A technique for interrater reliability evaluation of a mobile game aimed for executive functions stimulation

Bernardo B. Cerqueira[a], Débora N. F. Barbosa[a], João B. Mossmann[a], Jorge L. V. Barbosa[b]
[a]Universidade Feevale, Novo Hamburgo, RS 93525-075, Brazil.
[b]Universidade do Vale do Rio dos Sinos, São Leopoldo, RS 93020-190, Brazil.
Corresponding Author – Bernardo Benites de Cerqueira
Correspondent Author: Santa Cruz Street, 191. Bairro Campestre, São Leopoldo - RS, ZIP CODE 93044-516, Brazil. Email: bernardo@feevale.br

Abstract

This work presents an interrater reliability evaluation of a mobile game aimed for the executive function's stimulation, specifically the inhibitory control. The educational Exergame "The Incredible Adventures of Apollo & Rosetta in the Space" was presented in a previous piece of work, which showed the development of the game and its application in a researching intervention with children, in the 6 to 10 age-group, in a school environment. Subsequently, the game had its code reworked for being able to be used across different platforms, hence culminating in the present work. The methodology in this paper consists in a mixed qualitative-quantitative evaluation through questionnaires with four domain professionals experienced in the executive functions field. The statistical measurement used was based on the Kappa coefficient and average percentage among the judges. As the results indicate, there was a substantial agreement (k=0,659; P-value=0,000) between the raters, as well as a high percentage of agreement in general on the mobile game's capability of Executive Functions stimulation for children.

Keywords: Executive Functions; Inhibitory Control; Game Evaluation; Cognitive Stimulation; Digital Games;

Introduction

This work presents a game evaluation process utilized for the validation of a mobile game aimed for Executive Functions (EF) stimulation through a multiple interrater reliability. The development of the game "*The Incredible Adventures of Apollo & Rosetta in the Space*" (Apollo & Rosetta) was shown in a previous work (Mossmann et al, 2017), which consisted in the development of an Exergame aimed for EF stimulation, as well as an intervention program for children from 6 to 10 years old in the elementary school environment. Thereafter, the game Apollo & Rosetta was adapted into mobile devices in order to increase portability, as demonstrated in Barbosa et al (2018). As a sequel to the game adaptation, the present work features an interrater reliability evaluation with multiple judges, professionals with experience in the EF field.

In the context of programs aimed for EF stimulation, (Diamond, 2013) states that it is not clear whether computerized programs, such as digital games, are able to stimulate or not the EF. The core contradiction is that in order to assess whether the stimulation program was effective and beneficial, the benefit from the trained skill must be evidenced to untrained skills. In other words, there must be transference to other activities, promoting adaptive functioning of the player, not just their improvements on the practiced activities. Hence, further research aiming at the use of electronic programs and games designed for EF stimulation are needed, as well as methodologies for evidence-based research in this field (Diamond & Lee, 2011; Diamond, 2013; Diamond & Ling, 2016).

Also, according to a study by Baranowski et al (2016), few existing mobile apps incorporate videogame strategies for children. Because of this, it is necessary wider collaboration between game designers, health and behavior professionals, so that evidence-based behavior changing techniques can be granted on the applications (Baranowski et al, 2016). Besides, research on the effectiveness of these games are essential, as well as finding out what are the better combination of game mechanics and behavioral change processes in order to maximize behavior changes, along with lesser possible side effects on the participating subjects, such as children. In this matter, studies in the area associates the typical development of the EF and correlates them to matters such as school success (Bull et al, 2004), and also to the fact that good EF development is required to ensure proper school development and other behaviorial aspects (Jurado & Rosselli, 2007).

In this context, this work aims for the validation of the Apollo & Rosetta mobile game for EF stimulation with professionals having experience in the EF field, meeting current demands on needed research, such as involving videogame strategies aimed for health (Baranowski et al, 2016) and evidence-based research (Diamond & Ling, 2016). The objective of the game's validation is to evaluate with the judges if the mobile version of the game may present EF stimulation capabilities in each one of the seven activities that it is made up of. In addition, the judges will assess the comprehensibility, adequacy to the age group and consistency between the activities and proposed objective of the game, which is the stimulation of the EF, and primarily the Inhibitory Control.

