

Article

Promoting Optimal User Experience through Composite Challenge Tasks

Ricardo Cruz * and Luis A. Pineda * 

Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas,
Universidad Nacional Autónoma de México, Coyoacán 04510, Mexico

* Correspondence: crcruz@ciencias.unam.mx (R.C.); lpineda@unam.mx (L.A.P.);
Tel.: +52-55-3510-6331 (R.C.); +52-55-5622-3618 (L.A.P.)

Received: 10 August 2019; Accepted: 27 September 2019; Published: 1 October 2019



Abstract: Optimal user experience or *flow* is a theory with great impact on user experience. Promoting flow has become a competitive advantage for interactive systems, including rehabilitation. This can be achieved through an engaging interface that provides a rewarding experience and motivates the user to use the system again. This theory sustains that promoting a state of flow and improving task performance depends heavily on the balance between the challenges posed by the system and the skills deployed by the user. We further claim that balanced mental and motor skills demanded by the task improve flow and task performance. This paper presents an experiment supporting these claims. For this, we built two movement-interaction rehabilitation systems called SIBMER and Macoli (arm in Náhuatl). Both systems have two versions, one with a balanced load of mental and motor skills, and the other with an unbalanced one. Both versions are compared in terms of their potential to promote the state of flow and to improve task performance. Results show that a balance demand of mental and motor skills promotes flow, independently of the task complexity. Likewise, the experiment shows a correlation between flow and performance.

Keywords: interaction devices; tangible interaction; flow; optimal user experience; composite challenges; rehabilitation

1. Introduction

Optimal User Experience or *Flow* is a Human–Computer Interaction (HCI) theory focused on promoting a particular kind of human user emotional state that is desirable and can arise when users perform an intrinsically motivating activity [1,2]. This state, which is referred to here as *a state of flow*, is characterized roughly by being deeply involved and focused on a fully rewarding and satisfying experience. Although this theory originated in psychology it has been successfully adapted to computational environments [3–6] and more recently to systems supporting movement-based interaction, such as tangible and natural user interfaces, as well as virtual, augmented and mixed reality [7–9].

Promoting flow has become a competitive advantage for computer systems [10]. This can be achieved through an engaging interface that provides a rewarding experience and motivates the user to use the system again. Promoting flow improves positively the emotional state [5] and the learning process [3] too. In the setting of video games a state of flow improves performance, motivation, fosters an enjoyable experience and draws players to keep engaged [11]. For these reasons designers strive to integrate flow principles into their systems. However, there is still a lack of understanding of why some interfaces and environments promote more intensive flow experiences than others [12].

We subscribe the claim that the main condition to induce a state of flow is the balance between challenges and skills [13]. We consider further that the user needs to deploy physical and/or a mental

skill to meet the challenges, or that the challenges place a physical and/or a mental load upon the user. In case both kinds of loads are involved challenges are said to be *composite*. We hypothesized that an additional condition to improve flow is that composite challenges should also be balanced and developed an empirical study to test whether this can be sustained.

To carry out the experiment two interaction system called SIBMER and Macoli were designed and built. Both systems have two versions each implementing a balanced and an unbalanced challenge, so the question of whether the tasks involves composite or non-composite challenges can be settled intuitively. Both versions are compared to know which one promotes a better flow experience. Then the experimental setting and results are presented.

A warning note is that it is possible achieve a state of flow in purely physical or mental activities—e.g., chess or skiing—however, this paper focuses on tasks in which both kinds of loads are present, such as serious games and systems for supporting therapy. More generally, the scope of the theory can be characterized in terms of a plane with two orthogonal axes, one for the mental and the other for the physical load. Each dimension can be in turn divided into low and high load. The present theory focuses on the quadrants with the same value—i.e., low-low and high-high—of such plane. The quadrants with different values—i.e., low-high and high-low—have traditionally been the focus of standard flow studies. A more general approach would be to include an emotional dimension in addition to the physical and the mental, and to explore whether a balance between the three dimensions impact on flow, but we left such a study for further work.

2. Theory of Flow

The theory of flow sustains that there is an emotional state that arises when people perform a satisfying and rewarding experience. Csikszentmihalyi describes nine principal flow dimensions as factors to experience flow [14]: (1) Clear goals/objectives; (2) Unambiguous feedback; (3) Challenge–skills balance; (4) Action-awareness merging; (5) Concentration on the task at hand; (6) Feeling of control; (7) Loss of self-consciousness; (8) Time transformation; and (9) Autotelic experience. Novak et al. [15] present a state-of-the-art review and 13 dimensions including: (10) Excitement; (11) Exploratory behavior; (12) Interactivity; and (13) Telepresence, in addition to Csikszentmihalyi dimensions. Chen [16] reduces the number of constructs, but leaves out such important factors as clear goals, mix of action and alert and loss of self-consciousness, which are taken directly from the original work by Csikszentmihalyi [14,17]. However, it is not clear what concepts are necessary and which are contingent for promoting flow.

In recent studies Csikszentmihalyi states that conditions: (1), (2), and (3) promote flow in the course of performing a task [18]. In particular, if the activity is not challenging enough the user will become bored but if the challenge is too difficult he or she will get frustrated, so an adequate challenge level fosters the correct performance of the task [17]. In this study we focus on this last condition.

3. Balanced and Unbalanced Composite Challenges

Finneran and Zhang [19] hold that challenges are composite in movement-based interaction whenever they comprise physical and intellectual elements. Romero and Calvillo [20] consider that challenges are simple if they demand either particular physical or mental skill, and composite if they require both kinds of abilities. Hence, surfing the net, which is a mostly mental ability, is a simple challenge, but playing a dance video game, which involves motor coordination and kinesthetic memory, poses a composite challenge. Therefore, rehabilitation systems in which the patient/user carries out physical exercises supported by a computer system pose a composite challenge. This latter case requires deploying cognitive abilities—e.g., following an avatar on the screen—as well as coordinating motor behavior with virtual movements. In particular, two kinds of movement-based interaction systems that can benefit from composite challenges are rehabilitation systems and movement-based video games. In the former, composite challenges are inherent to the rehabilitation process; hence, encouraging a better flow state could lead to a better commitment and engagement. In the latter, the activity—mental

or physical—might require one dominating interaction mode, altering the balance between the composite challenge and the intensity of flow. However, the standard theory does not provide a precise definition or specific criteria to define physical or mental load in a challenge. We consider that challenges can be balanced or unbalanced depending on whether physical and mental abilities are demanded by the system for performing the tasks successfully in equal or unequal proportions.

A challenge needs an agent and a goal to be achieved; the agent possesses mental and physical skills, which are combined in varying degrees to overcome the challenge. The agent's skills can be visual, auditory, tactile, olfactory, gustatory and proprioceptive. Table 1 shows some examples of these kinds of skills. In desktop systems, challenges are traditionally considered simple because such systems demand mental skills mostly. In contrast, movement-based interaction systems make great use of physical skills. Hence, the scope of the study was to analyze challenges in movement-based interaction systems in which the activity goes beyond the use of a keyboard and mouse.

