | 0.930 | 0.055             |
| <b>H2</b> | Anxiety     | 0.910 | <b>0.018</b>      |
|           | Boredom     | 0.913 | <b>0.021</b>      |
|           | Engagement  | 0.952 | 0.201             |
|           | Frustration | 0.829 | <b>&lt; 0.001</b> |
| <b>H3</b> | Anxiety     | 0.914 | <b>0.022</b>      |
|           | Boredom     | 0.835 | <b>&lt; 0.001</b> |
|           | Engagement  | 0.975 | 0.712             |
|           | Frustration | 0.937 | 0.084             |
| <b>H4</b> | Anxiety     | 0.923 | <b>0.036</b>      |
|           | Boredom     | 0.905 | <b>0.013</b>      |
|           | Engagement  | 0.875 | 0.875             |
|           | Frustration | 0.932 | 0.061             |
| <b>H5</b> | Version A   | 0.920 | <b>0.031</b>      |
|           | Version B   | 0.910 | <b>0.018</b>      |
|           | Version C   | 0.914 | <b>0.022</b>      |
|           | Version D   | 0.923 | <b>0.036</b>      |
| <b>H6</b> | Version A   | 0.888 | <b>0.005</b>      |
|           | Version B   | 0.913 | <b>0.021</b>      |
|           | Version C   | 0.835 | <b>&lt; 0.001</b> |
|           | Version D   | 0.905 | <b>0.013</b>      |
| <b>H7</b> | Version A   | 0.963 | 0.382             |
|           | Version B   | 0.952 | 0.201             |
|           | Version C   | 0.975 | 0.702             |
|           | Version D   | 0.982 | 0.875             |
| <b>H8</b> | Version A   | 0.930 | 0.055             |
|           | Version B   | 0.829 | <b>&lt; 0.001</b> |
|           | Version C   | 0.937 | 0.084             |
|           | Version D   | 0.932 | 0.061             |

TABLE III: Chi-square test results for all hypotheses. Significant results are in bold.

| Hypothesis | $\chi^2(3)$ | p                 |
|------------|-------------|-------------------|
| <b>H1</b>  | 15.076      | <b>0.002</b>      |
| <b>H2</b>  | 20.341      | <b>&lt; 0.001</b> |
| <b>H3</b>  | 20.893      | <b>&lt; 0.001</b> |
| <b>H4</b>  | 39.418      | <b>&lt; 0.001</b> |
| <b>H5</b>  | 18.095      | <b>&lt; 0.001</b> |
| <b>H6</b>  | 3.474       | 0.324             |
| <b>H8</b>  | 43.798      | <b>&lt; 0.001</b> |

TABLE IV: Wilcoxon signed ranks test results for hypotheses H1, H2, H3, H4, H5 and H8.

|           | Control     | Other       | Z      | p                 | r      |
|-----------|-------------|-------------|--------|-------------------|--------|
| <b>H1</b> | Anxiety     | Boredom     | -2.476 | <b>0.013</b>      | -0.460 |
|           | Anxiety     | Engagement  | -1.494 | 0.135             | -0.277 |
|           | Anxiety     | Frustration | -0.807 | 0.420             | -0.150 |
| <b>H2</b> | Boredom     | Anxiety     | -0.579 | 0.562             | -0.108 |
|           | Boredom     | Engagement  | -3.006 | <b>0.003</b>      | -0.558 |
|           | Boredom     | Frustration | -1.361 | 0.173             | -0.253 |
| <b>H3</b> | Engagement  | Anxiety     | -2.196 | 0.028             | -0.408 |
|           | Engagement  | Boredom     | -3.882 | <b>&lt; 0.001</b> | -0.721 |
|           | Engagement  | Frustration | -1.125 | 0.261             | -0.209 |
| <b>H4</b> | Frustration | Anxiety     | -2.246 | 0.025             | -0.417 |
|           | Frustration | Boredom     | -4.468 | <b>&lt; 0.001</b> | -0.829 |
|           | Frustration | Engagement  | -2.449 | <b>0.014</b>      | -0.455 |
| <b>H5</b> | Version A   | Version B   | -2.641 | <b>0.008</b>      | -0.490 |
|           | Version A   | Version C   | -1.146 | 0.252             | -0.213 |
|           | Version A   | Version D   | -2.227 | 0.026             | -0.414 |
| <b>H8</b> | Version D   | Version A   | -2.933 | <b>0.003</b>      | -0.545 |
|           | Version D   | Version B   | -4.466 | <b>&lt; 0.001</b> | -0.829 |
|           | Version D   | Version C   | -3.174 | <b>0.002</b>      | -0.589 |