This paper contributes to the discussion by presenting a technique for videogames aimed for executive functions stimulation through evaluation and validation by domain professionals.

The proposed methodology is a qualitative-quantitative one, consisting of two distinct questionnaires for the judges (n=4) to analyze the features of the mobile game, judge and answer the questions. This work utilizes two different statistical calculations for the questionnaires: Kappa coefficient (Fleiss e Cohen, 1973), for overall agreement and agreement percentage. In order to perform the evaluation with the subjects, single sessions were held in loco and individually, in which the game was presented with a proposed script to the judges. The evaluation was audio-recorded for further analysis aiming to support or confront the results obtained and further discuss them. As the results show, the Kappa coefficient indicated substantial agreement between the judges; also, this study demonstrated high agreement percentages among the subjects.

This article is divided in 6 sections. Next section will describe the EF and the current knowledge on this field. In section 3, the development of Apollo & Rosetta mobile game is briefly introduced. Furthermore, section 4 will detail the methodology utilized for the interrater reliability evaluation, followed by the results, discussion and conclusions found in this work on section 5. Section 6 brings the final considerations.

Executive functions

The EF are a group of skills that guide individual's behavior towards task accomplishment, as well as inhibition of tendencies and thought regulation and control. Among these functions of the human brain, according to Diamond (2013) (Müller et al. 2008), there are three interrelated core skills, which are the basis of the other higher functions:

Working Memory (WM), responsible for managing, relating, connecting and handling any current to previous information, which is essential for reasoning, problem solving and creative thought of individuals; (Diamond, 2013)

Inhibitory Control (IC), responsible for braking and inhibiting impulses or behaviors, emotions and inappropriate or distractive thoughts, enabling self-control and nonhabitual responses according to the demands of each situation faced by the individual (*ibidem*);

Cognitive Flexibility (CF), a capacity of changing and alternating attentional focus between different tasks and adapting to the environment (*ibidem*).

In this context, studies associate the EF to the daily behavior of children with typical development, where the correlations found include questions such as school success and socio-affective functioning (Bull et al, 2004). Investigations in this context also confirms these findings and demonstrate that good development of the executive skills is required to ensure proper school development, a successful professional career and also many other daily aspects of the individual's life (Jurado & Rosselli, 2007). Still as regards the EF development, studies also have shown that the maturation of these skills occurs since early childhood (Bernier et al, 2010), on a long journey to adulthood (Conklin et al, 2007).

Currently it is also known that it is possible to help children to develop and improve their executive skills through ludic activities which stimulates reasoning, planning and IC (Diamond & Lee, 2011). Moreover, recent studies have shown that higher EF levels are related to self-control, creativity and task flexibility (Diamond, 2013), which are essential skills in many aspects of an individual's life. Such skills range from health and physiological issues, cognitive development and previously mentioned effects in one's social and professional spheres (Carlson et al, 2004; Diamond, 201; Hughes & Ensor, 2007). Hence, exercising and stimulating EF in order to improve it, could significantly increase the chances of success related to reading, writing and mathematics (Bull et al., 2008; Christopher et al., 2012; Monette et al., 2011; Toll et al., 2011; Welsh et al., 2010). The current scenario demonstrates the significance of investing in programs aimed for EF stimulation.