Table 1. The agent's skills characteristics.

Visual	Recognizing structures and models.
	Perceiving colors.
	Perceiving sizes and shapes.
	Recognizing position of objects.
	Estimating distances.
	Calculating weight or quantity.
	Calculating speed of objects in movement.
Auditory	Memorizing objects.
Auditory	Recognizing sound patterns or sequences.
	Recognizing the direction from which sounds come from.
Tactile	
	Recognizing soft and hard surfaces.
	Recognizing surfaces temperatures.
	Detecting the force applied on or by objects.
	Holding objects.
	Perform intense movements with hands.
	Perform light movements with hands.
Olfactory	Perform intense movements with your feet.
	Perform light movements with your feet.
Gustatory	Recognizing smells.
Proprioceptive	Recognizing tastes.
Proprioceptive	Recognizing inner feelings and stimuli.
	Perceiving movements of our own body, extension of extremities.
	State of equilibrium.

Composite challenges can be in turn balanced or unbalanced, as defined above. The operational use of the previous definitions relies on the capacity of measuring the level of motor and mental load demanded by the challenge. In the literature two methodologies are normally used for this purpose: (1) through physiological evaluations that measure cardiac, ocular, muscular and cortical activity, and (2) through psychological evaluations using questionnaires, surveys, scales and profiles. The problem associated with these kinds of methodologies is their generality, as these provide only global scores. Hence, assessing whether a challenge is more physical or mental and whether the required skills are balanced through these methodologies is a subjective exercise. In practice, the most common method is to apply a questionnaire to quantify the level of skills employed in a task. For example, The National Aeronautics and Space Administration Task Load Index questionnaire (NASA-TLX) [21].

In this paper, we present a methodology to measure mental and motor demands. This consists on applying a psychological evaluation using a Likert-type questionnaire with two scales, one for the mental and the other for the motor skills demands. For the construction of the questionnaire we first

reviewed: (1) The NASA-TLX which measures the mental and physical load, the temporal and effort demands and the performance and frustration levels; and (2) The Subjective Workload Assessment Technique (SWAT) which measures mental load, effort and stress. The questions were designed to assess the mental and physical load through the demands imposed by the task upon the modalities of perception (visual, auditory, gustatory, tactile, olfactory and proprioceptive). The questions were worded through a template of the following form:

How α demanding was the task to β

where α was either “physically” or “mentally” and β was of form verb + task. For instance “How physically demanding was the task to recognizing structures and models?”; “How mentally demanding was the task to recognizing structures and models?” The original questionnaire in Spanish and its translation into English are available at: https://gitlab.com/carlosricardocm/composite_challenges.

Modalities such as olfactory and gustatory were not included in the tasks. Questions related to such modalities were used as control—e.g., How physically demanding was the task to recognizing smells?—and subjects that assigned a significant value were discarded.

A validation study was carried out using the Cronbach’s alpha coefficient (Cronbach, 1951) and the SPSS software. The formula for Cronbach’s alpha is:

$$\alpha = \frac{N * \bar{c}}{\bar{v} + (N - 1) * \bar{c}} \quad (1)$$

where:

- N = the number of items
- $\bar{c}$ = average covariance between item-pairs
- $\bar{v}$ = average variance

Each response was used as an item/variable to calculate the Cronbach’s alpha coefficient in SPSS. The result for the alpha coefficient was 0.957. As a general criterion, an alpha coefficient > 0.9 is excellent, thus the validity and reliability of the questionnaire is well supported. The protocol was applied as follows:

1. The participant answered a questionnaire with personal information (Name, age, sex and education level).
2. Each participant was assigned to an interleaved group (Balanced–Unbalanced BG–NBG or Unbalanced–Balanced NBG–BG).
3. Each participant played a tutorial explaining the video game according to the assigned group version.
4. The participants played the video game corresponding to the assigned group version. The playing session was one minute per version.
5. The participants answered the personalized questionnaire.

We consider that a challenge is balanced when the difference between the global mental and motor demand is less or equal to a similarity threshold T , as shown in Equation (2):

$$\left| \sum_{i=1}^n MentalDemand_i - \sum_{i=1}^n MotorDemand_i \right| \leq T \quad (2)$$

The similarity threshold is 5% of the product of the total number of skills considered, the maximum score and the number of demands, as follows:

$$T = (A * M * D) * 0.05 \quad (3)$$

where A is the total number of abilities, M is the maximum score a user can assign to an ability and D is the number of kinds of demands (mental and motor).

The five percent is used as the standard error accepted in statistical analysis such as the alpha number which is a standard criteria to accept or reject the hypothesis of a significance test.

Additionally, the NASA-TLX questionnaire is applied. Results of the NASA-TLX provide a score for the mental load or demand. The overall average values of the Cumulative Frequency Distributions of NASA-TLX Global Workload Score by Task Type [22] for the relevant tasks are shown in Table 2.

Table 2. Minimum, 50th percentiles and Maximum averages values grouped by task.

Tasks	Average Minimum Values	Average 50th Percentiles	Average Maximum Values
Cognitive Activities	13.08	46	64.90
Video games	14.08	56.60	78.00
Physical activities	40.83	62	75.19
Computer activities	7.46	54	78.00
Overall average	18.86	54.62	74.02

Based on Table 2 we propose Table 3 to assess the results of the global NASA-TLX score.

Table 3. Assessment table for the global NASA-TLX score.

Global Score	Level of Mental Load/Demand
From 18.85 to 54.62	Low
From 54.62 to 74.02	Moderate

In the present methodology three scores are obtained: the motor demand global score; the mental demand global score; and the NASA-TLX questionnaire score, which is used to verify the results obtained in the mental demand of the personalized questionnaire.

In the following section we describe SIBMER and Macoli along with their game-like environment that implements two versions of composite challenges, one involving balanced and the other unbalanced mental and motor skills.

4. SIBMER

SIBMER is a movement-based interaction system based on virtual reality. It was developed with two main aims: (1) to analyze the relationship between composite challenges and movement-based interaction systems, and (2) to rehabilitate patients with shoulder injuries doing flexion and extension arm exercises.

4.1. System Operation

The main objective in SIBMER is to control a spaceship that crosses portals while avoiding asteroids and collects time spheres in a virtual 2.5D environment. The movements of the hand are mapped into the movements of the ship through the Microsoft Kinect sensor. The score is increased each time the ship crosses a portal and is reduced each time it collides with an asteroid; the ship has a shield whose intensity is reduced each time the spaceship collides with an asteroid. In the Head-Up Display the intensity of the shield is shown on the right side (see Figure 1). The game ends in two ways: when the ship reduces the intensity of the shield to zero or when time runs out; however, the user can obtain a time bonus every time he captures a time sphere. There is also auditory feedback through sounds that are reproduced when the user crosses a portal or captures a time-lapse, and visual feedback through a graphical face that smiles, grieves or remains expressionless depending on the user performance.



Figure 1. Graphical Items. (1) Spaceship; (2) Portal; (3) Asteroid; (4) Score; (5) Shield intensity (6) Timer remaining; (7) Hand used; (8) Feedback face image.