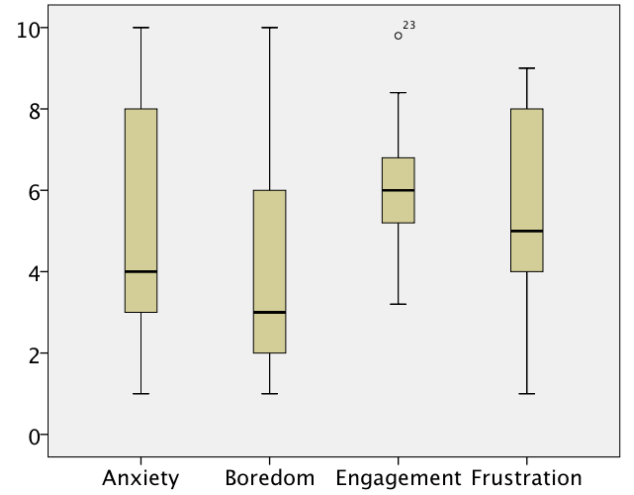


Fig. 5: Boxplots of the medians of the mental states for version A.

It is also noteworthy that boredom and frustration have a similar and larger IQRs compared to anxiety and engagement. Moreover, engagement is again the mental state with smaller IQR.

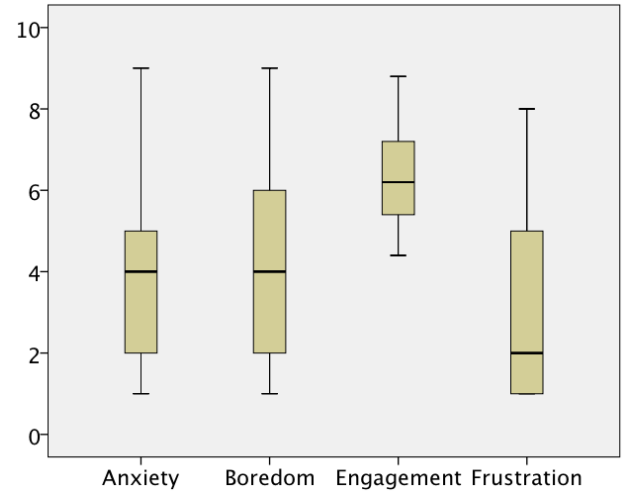


Fig. 6: Boxplots of the medians of the mental states for version B.

We can conclude that for **H3** playing version C led to higher engagement compared to boredom, and this effect was large in size. In addition, version C did not produce any substantial differences regarding engagement and anxiety or frustration. Therefore, we find this test inconclusive for this hypothesis. Although the test for this hypothesis is inconclusive, we can verify in Figure 7 that engagement had the highest median ( $Mdn = 6.4$ ) and the smaller IQR, compared to anxiety ( $Mdn = 4$ ), boredom ( $Mdn = 2$ ) and frustration ( $Mdn = 6$ ).

Results of **H4** show that playing version D led to a higher frustrated feeling compared to boredom and engagement and this effect was large and medium in size, respectively. There-

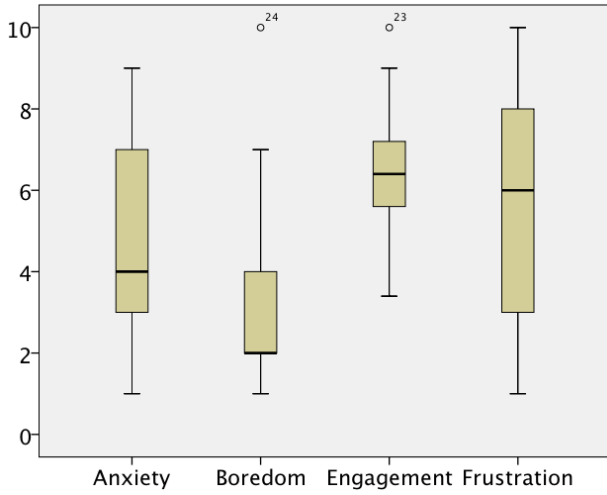


Fig. 7: Boxplots of the medians of the mental states for version C.