In the context of programs and interventions aimed for EF stimulation for children, a study by Diamond & Lee (2011) offers six different approaches for this age range. Among these approaches, computerized training programs were assessed, including digital games as a tool for cognitive stimulation and not only entertainment purposes. Although stimulation evidence was found on WM and reasoning for children, inhibitory control stimulation benefits remained unclear. Yet, according to a study by de Jong (2014) on the mostly investigated computerized WM training game CogMed®, a stimulation program which requires a certified mentor for the application, little attention on mentors roles in the final results of the training were assessed, and seems to account as well for the benefits of the training, more than computerized games. A recent study (Diamond & Ling, 2016) also indicates the current contradictions on the presented evidences regarding the effectiveness of computerized program approaches in EF training, demonstrating the necessity to invest in areas dedicated to EF stimulation. The utilization of adapted software, digital games included, for stimulation, diagnostic and rehabilitation potentials of EF is currently under study. Besides, the controversies in this field remains on the resulting stimulation of the participating subjects (Buelow et al, 2015), as well as the methodologies employed (Mansur-Alves et al, 2017) and the reproduction of the results in other studies (Holmes et al, 2009) (Diamond & Ling, 2016).

GAME ADAPTATION BETWEEN PLATFORMS – *APOLLO & ROSETTA*

This session summarizes the development process of the Exergame Apollo & Rosetta for the mobile platform, specifically Android Tablets (Barbosa et al, 2018). The educational Exergame Apollo & Rosetta was developed by Mossmann (in press) which investigated exergames as mediators for EF stimulation, specifically IC, for elementary school children in the school environment. The Exergame consisted in 7 different activities, each one aimed for a different aspect of the IC. Also, the narrative is set in a space environment, where the player has to help the characters to become a space explorer, being each activity necessary for the character to develop its skills for space exploring.

The exergame's development team included many professionals and students from areas such as computer engineering, psychology, game design (specific professional formation), programming, mathematics, pedagogy, experts in the neuropsychology field, voluntary subjects for game tests etc. In addition, a pilot study was performed in the school environment, which consisted in a 3-month program aimed for children in the age range of 6 to 10 years old. Subsequently, the game was adapted from Exergame into mobile devices, aiming to increase mobility for the application in the school environment. In this context, the necessary planning for project management in software development aimed for games is an extremely important and complex task, which requires the foresight of many aspects in its conception. Hence, the scope of the project has to be specified beforehand, as well as its duration, complexity, production agenda, cost planning, development estimates and, in the case of the industry, risk forecast and lucrative returns of the final product, in order to raise project's success chances as much as possible (Baba & Tsang 2001) (Tsang, 2005). Based on the proposed model by Baba and Tsang (2001), the game Apollo & Rosetta was adapted from an Exergame into mobile devices (Barbosa et al, 2018).

Also, based on the cyclic evolutionary model aimed for game development (Baba & Tsang 2001), the adaptation followed the technique similarly, which has its utilization and application detailed and documented in (Barbosa et al, 2018). The technique can be summarized in 5 stages, on which the development team relies on the discussion, analysis, conceptualization, definition of which changes should be made and then schedule the development, followed by evaluation with the target audience. At the end of the evaluation, the results are analyzed and the process starts again, according to the analysis by the development team, addressing issues raised by the testing. In total, the development of the adaptation project lasted 5 months. Also, it is important to emphasize that, regarding this cyclical development model, its application could keep on going indefinitely, aiming for its improvement and game testing, since the game will always be a “prototype” in this technique. However, Baba & Tschang (2001) indicates that each cycle restart also entails a gradual increase of the cost and production time of the project.

In Figure 1, the game tests with voluntary subjects are demonstrated bellow:

Figure 1 – Game tests with voluntary subjects.

The version of the game resulting from the application of the technique is described below, with details of each activity involving the game Apollo & Rosetta for mobile devices, which were further evaluated by the judges:

Activity 1 – Explorer: In this minigame, the player’s character surfs through a space tunnel towards the horizon, where the goal is to collect with the character’s hands multiple objects that appear along the route, using on-screen buttons, and to dodge obstacles. At this point, a list is found on the edge of the screen, indicating which objects are able to be collected or not, being updated randomly in each stage of the activity. In order to win this round, the player has to make the character catalog only the correct items, which are indicated on the list, and to dodge different obstacles or incorrect items. Originally intended to be played in Exergame with body movements for laterality and the player hands to reach the collectable items, the mobile version uses the device rotation for the character’s movement and onscreen buttons for the character’s hands to be used ingame.