4.2. Architecture

SIBMER consists of three main components: (a) a video game; (b) an administrative software; and (c) a Kinect sensor. This sensor detects the user and creates a 3D skeleton, which is sent to the video game to map the movements of the hand to the movements of the ship (Figure 2).

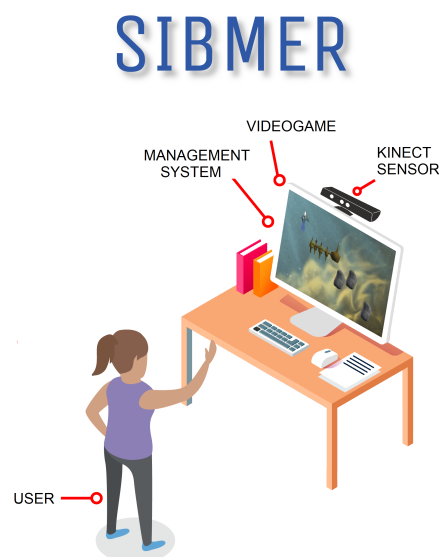


Figure 2. Architecture of SIBMER.

The video game was built using the Microsoft XNA framework and programming language c#, using Visual Studio 2012 as IDE. The video game and the management system run in a PC with windows 8.1.

4.3. Video Game and Composite Challenges

The video game implements two challenges: one balanced and the other unbalanced. In both versions the objective is to perform flexion, abduction and adduction exercises in one minute (which can be extended to a maximum of two and a half minutes if the user gets all the time bonuses).

An experiment was carried out with the purpose of validating that the challenges correspond to the expected type (balanced or unbalanced) as follows: 20 subjects were selected, 10 women and 10 men aged between 18 and 28 years, with a minimum level of high school. The users were not aware of the purpose of the experiment in advance or the scenery type. The independent variable is the type of challenge, while the dependent variables are the global score of the NASA-TLX questionnaire and the total mental and motor demand of the questionnaire designed for this investigation. The participants played the two versions of the game interspersed.

The main objective is to control a spaceship that crosses portals while dodging asteroids and taking time spheres in a virtual 2.5D environment in both versions. In version V1 (Watch video <https://youtu.be/q3udhkCiS6A>), the ship only must cross the largest number of portals avoiding colliding with asteroids. However, in version V2 (Watch video <https://youtu.be/j0SfNI3pvME>), the user is notified beforehand of a sequence of colors and the spaceship has to cross the portals following such specific color order, as well as avoiding to collide with the asteroids. The color sequence has a variable length from three to seven, which is the maximum sequence that a normal user could remember.

The abilities and demands considered in versions V1 and V2 are shown in Tables 4 and 5 respectively; both values correspond to the mean of the results assigned by the 20 users taking off outliers. The excel formula “TRIMMEAN” was used to discard outliers excluding a percentage of data points from the top and bottom tails of a data set. A percentage of 20% was used so 1 value will be discarded from each end of the range before calculating the mean of the remaining values.

Table 4. Mean for the abilities and demands of SIBMER version V1.

Modality	Ability	Motor Demand	Mental Demand
		0 1 2 3 4 5 Low - Moderate - High	0 1 2 3 4 5 Low - Moderate - High
Visual	Recognizing structures and models.	1.36	2.05
	Perceiving colors.	0.2	1.91
	Perceiving sizes and shapes.	0.43	1.81
	Recognizing position of objects.	1.47	2.63
	Estimate distance.	2.23	2
	Calculating weight or quantity.	1.15	1.9
	Calculating the velocity of objects in movement.	2.16	3
	Memorizing objects.	0	0.22
Auditory	Recognizing sound patterns or sequences.	0.25	1.27
	Recognizing the direction from which sounds come from.	0.13	0.81
Tactile	Recognizing soft and hard surfaces.	0.26	1.41
	Recognizing surface temperatures.	0.23	0.43
	Detecting the force applied on or by objects.	1	1.55
	Holding objects.	1.23	1.66
	Perform intense movements with your hands.	2.6	2.33
	Perform light movements with your hands.	2	1.55
	Perform intense movements with your feet.	0	0
	Perform light movements with your feet.	0	0.66
Olfactory	Recognizing smells.	0	0.66
Proprioceptive	Recognizing inner feelings and stimuli.	1.2	1.5
	Perceiving movements of our own body, extension of extremities.	1.68	1.89
	State of equilibrium.	1.73	1.5
	TOTAL	21.37	31.62

Table 5. Mean for the abilities and demands of SIBMER version V2.

Modality	Ability	Motor Demand 0 1 2 3 4 5	Mental Demand 0 1 2 3 4 5
		Low - Moderate - High	Low - Moderate - High
Visual	Recognizing structures and models.	1.11	2.55
	Perceiving colors.	0.66	2.4
	Perceiving sizes and shapes.	0.84	2.05
	Recognizing position of objects.	0.625	2.4
	Estimate distance.	1.7	2.4
	Calculating weight or quantity.	1.55	1.5
	Calculating the velocity of objects in movement.	2.15	2.8
	Memorizing objects.	0.89	5
Auditory	Recognizing sound patterns or sequences.	0.94	2.65
	Recognizing the direction from which sounds come from.	0.38	1.8
Tactile	Recognizing soft and hard surfaces.	0.88	1.6
	Recognizing surface temperatures.	0.68	1.3
	Detecting the force applied on or by objects.	1.4	1.85
	Holding objects.	1.75	1.35
	Perform intense movements with your hands.	2.75	1.85
	Perform light movements with your hands.	1.95	1.68
	Perform intense movements with your feet.	0	0
	Perform light movements with your feet.	0.23	0
Olfactive	Recognizing smells.	0.33	0.4
Proprioceptive	Recognizing inner feelings and stimuli.	1.65	1.8
	Perceiving movements of our own body, extension of extremities.	1.25	2
	State of equilibrium.	1.6	1.65
	TOTAL	25.36	41.03

For the similarity threshold in SIBMER, $A = 22$, $M = 5$ and $D = 2$, hence:

$$T = (22 * 5 * 2) * 0.05 = 11 \quad (4)$$

The version V1 is balanced because:

$$\left| \sum_{i=1}^n MentalDemand_i - \sum_{i=1}^n MotorDemand_i \right| \leq T$$

$$|31.62 - 21.37| < 11$$

The version V2 is unbalanced because:

$$\left| \sum_{i=1}^n MentalDemand_i - \sum_{i=1}^n MotorDemand_i \right| \leq T$$

$$|41.03 - 25.36| > 11$$

The V1 is balanced because the motor demand –performing intense movements to move the spaceship with the hand– and the mental demand –recognizing the position of objects, estimating distances, estimating the speed of objects in movement– occur in a similar proportion. The V2 is unbalanced because the extra mental activities –memorizing objects and perceiving colors– affect the balance between the rest of the abilities.

The results of the NASA-TLX in turn are shown in Table 6 as follows:

Table 6. Global score results of the NASA-TLX questionnaire.