fore, this test inconclusive for this hypothesis. Figure 8 pictures the boxplots of the medians of the self-assessed mental states. Both anxiety and frustration have the highest median ( $Mdn = 7$ ), compared to the ones of boredom ( $Mdn = 3$ ) and engagement ( $Mdn = 5.8$ ). In particular, frustration's IQR has higher values than the anxiety's, therefore we can conclude that, although not at a significant level, version D lead users to a frustrated mental state.

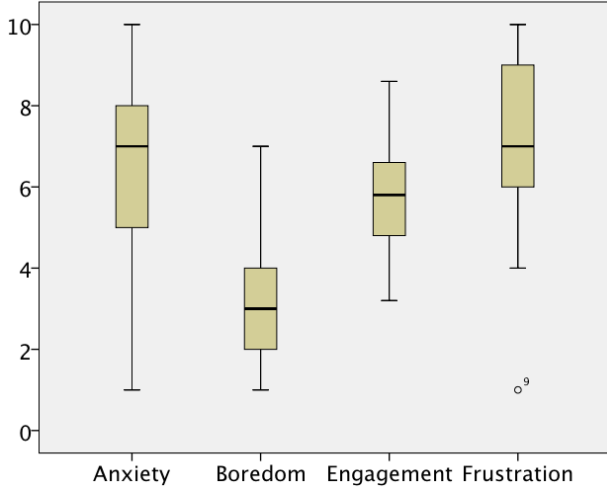


Fig. 8: Boxplots of the medians of the mental states for version D.

Concerning **H5**, we can conclude that playing version B led to a less anxious feeling and this effect was medium in size. Nevertheless, versions C and D didn't produce any substantial difference of anxiety relative to the control programme. In Figure 9, version D has the highest median ( $Mdn = 7$ ), compared to the ones of versions A ( $Mdn = 4$ ), B ( $Mdn = 4$ ) and C ( $Mdn = 4$ ). This suggests that, while not significantly, users were more anxious playing version D compared to any

other. Moreover, versions B and D have similar IQR, which suggests that they have a similar variation.

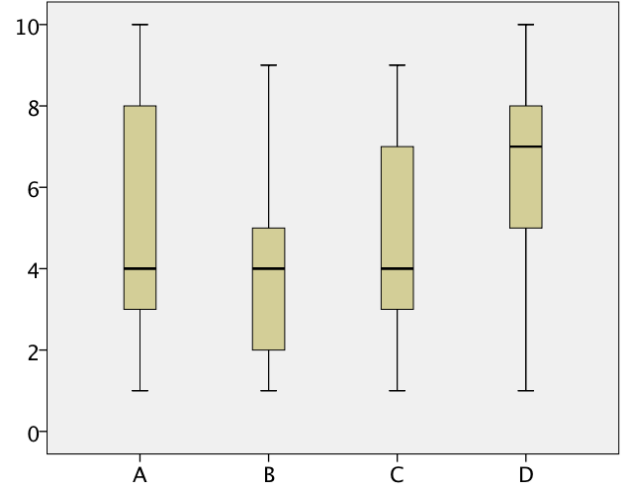


Fig. 9: Boxplots of the medians for the "I felt anxious" item.

Since the version the tester played didn't significantly affected how they answered the item "I felt bored", we conclude that this test is inconclusive for **H6**. Figure 10 pictures the boxplots of the medians of the answers for this item across different versions. Version B has the highest median ( $Mdn = 4$ ), compared to versions A ( $Mdn = 3$ ), C ( $Mdn = 2$ ) and D ( $Mdn = 3$ ). This suggests that, while it is not significant, users were more bored playing version B compared to any other. It is also noteworthy that versions A and B have a similar and larger IQR compared to versions C and D, which also have a similar IQR.