Activity 2 – Deciphering Codes: In this minigame the player must be attentive to a panel on the screen which has 4 letters in it, each of which related to 4 buttons containing the same letters in the game’s interface. The goal of this activity is to press the indicated letters that are highlighted on the panel, each at play respectively; in the meantime, a space television voices a word occasionally. Every time the space TV spells a word, the player must pay attention to the sound stimuli and compare whether the word’s first letter matches the indicated letter on the panel or not. In case these letters match, the player must inhibit its action of pressing the panel’s indicated button and press a special button on the screen, and if they do not match, the player must ignore the sound stimuli and proceed as usual. In the Exergame it was originally intended for the player to interact with ingame buttons with their hands and feet, whereas in the mobile version the buttons were rearranged for an ergonomic interaction when the player holds the device with both hands.

Activity 3 – Particle Accelerator Tunnel: The goal of this minigame is to guide the character through a tunnel,making it dodge obstacles that appear along the routeand piloting its flyer hoverboard from left to right by inclining the device about. This activity consists of two alternating moments, which results in different device maneuvering by the player. In the first moment, the player's third person camera aims for the horizon showing the back of the character and the route's obstacles. Meanwhile, from time to time the camera inverts its position during a short time period, which consists in the second moment, resulting in a perspective that shows the character's front as the incoming obstacles are displayed in the game interface. During the inverted camera position state, the player's movements are also inverted, which means that, whenever the player inclines the device to the right, the character will move to the left, and vice-versa. In the Exergame format, the player had to interact moving his whole body to dodge the ingame obstacles, as it was replaced by device rotation in the mobile version.

Activity 4 – Jumping Asteroids: In this activity, the player has to focus on four asteroids beneath the character's feet, which corresponds to four buttons on the game interface. One pair of these asteroids is colored in each round, and the player's goal is to press the respective buttons to match the indicated move by the game. However, there are colors that must not be touched and which are indicated on a panel in the game interface; hence, the player must press the opposite pair of buttons instead. Therefore, whenever the color of the asteroids matches the colors shown on the list, the player must act on the opposite pair of buttons, aiming for the uncolored pair of asteroids in this case. In the Exergame format, the player had to jump and fall into the right spot with their feet, according to the demanded move and the asteroids, whereas in the mobile version, buttons were placed and arranged ergonomically onscreen for the user to interact with their thumbs, for the sake of better playability.

Activity 5 – Galactic Art: In this minigame, the player has to be attentive to flying colored balls, as space pipes toss them about in mid-air constantly. The goal of the activity is to hit with the player's finger the correct colored balls and to avoid black or white ones, in order to paint a canvas present in the scenario. Meanwhile, from time to time, space-flies come in to the scene and starts to mess with the canvas and make annoying noises as the player's score diminishes. Hence, the player has to decide whether to scare them away by pressing a finger over the space-fly, losing some colored balls in the process, or keep hitting the colored balls and perform that action later. In the Exergame version the player has to interact with the objects using their hands to reach the colored balls, while in the mobile version this movement was replaced by touch interaction with the device's screen.

Activity 6 – Stellar Laboratory: The goal of this minigame is to collect space elements (alien vitamins) that come through 4 colored and numbered tubes. In order to collect the vitamins, each tube has a respectively numbered button on the game interface, which the player must press to collect the desired elements. However, each vitamin has its own number and color and must be disposed of if they do not match its tube color and number. In the Exergame format, the player had to interact with the buttons on the game interface with their hands and feet simultaneously, while in the mobile version the buttons were recreated and rearranged ergonomically for the player to interact with them easily while holding the device.