Global Score for NASA-TLX V1	Global Score for NASA-TLX V2
47.85	55.486

The mental scores of the personalized questionnaire with the TLX NASA global scores show that the mental load is low in the balanced version while moderate in the unbalanced one.

5. Flow Experiment

The previous measure of whether the motor and mental demands of the tasks are balanced or not allows testing of the hypothesis of whether such balance promotes a state of flow and an experiment to find out whether this is in fact the case was carried out. The experimental protocol is as follows:

1. The participant answered a questionnaire with personal information (Name, age, sex, degree and hours of experience per day playing Kinect).
2. Each participant was assigned to an interleaved group (Balanced–Unbalanced BG–NBG or Unbalanced–Balanced NBG–BG).
3. Each participant played a tutorial explaining the video game according to the assigned group version.
4. The participant played the video game corresponding to the assigned group version. The playing session was one minute per version.
5. The participant answer the flow status questionnaire.
6. The participant played the second video game.
7. The participant answer again the flow status questionnaire.

Both groups are compared in terms of their potential to promote the state of flow and to improve task performance. The hypotheses are as follows:

H0: *There is no significant difference between the means of the flow score.*

H1: *There is significant difference between the means of the flow score.*

The assessment was based on the Flow State Scale or FSS [23], which measures quantitatively the flow level achieved by people participating in physical activities in general; however, this scale has been adapted to other contexts such as work, learning and computing environments [24]. In this study, we used a Spanish version of the scale translated by Calvo et al., which is a coherent and valid equivalent of its English version [25]. The FSS consists of 36 Likert-type items/questions that go from 1 (strongly agree) to 10 (strongly disagree). The present scale provides a score to measure the total flow and nine second-order elements that refer to the nine flow dimensions (4 questions for each dimension). Table 7 shows a mapping between the flow dimensions and the FSS question numbers.

Table 7. Flow Dimensions and FSS Questions Mapping.

Type	Dimension	Question
Type 1	Challenge-skill balance	1, 10, 19, 28
Type 2	Action-awareness merging	2, 11, 20, 29
Type 3	Clear goals/objectives	3, 12, 21, 30
Type 4	Clear feedback	4, 13, 22, 31
Type 5	Full concentration on the task	5, 14, 23, 32
Type 6	Feeling of control	6, 15, 24, 33
Type 7	Loss of consciousness	7, 16, 25, 34
Type 8	Time transformation	8, 17, 26, 35
Type 9	Autotelic experience	9, 18, 27, 36

The following are some example of items on the scale:

- Type 1. I knew that I was competent enough to meet the high demands of the situation.
- Type 2. I performed automatically.
- Type 3. I knew clearly what I wanted to do.
- Type 4. It was really clear to me what I was doing well.

- Type 5. I was completely focused on what I was doing.
- Type 6. I felt in total control of what I was doing.
- Type 7. I was not concerned with what others might have been thinking of me.
- Type 8. Time seemed to be altered.
- Type 9. I really had fun with what I was doing.

The full set of questions is available in [23,25].

The sample included 38 people, 19 men and 19 women, aged from 19 to 56 years old, with under and graduate degrees, all clinically healthy, that did not participate in the previous experiment. They had neither previous knowledge of the experiment objectives nor of the type of scenarios in which they had to perform. The independent variable was Balanced Game (BG) and Non-Balanced Game (NBG) while the dependent variables were the FSS answers and the performance scores.

Results

The Student's *t*-test was applied to check the statistical significance of the experiment. This test is used for verifying a statistical hypothesis about the mean of a normal population with dependent samples of a relatively small size [26]. As the data is dependent—i.e., all the participants played both versions but in different order—the Student's *t*-test was split in two parts: in the first the participants played first the BG and then the NBG versions and in the second played the reversed order. These results are shown in Tables 8 and 9 respectively. The parameter *t* is the computed *t* statistic, as explained below, and *df* stands for degrees of freedom. The significance level was $\alpha = 0.05$ and the analysis was performed using the GNU PSPP software [27].

The Student's *t*-test holds only when the data has a normal distribution. The Shapiro–Wilk test [28] is used to verify this latter condition for small samples (less than 50). If the Sig. value of the Shapiro–Wilk test is greater than 0.05, the data is normal; otherwise the data deviate significantly from a normal distribution. The Sig. value for the present data is 0.952, hence it follows a normal distribution and the Student's *t*-test is applicable.

Table 8. Result of the Student's *t*-test for the BG–NBG group.

Related Differences								
	Mean	Deviation Standard	Standard Error	95% Confidence Interval for the Difference		<i>t</i>	df	Bilateral Significance
				Lower	Upper			
BG–NBG	0.24750	0.27059	0.09567	0.2128	0.47372	2.587	18	0.036

Table 9. Result of the Student's *t*-test for the NBG–BG group.

Related Differences								
	Mean	Deviation Standard	Standard Error	95% Confidence Interval for the Difference		<i>t</i>	df	Bilateral Significance
				Lower	Upper			
NBG–BG	−0.07000	0.36174	0.12790	−0.372242	0.23242	−0.547	18	0.601

The tested hypothesis is accepted—and the null hypothesis is rejected—if the bilateral significance or *p*-value is ≤ 0.05 and vice versa. The *p*-value in Table 8 is 0.036; hence, there is a statistical significance to support that the average in the global BG flow score is higher than the average of the NBG when the users played the game in this order. The mean is calculated by subtracting the mean of the second group from the mean of the first. The sign of the mean difference corresponds to the sign of the *t* value. A positive *t* value, as in this case, indicates that the mean for the BG is significantly greater than the mean for NBG.

However, Table 9 shows that the p -value is 0.601 and null hypothesis is accepted. This indicates that there is no statistical significance to affirm that the average in the global BG flow score is higher than the average of the NBG when the users played the NBG game first.

These results accord with the intuition that playing first a balanced game improves flow, and when the setting changes to an unbalanced one the flow is diminished. The intuition on the second setting should be that the flow is low playing a non-balanced game, but it should improve when the balanced one is played afterwards. However, the result is not significant in the latter order. We noticed that participants rated the unbalanced version better in 9 out of the 19 cases, so they tend to prefer the unbalanced version slightly regardless of the challenge. This suggests that the evaluation of the second game played is biased by the first or perhaps that the participants felt more comfortable after playing the first game, or that they liked the unbalanced one better because it is more complex and attractive.

Hence, the experiment with SIBMER was inconclusive and further research and experiments were needed. We could have tested the two versions with independent groups to rule out order effects, but there remains the fact that the unbalanced version is significantly more complex than the balanced one, and this factor should also be controlled. For these reasons, a second system (Macoli) was developed and evaluated. The purpose of this latter exercise was to assess whether flow is due to the balance between motor and mental demands, independently of the order of play, and also of the complexity or how attractive a system may be in relation to others.

6. MACOLI

Macoli is a movement-based interaction system. It was developed with two aims: (1) to analyze the relationship between composite challenge and movement-based interaction systems targeted to rehabilitation scenarios and (2) to solve the inconclusive research and experiments made in SIBMER.