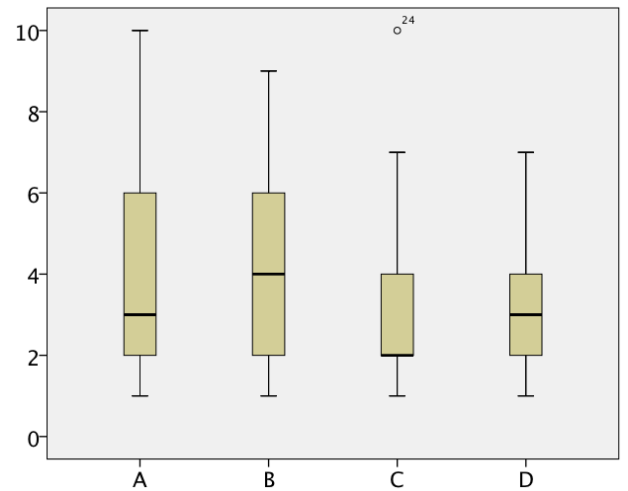


Fig. 10: Boxplots of the medians for the "I felt bored" item.

For **H7**, we conducted a one-way repeated measures Analysis of Variance (ANOVA) to compare the effect of the gaming version on value of Flow Degree in versions A, B, C and D. Mauchly's Test indicated that the assumption of sphericity has been violated,  $\chi^2(5) = 18.367, p = .003$ . There was



a significant effect of which version the participant played,  $Wilks' \Lambda = 0.741, F(3, 26) = 3.025, p = 0.048$ . Greenhouse-Geisser correction determined that mean Flow Degree value did not differ statistically significantly between the gaming versions  $F(2.029, 56.826) = 1.961, p = 0.150$ . Although the test for this hypothesis is inconclusive, we can verify in Figure 11 that version C had the highest median ( $Mdn = 6.4$ ), compared to the ones of versions A ( $Mdn = 6$ ), B ( $Mdn = 6.2$ ) and D ( $Mdn = 5.8$ ). Also, all versions have values with similar IQR.

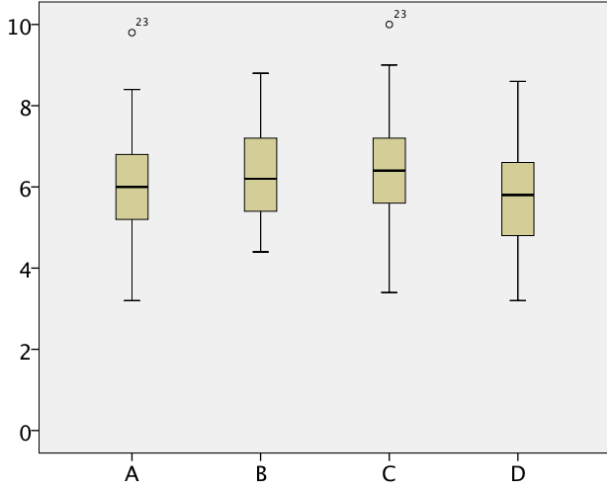


Fig. 11: Boxplots of the medians for the normalized Flow Degree values between versions.

Finally, we conclude that playing versions A and C led to a less frustrated feeling compared to version D, and this effect was medium in size for both versions, and that playing version B led to a less frustrated feeling compared to the control one with a effect large in size. Thus, hypothesis **H8** is accepted. Moreover, Figure 12 pictures the boxplots of the medians of the answers for this item across different versions. As expected, version D has the highest median ( $Mdn = 7$ ), compared to the ones of versions A ( $Mdn = 5$ ), B ( $Mdn = 2$ ) and C ( $Mdn = 6$ ). Versions A and B have similar IQR. In contrast, version C and version D have large and small IQR.

## V. DISCUSSION

This section discusses the validity of the guidelines presented in Table I. Tests for **H1** and **H5** were inconclusive, but we were able to find out that version A lets users significantly more anxious than the version B and less anxious than the version D. It is normal that the user felt more anxious in the version D than in the version A, since the difference between challenge and skills was greater in the former than in the later. Thus, users usually feel greater anxiety when they are frustrated rather than when they are simply anxious. We can conclude that our design guidelines for this version were accurate when distinguishing boredom from anxiety. Regarding distinguishing engagement from anxiety, we were not able to assess if the design guidelines provided a significant