Activity 7 – Opposites Challenge: In this minigame, the player has to collect different elements that appear on the scene using the characters hands or feet, commanding it through 4 buttons on the game interface respectively. The goal is to collect the elements indicated by Tivo, a computer present in the scenario which voices what side, size or specific object to collect. However, from time to time Tivo's brother, Ovit, hacks the computer and takes its brother's place for a short time period, indicating the wrong objects on the tunnel to collect, causing the player to act in opposition to Ovit's indication. In the original Exergame version, the player had to reach for the objects with their hands and feet, whereas in the mobile version this move was replaced by onscreen buttons corresponding to the character's hands and feet animation, in order to reach the objects.

Materials and methods

The proposed methodology in this paper is a mixed qualitative-quantitative one, consisting of an evaluation with two distinct questionnaires (Mossmann, in press) for domain professionals of the executive functions field, which analyzed and judged the mobile game according to the proposed questions. The set-up for the application of the evaluation was individual, single sessions, held in loco, accompanied by a researcher who followed a script for the presentation of the game and the seven activities for the judges. The following subsections describe the subject's participation in the evaluation and the analysis parameters utilized for this methodology.

In the application context, the evaluation consisted in two stages for the judges to analyze, judge and assess the activities in Apollo & Rosetta. In the first stage, the judges had to evaluate which EF was predominantly present in each one of the 7 activities of the mobile game, an evaluation stage that aims for the validation of the predominant EF which is required to obtain success in the referred activity. The judges had to choose between four numbers in this first questionnaire for each activity:

1) *Planning* – Refers to the ability to identify and organize diverse elements towards reaching a goal. It also consists in studying the model/idea and gather every material required to accomplish the task;
2) *Inhibitory Control* – Refers as much to the ability to inhibit automatic or impulsive actions (self-control) as to avert player's attention being diverted when they are confronted with distraction factors or thought processes (interference control);
3) *Working Memory* – Refers to the ability to keep information for a short time period while performing a complex activity. Also, this process enables mental manipulation of previously acquired information with current ones;
4) *Cognitive Flexibility* – Refers to the ability to change attentional focus, perspectives, priorities or rules and to helping players adapt to environment demands. It relates to the ability to take or consider different approaches to a situation or problem.

In the second stage of the evaluation, the judge's objective is to assess and answer three different questions of a questionnaire related to the game's target audience adequacy, as well as to the comprehension and coherence between the activity and the proposed objective. The questions involved are described below:

a) Q1 - The activity is adequate for children in Elementary School (6 to 10 years old).
b) Q2 - The description and instructions are comprehensive and clear.
c) Q3 - There is coherence between the activity and the proposed objective.

In order to judge the questions in this second stage, the judges had to give their answers using the Likert Scale (Likert, 1932) from 0 to 5 for each question regarding the activity (0 – Totally disagree; 1 – Disagree; 2 – Partially disagree; 3 – Partially agree; 4 – Agree; 5 – Totally agree;). Replying to a questionnaire based on this scale the participants specifies their agreement level to each question.

Afterwards, the obtained data with these two questionnaires will be crossed through statistical measurement to rate the agreement reliability between the judges utilizing the Kappa coefficient (Fleiss e Cohen, 1973), as this is the most used coefficient for data classification in nominee categories (Chen & Krauss, 2004). In situations involving more than two judges for the agreement coefficient, different approaches can be adopted, related to the participant's n size (Posner, 1990). In case the agreement rate is considered satisfactory related to the evaluated questions, this could indicate the validation between these domain professionals. The agreement results of the evaluation among the judges could also indicate issues showing the necessity for further revision in the game.

Participants

In total, the evaluation relied on four participants (n=4), and the inclusion criteria for the judges was at least two years of experience in the EF field. The evaluation was performed individually by a researcher with each judge, at a location and time viable for the judge to participate. Each session had between 1 hour and 30 minutes to 2 hours long for the evaluation to be completed, in which the judge had to answer an identification field and then follow the proposed script for the game to be presented in an Android® tablet by the researcher. Following comes the description of the participant's experience and area:

Judge A – has 5 years of experience in EF field (children and adolescent neuropsychological evaluation);

Judge B – has 9 years of experience in EF field (children and elderly neuropsychological evaluation);

Judge C – has 2 years of experience in EF field (children neuropsychological evaluation);

Judge D – has 9 years of experience in EF field (neuropsychological evaluation);

Analysis parameters for the evaluation

In order to evaluate the overall agreement among the 4 judges, the Kappa coefficient (Fleiss e Cohen, 1973) was calculated to obtain overall agreement among the subjects, although it was not possible to perform the calculation separately in the second stage due to low sample size and low response variability.