6.1. System Operation

Macoli is used to do flexion exercises through a series of movements during which the user places a bar on each step of the wall-bar/ladder technological device. The video game displays a musical staff where each line corresponds to a step on the ladder. The number of lines in the staff is defined according to the limitations of the user's arm reach. Each line shows a musical note at each time interval. The notes appear on the checkpoint at the far right of the staff or appear at the left extreme of the staff and move towards the checkpoint according to the version of the game, as explained below. The objective of the video game is to synchronize the placement of the bar on the corresponding step with the time when the note appears on or reaches the checkpoint (push-button) (Figure 3) (See video at <https://youtu.be/SwyO2Q9m09o>). If the user places the bar on the right step at the right time the system plays the melody right. Therefore, good performance is rewarded with a fine melody while a diminished performance is penalized with a poor tune. Complementing the device with a serious video game brings a ludic component to rehabilitation exercises, promoting the engagement of the patient in his or her rehabilitation.

Each user carry out a calibration session in which two parameters are set up: (1) The highest step that the user can reach (out of seven), and (2) The height of the lowest step which is determined according to the user's arm mobility level. Using these parameters, the video game creates staff lines dynamically; for example, whenever a user is able to place the bar on the first and second steps, the system generates and renders the corresponding two-line staff. As the sessions progress, the system provides new challenges to maintain the user engaged; for example, the system can increase the speed on the notes, create a new line and occasionally display notes on it, and add extended notes so the user has to keep the bar on the same step for longer periods of time. Each session consists of the calibration step, a training exercise (supported by a tutorial) and the exercise session properly in which the user plays the video game.

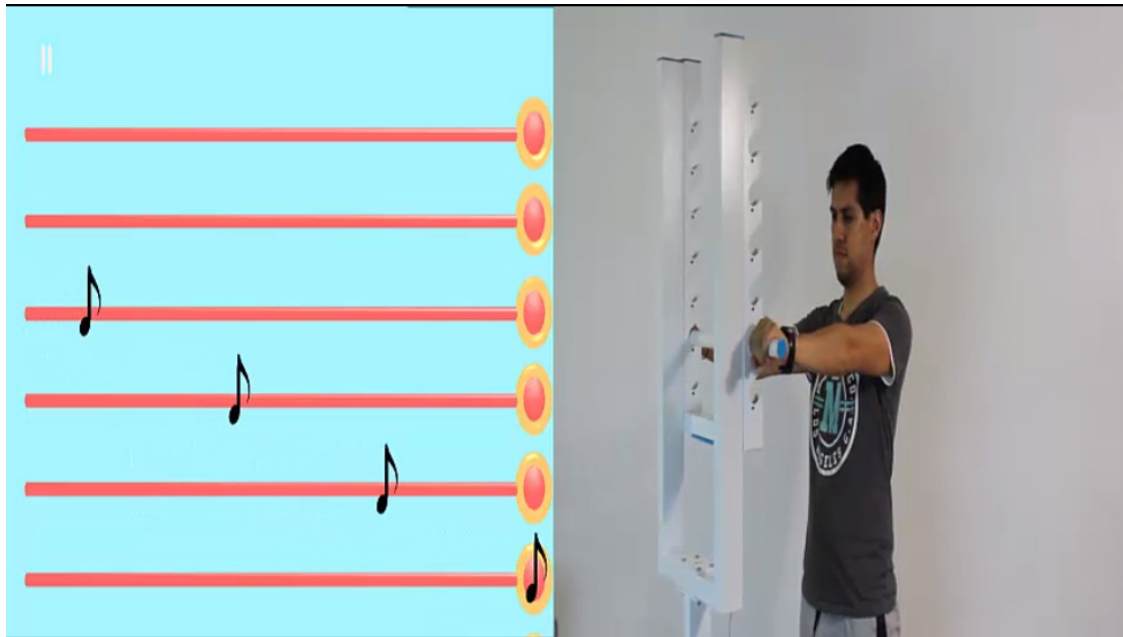


Figure 3. Left: Serious video game screenshot. Right: User using the ladder technological device.

6.2. System Architecture

Macoli consists of three main components: a video game-software (Figure 4-1), a management system for users and administrative data-software (Figure 4-2) and a technological device in the form of wall-bars or gymnastics ladder-hardware (Figure 4-3).

macoli

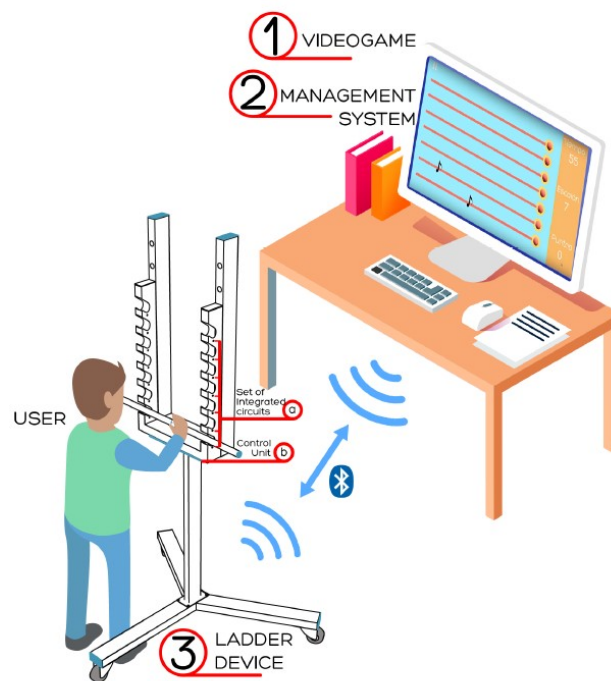


Figure 4. Macoli Architecture.

The ladder device has two main components: a set of integrated circuits (Figure 5) for sensing the step in which the PVC bar is placed (one for each side of the ladder) and the control unit (Figure 6) that receives and transforms each integrated circuit signal into a number that indicates the step in which the ladder was placed. Once this number is processed it is sent to the video game through a Bluetooth connection.

Each custom integrated circuit (Figure 5(a3)) has a proximity sensor and two resistance, an acrylic base (Figure 5(a2)) and an acrylic cover (Figure 5(a1)). The control unit has an Arduino mega (Figure 6(b1)), a Bluetooth module (Figure 6(b2)), a power supply (Figure 6(b3)), an on/off switch (Figure 6(b4)) and a main support base (Figure 6(b5)). All software and hardware elements were developed in-house.

The video game was built using the Unity Game Engine and programming language c#. The management system was developed with Visual Studio 2013 and programming language XAML and c#. The video game and the management system run in a PC with windows 8.1. The video game has screens to select a song, the difficulty level and time session. Once this setup is finished, the musical staff, score and session time are rendered on the screen.

The video game has two versions: In the first the notes appears immediately above the push-buttons and in the second the notes appear at the far left extreme of the staff and move towards the push-buttons. In both cases, the video game senses the numbers received by the ladder device (via Bluetooth) and checks if the input is right and the score is incremented. At the end of session, the video game provides a summary of the user's performance.