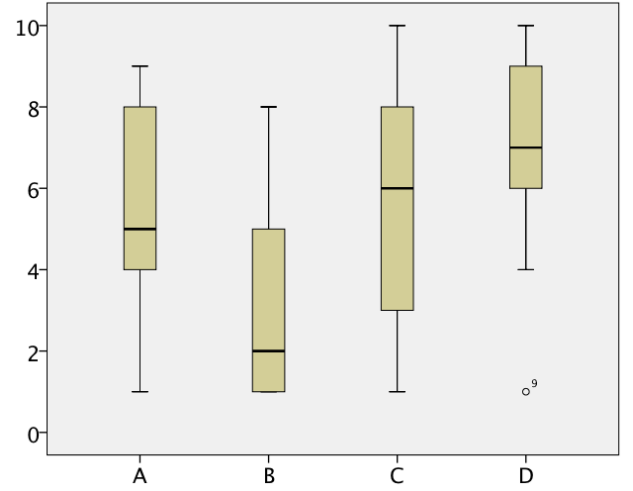


Fig. 12: Boxplots of the medians for the "I felt frustrated" item.

difference. Note that the median for predominant mental state was larger for engagement compared to anxiety's. We think that it happened because the enemies had more health points that they should have and it led users to feel more anxious because it took too long to kill one enemy.

Version B did not provide significant results. While the median of the item "I felt bored" was larger on that version compared to any other, the predominant mental state while playing this version was engagement. This may be given to the fact that a significant number of participants did not have regular experience with FPS, which led to a higher engagement since the difficulty was the lowest among all versions. Nonetheless, the median of the item "I felt bored" for every version was below five in a ten-point Likert scale, which tells us that users did not feel boredom. Therefore, we conclude that we were not able to successfully induce boredom on the participants with our guidelines.

Flow degree analysis shows that there was no significant difference between versions. In spite of not being able to verify that our guidelines were statistically significant, engagement was the predominant mental state while playing version C. Also, we see that the Flow Degree median was higher for version C and that all versions showed values above five in a ten-point Likert scale, meaning that, in general, people felt engagement in every version they played. We can conclude that our guidelines were able to induce approximately the same engagement level in all versions.

Lastly, the guidelines used for version D were a success, since there were significant differences between the values in the answers for the "I felt frustrated" item between versions. Moreover, playing version D led a predominant frustrated mental state with a median equal to anxiety. Therefore, we find that we successfully induced frustration with our guidelines, which are the following:

- Low enemy health;
- High enemy speed;

- Low enemy spawn time;
- Low shooting probability.

There are some important factors that may explain the lack of significance observed in some of our results. First, the number of participants in this experience was small. A larger number of participants would allow conclusions with a stronger impact. Second, we might have taken problematic design decisions and experimental methodology problems in the creation of the versions for mental state induction because the values were not validated beforehand and by assuming that three minutes were enough to record a calm mental state. Finally, our participants did not cover the whole range of player expertise, as most players did not play the game at least once a week.

Our findings provide additional evidence for inducing and investigating different mental states achieved while playing computer games, specially FPS. In particular, game developers may use the game design options to induce anxiety, boredom, engagement and frustration. These mental states can be useful to provide a better gaming experience, since game designers can change how the player feels accordingly to a game element or adapt the game based on the mental state of the user.

## VI. CONCLUSION

We investigated whether our guidelines are able to induce certain mental states. If so, leading gamers to certain mental states or a mental-based adaptation can be a further step in game development by designing game elements based on those guidelines.

In future studies, a larger sample size must be acquired in order to replicate these findings, with an increased female presence and number of participants that play FPS as well, since they are used to the game type and can provide a better feedback than players who only play mobile games, for example. Furthermore, we can validate the plots created in Section II-B, so that we can extend them to all the FPS genre. Lastly, we can use these guidelines to create game versions that induce specific mental states and record biofeedback of gamers while they are playing. This may allow us to model a classifier that can distinguish how the player is feeling, thus we can provide a way to address real-time gameplay adaptability.

## ACKNOWLEDGMENT

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013.