For Kappa coefficient interpretation, it was utilized the scale proposed by Landis and Koch (1977), according to the description in Table 1.

Table 1 – Kappa coefficient scale interpretation (Landis & Koch, 1977)

Kappa	**Agreement level**
< 0,00	Inexistent agreement
0,00 - 0,20	Minimum agreement
0,21 - 0,40	Reasonable agreement
0,41 - 0,60	Moderate agreement

0,61 - 0,80	Substantial agreement
0,81 - 1,00	Perfect agreement

For the data analysis it was utilized R (version 3.4.4) (R Core Team, 2015).

Also, in order to measure the agreement among the 4 judges in general and in each stage, the average agreement percentage between the judges was calculated from the simple ratio on the number of times the judges agreed exactly.

By way of example, supposing that 3 judges classified 5 activities, the analysis will verify the data as follows:

Judge A – 2, 2, 2, 2, 2

Judge B – 3, 2, 2, 2 ,2

Judge C – 2, 3, 3, 2, 2

To calculate the agreement percentage among these raters, the number of times they agreed is computed and the resulting number is divided by the number of total evaluations:

Percentage of agreement between Judges A and B = 4/5 = 80%;
Percentage of agreement between Judges A and C = 3/5 = 60%;
Percentage of agreement between Judges B and C = 2/5 = 40%;

The average agreement percentage between these example judges is the average ratio among them: (80 + 60 + 40)/3 = 60%.

Interrater reliability results and discussion

As the objective set previously was "to evaluate with judges having experience on the EF field if the mobile version of the game may present EF stimulation capabilities in the activities", the overall Kappa coefficient resulted in 0,659 (p-value=0,000), indicating a substantial agreement between raters according to the parameters of the Kappa coefficient scale shown in Table 1. Moreover, Table 2 presents the general agreement percentage results between judges. It can be verified that the average agreement percentage of the four judges was 84.4% [74.3%; 94.5%], implying a high percentage of agreement among the raters. The coefficient and average percentage refers to the predominant EF in the activities, as well as the previously stated questions in section 4.

Table 2 – General Agreement Analysis

Judge	0	1	2	3	4	5	Agreement percentage
Judge A	0	0	7	0	0	24	84,4% [74,3%; 94,5%][1]
Judge B	0	1	7	4	3	16	
Judge C	0	0	6	0	2	23	
Judge D	0	0	7	0	0	24	

[1] Confidence interval.

As seen in Table 2 and according to previously mentioned methodology, it can be noticed that higher numbers appear in the categories "2" and "5". In Table 2, Stage 1 and 2 of the evaluation were mixed in the categories, as Stage 1 numbers represent predominant EF: Categories "1 – Planning; 2 – Inhibitory Control; 3 – Working Memory; 4 – Cognitive Flexibility"; and Stage 2 numbers represent the rating according to the questionnaire presented in section 4: Categories "0 – Totally Disagree; 1 – Disagree; 2 – Partially disagree; 3 – Partially Agree; 4 – Agree; 5 – Totally Agree". The agreement percentage is calculated according to the number of times the judges agreed, considering that the total numbers of ratings possible in Stage 1 is 7, while in Stage 2 it is 24. In this perspective, Judge A and D agreed totally regarding which predominant EF stimulation capability can be seen on the activities as category "2 – "Inhibitory Control", consisting in 7 ratings for each judge. On the other hand, e.g., the same judges also completely agreed on category 5 of the Stage 2: "5 – "Totally Agree", therefore totalizing 7 ratings in category 2 and 24 in category 5.