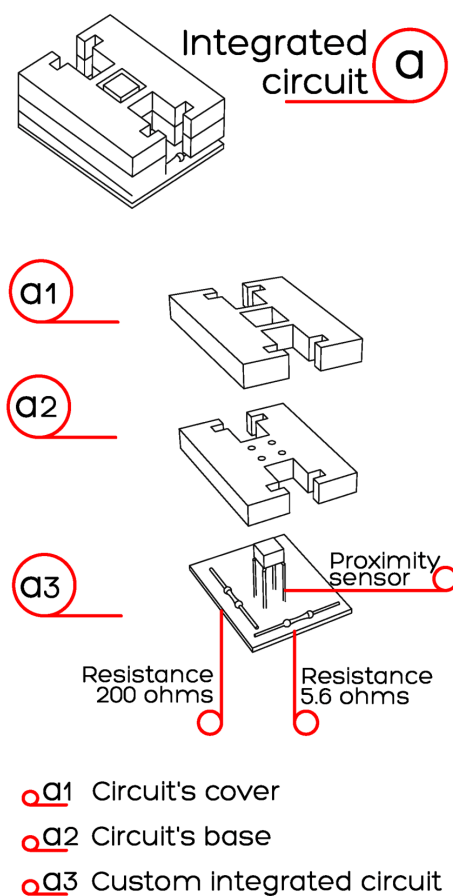


Figure 5. Integrated Circuit detail.

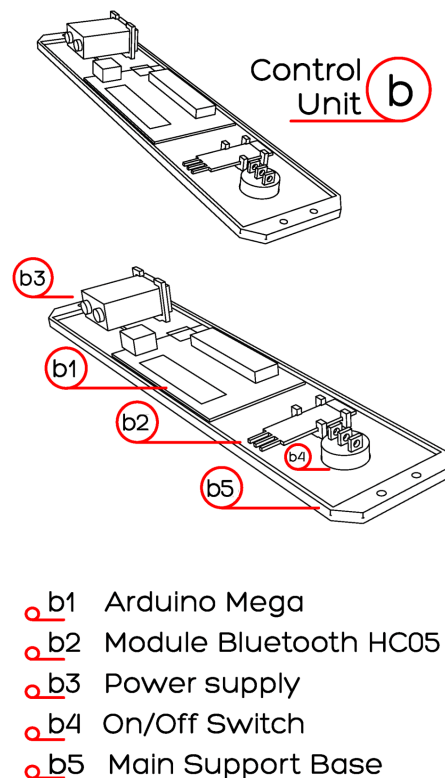


Figure 6. Control Unit detail.

6.3. Video Game and Composite Challenges

The objective of the game in its two versions is to perform the same number of flexion exercises as notes appear on the staff by placing the bar on the device for a period of time, which was set to one minute.

In the first version (V1) the notes move along the staff at a constant speed. The pattern forces the user/patient to coordinate the exact time of placement the bar with the moment the note goes through the checkpoint. The user makes use of his or her tactile skills in order to move the bar from one step to another and also of his or her visual and auditory skills in order to estimate the speed of musical notes and to recognize the melody (Table 1) (See video at <https://youtu.be/mQ-T2pfitWk>).

In the second version (V2), the notes appear directly on the checkpoints, so the user places the bar from one step to another without worrying about synchronizing the time when the note reaches the push-button. In this setting the demand of visual skills is lower than in the first one because there is no need to estimate the time when the note reaches the checkpoint and the auditory key guides were removed (See video <https://youtu.be/62OMuFrWzU>).

To validate the balance of each challenge an experiment was carry out. 20 subjects were selected, 10 women and 10 men with ages between 18 and 28 years, with a minimum level of bachelor's degree. The users were not aware of the purpose of the experiment in advance nor of the type of scenario. The independent variable is the type of challenge, while the dependent variables are the global score of the NASA-TLX questionnaire and the total mental and motor demands of the questionnaire designed for this investigation, as explained above. The participants played the two versions of the game (V1 and V2) interspersed.

The abilities and demands considered in versions V1 and V2 are shown in Tables 10 and 11 respectively; both values correspond to the mean of the results assigned by 20 users taking off outliers. The excel formula TRIMMEAN was used for this purpose.

Table 10. Mean for the abilities and demands of Macoli version V1.

Modality	Ability	Motor Demand	Mental Demand
		0 1 2 3 4 5 Low - Moderate - High	0 1 2 3 4 5 Low - Moderate - High
Visual	Recognizing structures and models.	2.5	2.7
	Perceiving colors.	1.05	2
	Perceiving sizes and shapes.	1.05	1.9
	Recognizing position of objects.	1.45	2.1
	Estimate distance.	2.8	3.1
	Calculating weight or quantity.	1	1.7
	Calculating the velocity of objects in movement.	2.1	2.5
	Memorizing objects.	2.1	2.5
Auditory	Recognizing sound patterns or sequences.	1.2	4.7
	Recognizing the direction from which sounds come from.	1.9	2.65
Tactile	Recognizing soft and hard surfaces.	1.8	3.3
	Recognizing surface temperatures.	0.9	1.6
	Detecting the force applied on or by objects.	1.4	1.3
	Holding objects.	1.75	1.85
	Perform intense movements with your hands.	2.75	1.85
	Perform light movements with your hands.	2.5	1.6
	Perform intense movements with your feet.	2.75	1.85
	Perform light movements with your feet.	2	1.7
Olfactory	Recognizing smells.	0.75	0.55
Proprioceptive	Recognizing inner feelings and stimuli.	1.1	0.8
	Perceiving movements of our own body, extension of extremities.	1.25	2
	State of equilibrium.	1.6	1.65
	TOTAL	35.8	43.95

Table 11. Mean for the abilities and demands of Macoli version V2.

Modality	Ability	Motor Demand	Mental Demand
		0 1 2 3 4 5 Low - Moderate - High	0 1 2 3 4 5 Low - Moderate - High
Visual	Recognizing structures and models.	2.1	3
	Perceiving colors.	0.9	2.4
	Perceiving sizes and shapes.	0.8	2.05
	Recognizing position of objects.	0.8	2.6
	Estimate distance.	2.4	3.5
	Calculating weight or quantity.	1.55	1.85
	Calculating the velocity of objects in movement.	2.15	2.8
	Memorizing objects.	0.4	4.4
Auditory	Recognizing sound patterns or sequences.	0.1	0.3
	Recognizing the direction from which sounds come from.	0.1	0.3
Tactile	Recognizing soft and hard surfaces.	1.15	2.6
	Recognizing surface temperatures.	1.4	1.3
	Detecting the force applied on or by objects.	2	1.85
	Holding objects.	2.5	1.5
	Perform intense movements with your hands.	2.6	1.9
	Perform light movements with your hands.	2.8	2.1
	Perform intense movements with your feet.	2.1	1.8
	Perform light movements with your feet.	1	1.2
Olfactory	Recognizing smells.	0.5	0.55
Proprioceptive	Recognizing inner feelings and stimuli.	1.1	0.8
	Perceiving movements of our own body, extension of extremities.	1.25	2
	State of equilibrium.	1.6	1.65
	TOTAL	31.3	42.45