## REFERENCES

- [1] R. Lara-Cabrera and D. Camacho, "A taxonomy and state of the art revision on affective games," *Future Generation Computer Systems*, 2018.
- [2] R. W. Picard, "Affective Computing," *Pattern Recognition*, vol. 73, no. 321, p. 304, 1997.
- [3] A. Dekker and E. Champion, "Please biofeed the zombies: Enhancing the gameplay and display of a horror game using biofeedback," in *DiGRA Conference*, 2007.
- [4] G. N. Yannakakis, H. P. Martínez, and A. Jhala, "Towards affective camera control in games," *User Modeling and User-Adapted Interaction*, vol. 20, no. 4, pp. 313–340, 2010.
- [5] T. Alves, S. Gama, and F. S. Melo, "Flow-z: A flow-based adaptable game to maintain optimal challenge," MSc dissertation, Universidade de Lisboa - Instituto Superior Técnico, 2017.
- [6] B. Reuderink, A. Nijholt, and M. Poel, "Affective pacman: A frustrating game for brain-computer interface experiments," in *International conference on intelligent technologies for interactive entertainment*. Springer, 2009, pp. 221–227.
- [7] A. Drachen, L. E. Nacke, G. Yannakakis, and A. L. Pedersen, "Correlation between heart rate, electrodermal activity and player experience in first-person shooter games," in *Proceedings of the 5th ACM SIGGRAPH Symposium on Video Games*. ACM, 2010, pp. 49–54.
- [8] L. E. Nacke, M. N. Grimshaw, and C. A. Lindley, "More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game," *Interacting with Computers*, vol. 22, no. 5, pp. 336–343, 2010.
- [9] L. S. Colzato, P. J. Van Leeuwen, W. P. van den Wildenberg, and B. Hommel, "Doom'd to switch: superior cognitive flexibility in players of first person shooter games," *Frontiers in psychology*, vol. 1, 2010.
- [10] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an Understanding of Flow in Video Games," *Computers in Entertainment*, vol. 6, no. 2, pp. 20:1–20:27, 2008.
- [11] L. E. Nacke and C. A. Lindley, "Flow and Immersion in First-person Shooters: Measuring the Player's Gameplay Experience," *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*, pp. 81–88, 2008.
- [12] M. Grimshaw, *The acoustic ecology of the first-person shooter*. VDM Verlag, 2008.
- [13] H. Qin, P.-L. Patrick Rau, and G. Salvendy, "Measuring player immersion in the computer game narrative," *Intl. Journal of Human-Computer Interaction*, vol. 25, no. 2, pp. 107–133, 2009.
- [14] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. NY Harper & Row, 1990.
- [15] —, *Finding Flow: The Psychology of Engagement with Everyday Life*. Basic Books, 1997.
- [16] L. Harmat, Ö. de Manzano, T. Theorell, L. Högman, H. Fischer, and F. Ullén, "Physiological correlates of the flow experience during computer game playing," *International Journal of Psychophysiology*, vol. 97, no. 1, pp. 1–7, 2015.
- [17] M. Csikszentmihalyi, *Beyond boredom and anxiety*. Jossey-Bass, 2000.
- [18] K. M. Gilleade and A. Dix, "Using frustration in the design of adaptive videogames," *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '04*, pp. 228–232, 2004.
- [19] K. Poels, Y. De Kort, and W. Ijsselstein, "It is always a lot of fun!: exploring dimensions of digital game experience using focus group methodology," in *Proceedings of the 2007 conference on Future Play*. ACM, 2007, pp. 83–89.
- [20] L. Runciman, "Fun?" *College English*, vol. 53, no. 2, pp. 156–162, 1991.
- [21] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. Mcbroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the Game Engagement Questionnaire : A measure of engagement in video game-playing," *Journal of Experimental Social Psychology*, vol. 45, no. 4, pp. 624–634, 2009.
- [22] P. J. Lang, M. M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Affective ratings of pictures and instruction manual," *Technical report A-8*, 2008.
- [23] P. J. Lang, "The emotion probe: Studies of motivation and attention," *American psychologist*, vol. 50, no. 5, p. 372, 1995.
- [24] P. Lang and M. M. Bradley, "The international affective picture system (iaps) in the study of emotion and attention," *Handbook of emotion elicitation and assessment*, vol. 29, 2007.
- [25] P. J. Lang and M. M. Bradley, "Emotion and the motivational brain," *Biological psychology*, vol. 84, no. 3, pp. 437–450, 2010.