For the similarity $A = 22$, $M = 5$ and $D = 2$, hence:

$$T = (22 * 5 * 2) * 0.05 = 11 \quad (5)$$

The version V1 is balanced because:

$$\left| \sum_{i=1}^n MentalDemand_i - \sum_{i=1}^n MotorDemand_i \right| \leq T$$

$$|43.95 - 35.8| < 11$$

The version V2 is unbalanced because:

$$\left| \sum_{i=1}^n \text{MentalDemand}_i - \sum_{i=1}^n \text{MotorDemand}_i \right| \leq T$$

$$|31.3 - 42.45| > 11$$

These results are due to the fact that in the balanced version the motor abilities—moving the bar through the steps—are considered similar to the visual and auditory abilities—estimating the speed of the notes and recognizing the sound pattern—while in the unbalanced version there are no such restrictions; for example, in V2 there was a small time interval to place the bar on the step—versus the exact moment in which the note reaches the checkpoint in V1—and the mental activities that affected the balance—recognize sound patterns, estimate speeds, etc.—were not present.

Overall, the balanced version V1 is more complex and more attractive than the unbalanced version V2 which is simpler and perhaps less exciting.

The results for evaluating the NASA-TLX are shown in Table 12. These show that the mental load is low for both versions of the game.

Table 12. Global score results of the NASA-TLX questionnaire.

Global Score for NASA-TLX V1	Global Score for NASA-TLX V2
49.7	46.87

The mental scores of the personalized questionnaire and the TLX NASA global scores taken together show that the mental load in version V1 is greater than the load of version V2.

7. Experiment and Results

As before, we conducted an experiment to determine the extent to which tasks V1 and V2 promote a state of flow. The participants were divided in two groups: (1) the balanced challenge (BG) and (2) the unbalanced one (NBG).

The hypotheses are as follows:

H0: *There is no significant difference between the means of the BG and NBG.*

H1: *There is a significant difference between the means of the BG and NBG.*

Performance was registered through a score. In the BG scenario, the score is the number of times the participant places the bar on the device at the exact moment that the note passes through the checkpoint; and in the NBG scenario, is the number of times the participant places the bar on the device at the moment a note appears on the checkpoint (maximum score of 53 in both settings).

The hypotheses in this aspect of the experiment is whether performance correlates with flow:

H0: *Performance is not correlated with flow.*

H1: *The performance is correlated with flow.*

The assessment was based on the Flow State Scale or FSS [23]. A new sample with the same characteristics of the one used for the experiments with SIBMER was selected, with the only difference that the age range was from 19 to 52 years in this latter case. The independent and dependent variables were also as before.

The experiment was conducted as follows:

1. The participants answered a personal data questionnaire and read the Macoli operating guide.
2. BG or NBG scenarios were interspersed assigned to each participant.
3. Each participant performed a tutorial session.
4. Parameters for level of difficulty, number of steps, duration time and musical melody information were set up (See video at <https://youtu.be/SwyO2Q9m09o>, minute 2:40). For the validation conditions to be equal during the experiment, all the participants performed the test with an initial difficulty level of one—i.e., how fast the notes appear on the staff, the seven steps, one minute session time and the melody *jingle bells*.
5. The experimental session was carried out.

Results

The hypotheses of flow and performance were analyzed independently. The results with the values of the means for the global flow score are shown in Table 13 where the global BG mean is greater than the global NBG mean. The Student's *t*-test for independent samples is used to validate whether this result is significant [26].

Table 13. Mean and deviation standard of the Balanced and Unbalanced challenges.

Challenge	N	$\bar{x}$	σ	Error $\bar{x}$
BG	19	8.3955	0.78433	0.17994
NBG	19	7.7759	0.97911	0.22462

The Student's *t*-test for independent samples assumes the homogeneity of the variance. Levene's test [29] is used to verify this latter condition. Table 14 shows the results of the Student's and Levene's tests considering whether equal variances are assumed or not, as indicated in the corresponding lowest two rows. A Sig. value greater than 0.05 indicates that the equality of variances is assumed and vice versa. In this case the value is 0.23; hence the variances are equal. Then the bilateral significance or *p*-value is considered. If this value is ≤ 0.05 the hypothesis is accepted and vice versa. In this case the value is 0.038; hence there is a significant difference between the means of the BG and NBG and the hypothesis is accepted—i.e., the mean of flow in BG is greater than the mean of flow in NBG.

Table 14. Student's *t*-test result for Balanced vs Unbalanced challenges.

	Levene's Test for Equality of Variances		t-test for Equal Means						
	F	Sig.	t	df	Bilateral Significance	Mean Difference	Std. Error Difference	Confidence Interval	
								Lower	Upper
Equal variances assumed	1.486	0.231	2.15	36	0.038	0.62	0.29	0.4	1.2
Equal variances not assumed			2.15	36	0.038	0.62	0.29	0.4	1.2

To assess performance, the condition was divided in two parts: (1) performance (BG) correlated with flow (BG), and (2) performance (NBG) correlated with flow (NBG). Each part was analyzed to obtain Pearson's correlation coefficient using PSPP. The BG data analysis revealed a correlation coefficient $r = 0.87$ while the NBG data analysis revealed a correlation coefficient $r = 0.50$. Therefore, the first and second condition show a high and low correlation, respectively. This means that indeed flow is correlated with high performance but this is not the case for a low state of flow.

8. Discussion and Conclusions

In this paper, we claim that interactive systems that enforce balanced challenges foster a state of flow and improve performance. A balanced challenge is one in which the mental and motor skills that the user needs to deploy to achieve the task successfully are demanded by the system in similar proportions.

We first presented and tested a methodology to measure the mental and the motor load involved in interactive tasks, and hence to assess whether the challenges in a human–computer interaction task are balanced. Then, we developed a methodology to assess whether balanced challenges do promote a state of flow better than unbalanced ones, and to evaluate whether flow promotes performance better.

To test these hypotheses, we developed two interactive systems with tangible interfaces implementing serious games oriented to rehabilitate shoulder injuries: SIBMER and Macoli. Both systems had two versions, one implementing balanced and the other unbalanced challenges. The experiments confirmed that in both systems balanced challenges promote a better state of flow, and in the case of Macoli flow was correlated with high performance. These results suggest that balanced challenges promote flow and flow improves performance.

However, in SIBMER the unbalanced version was more complex and attractive than the balanced one, and playing one version of the game first biased playing the other. However, this contrasts strongly with Macoli in which the balanced version is more complex and attractive than the unbalanced one, and playing one version of the game does not influence playing the other. These observations taken together suggest that balanced challenges do promote flow and flow promotes better performance independently of the mental or motor load, and of the complexity and attractiveness of the system. In summary, the experimental study presented in this paper provides support to the hypothesis that HCI systems involving balanced challenges are more likely to induce a state of flow than systems with the same or similar goals that involve simple or unbalanced challenges.

These results are very positive and encouraging but need to be corroborated with further systems and experiments in different settings and application domains.

A more general approach would be to include an emotional dimension in addition to the physical and the mental, and to explore whether a balance between the three dimensions impact on flow.

The present study was also restricted to the user-experience aspect of the systems, and the application to the rehabilitation of actual patients with shoulder injuries is left for further work.

Author Contributions: Conceptualization and methodology R.C. and L.A.P.; software, validation, formal analysis, experiments and data collections, R.C.; writing—original draft preparation, writing—review and editing, R.C. and L.A.P.; supervision, L.A.P.

Funding: R.C. was funded by a CONACyT's doctoral research grant No. 218241 and CONACyT's Project 178673; L.A.P. acknowledges the support of CONACyT's Project 178673 and PAPIIT-UNAM Projects IN109816 and IN112819.

Acknowledgments: The authors gratefully acknowledge Jesús Favela, Enrique Sucar, Tom Froese, Fernando Gamboa and Pablo Romero for useful comments and suggestions. We also thank the reviewers of this and previous versions of the paper.

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

References

1. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*; Harper Perennial: New York, NY, USA, 1991.
2. Romero, P.; Calvillo-Gómez, E. An Embodied View of Flow. *Interact. Comput.* **2014**, *26*, 513–527. [[CrossRef](#)]
3. Ghani, J.A.; Deshpande, S.P. Task Characteristics and the Experience of Optimal Flow in Human–Computer Interaction. *J. Psychol.* **1994**, *128*, 381–391. [[CrossRef](#)]
4. Ghani, J.A.; Supnick, R.; Rooney, P. The Experience of Flow in Computer-mediated and in Face-to-face Groups. In Proceedings of the Twelfth International Conference on Information Systems, New York, NY, USA, 16–18 December 1991; University of Minnesota: Minneapolis, MN, USA, 1991; pp. 229–237.
5. Trevino, L.K.; Webster, J. Flow in Computer-Mediated Communication: Electronic Mail and Voice Mail Evaluation and Impacts. *Commun. Res.* **1992**, *19*, 539–573. [[CrossRef](#)]
6. Webster, J.; Trevino, L.K.; Ryan, L. The dimensionality and correlates of flow in human-computer interactions. *Comput. Hum. Behav.* **1993**, *9*, 411–426. [[CrossRef](#)]
7. Nijholt, A.; van Dijk, B.; Reidsma, D. Design of Experience and Flow in Movement-Based Interaction. In Proceedings of the Motion in Games: First International Workshop, MIG 2008, Utrecht, The Netherlands, 14–17 June 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 166–175.

8. Pasch, M.; Bianchi-Berthouze, N.; van Dijk, B.; Nijholt, A. Immersion in Movement-Based Interaction. In Proceedings of the Intelligent Technologies for Interactive Entertainment: Third International Conference, INTETAIN 2009, Amsterdam, The Netherlands, 22–24 June 2009; Springer: Berlin/Heidelberg, Germany, 2009; pp. 169–180.
9. Nijholt, A.; Pasch, M.; van Dijk, B.; Reidsma, D.; Heylen, D. Observations on Experience and Flow in Movement-Based Interaction. In *Whole Body Interaction*; Springer: London, UK, 2011; pp. 101–119.
10. Hassenzahl, M.; Tractinsky, N. User experience—A research agenda. *Behav. Inf. Technol.* **2006**, *25*, 91–97. [[CrossRef](#)]
11. Chen, J. Flow in Games (and Everything else). *Commun. ACM* **2007**, *50*, 31–34. [[CrossRef](#)]
12. Finneran, C.; Zhang, P. The challenges of studying flow within a computer-mediated environment. In Proceedings of the AMCIS 2002 Proceedings, Dallas, TX, USA, 9–11 August 2002; Volume 416.
13. Fullagar, C.; Knight, P.; Sovern, H.S. Challenge/Skill Balance, Flow, and Performance Anxiety. *Appl. Psychol.* **2013**, *62*, 236–259. [[CrossRef](#)]
14. Csikszentmihalyi, M. *The Evolving Self: A Psychology for the Third Millennium*; HarperCollins Publishers: New York, NY, USA, 1993.
15. Novak, T.P.; Hoffman, D.L.; Yung, Y.F. Measuring the Customer Experience in Online Environments: A Structural Modeling Approach. *Mark. Sci.* **2000**, *19*, 22–42. [[CrossRef](#)]
16. Chen, H.; Wigand, R.; Nilan, M. Optimal experience of Web activities. *Comput. Hum. Behav.* **1999**, *15*, 585–608. [[CrossRef](#)]
17. Nakamura, J.; Csikszentmihalyi, M. The concept of flow. In *Handbook of Positive Psychology*; Oxford University Press: Oxford, UK, 2002; pp. 89–105. [[CrossRef](#)]
18. Csikszentmihalyi, M.; Nakamura, J. *Effortless Attention in Everyday Life: A Systematic Phenomenology*; MIT Press: Cambridge, MA, USA, 2010; pp. 179–190.
19. Finneran, C.M.; Zhang, P. Flow in Computer-Mediated Environments: Promises and Challenges. *Commun. Assoc. Inf. Syst.* **2005**, *82*, 101. [[CrossRef](#)]
20. Romero, P.; Calvillo-Gómez, E. Effortless Attention and Composite Challenges in Movement Interaction. In Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers, Birmingham, UK, 12–14 September 2012; British Computer Society: Swinton, UK, 2012; pp. 157–164.
21. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Adv. Psychol.* **1988**, *52*, 139–183. [[CrossRef](#)]
22. Grier, R.A. How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, USA, 26–30 October 2015; Volume 59, pp. 1727–1731.
23. Jackson, S.A.; Marsh, H.W. Development and Validation of a Scale to Measure Optimal Experience: The Flow State Scale. *J. Sport Exerc. Psychol.* **1996**, *18*, 17–35. [[CrossRef](#)]
24. Hamari, J.; Koivisto, J. Measuring Flow in Gamification. *Comput. Hum. Behav.* **2014**, *40*, 133–143. [[CrossRef](#)]
25. Calvo, T.G.; Castuera, R.J.; Ruano, F.J.S.R.; Vaíllo, R.R.; Gimeno, E.C. Psychometric Properties of the Spanish Version of the Flow State Scale. *Span. J. Psychol.* **2008**, *11*, 660–669. [[CrossRef](#)]
26. Kalpić, D.; Hlupić, N.; Lovrić, M. Student's t-tests. In *International Encyclopedia of Statistical Science*; Lovric, M., Ed.; Springer: Berlin/Heidelberg, Germany, 2011; pp. 1559–1563.
27. GNU Project. *GNU PSPP for GNU/Linux (Version 0.8.5) [Computer Software]*; Free Software Foundation: Boston, MA, USA, 2015.
28. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611. [[CrossRef](#)]
29. Gastwirth, J.L.; Gel, Y.R.; Miao, W. The Impact of Levene's Test of Equality of Variances on Statistical Theory and Practice. *Stat. Sci.* **2009**, *24*, 343–360. [[CrossRef](#)]