In addition to the general results, an agreement percentage was utilized for analyzing the results of the assessed questions individually, which are shown in Table 3. Due to low sample size and low response variability, it was not possible to compute the Kappa value in this individual analysis. Table 3 also represents the same pattern explained as regards Table 2, but the data show the judge's ratings per question assessed; therefore, the agreement results are presented separately, question by question.

Table 3 – Agreement analysis by stage

Stage (Question)	Judge	0	1	2	3	4	5	Agreement percentage
Stage 01	Judge A	0	0	7	0	0	0	92,9% [86,6%; 99,1%][1]
	Judge B	0	0	7	0	0	0	
	Judge C	0	0	6	0	1	0	
	Judge D	0	0	7	0	0	0	
Stage 02 (Q01)	Judge A	0	0	0	0	0	8	93,8% [88,3%; 99,2%][1]
	Judge B	0	1	0	0	0	7	
	Judge C	0	0	0	0	0	8	
	Judge D	0	0	0	0	0	8	
Stage 02 (Q02)	Judge A	0	0	0	0	0	8	58,3% [28,9%; 87,8%][1]
	Judge B	0	0	0	4	2	2	
	Judge C	0	0	0	0	1	7	
	Judge D	0	0	0	0	0	8	
Stage 02 (Q03)	Judge A	0	0	0	0	0	8	93,8% [88,3%; 99,2%][1]
	Judge B	0	0	0	0	1	7	
	Judge C	0	0	0	0	0	8	
	Judge D	0	0	0	0	0	8	

[1] Confidence interval.

As Table 3 shows, agreement regarding the Stage 1 – "Predominant EF component of each activity" was 92,9%, on a confidence interval of 86,6% - 99,1% among raters. Therefore, out of the 7 activities judged, a high agreement among the judges indicated that the predominant EF component was the Inhibitory Control (2). In this context, Judge C answered that the Activity 3 "Particle Accelerator Tunnel" had a predominant Cognitive Flexibility over the Inhibitory Control, caused by the variable environment of the "reverse movements" mechanics and dodging the obstacles, explained previously in section 3. As for the other judges, Judge B stated that "sometimes, it is difficult to think of inhibitory control without the cognitive flexibility due to its interrelation regarding the cognitive functioning." In addition to this, the judge also stated that, "taking in consideration that the objective of the game is to stimulate the laterality of the player, it makes sense to interpret that it relates to the inhibition of the action over the flexibilization". On the other hand, also in this context, Judge D stated that "inhibitory control and attention are mostly demanded in this activity".

Regarding agreement percentage on Stage 2 (Q01), Table 3 shows that most raters fully agreed, except for Judge B. For this judge, the age range (6 to 10 years old) does not seem fit, taking in consideration that, in order to play the activity "Deciphering Codes", its user depends on alphabetization knowledge for reaching success in the activity. The judge stated that, in some cases, children are not subjected to enough time of school learning so as to be fully alphabetized, which is the case of younger children (6 years old) in vulnerable social environments; therefore, it may not be possible for some players in this age range to be fully able to play this activity.

Regarding the Stage 2 (Q02) on Table 3, it can be verified that the agreement percentage is the lower among the other results. This is due to higher disagreement according to Judge B evaluation; for this subject, the tutorial's model present in the game to explain the rules and goals of the activities lacked "modelling", which means a tutorial structure in which the player plays the activity while learning it, instead of going through a narrative and demonstrative tutorial. Also, for this judge, the tutorial buttons should be heavily highlighted to grasp children's attention, and due to these observations, the judge's rating decreased overall agreement in this question as the tutorials could be improved for better understanding. At the same time, in Stage 2 (Q03), according to Table 3, every judge agreed or totally agreed that the activity and the objective of the game were coherent. Figure 2 qualitatively illustrates the agreement threshold between the judges in each stage of the evaluation, which were explained hereby in this analysis. Each colored line in Figure 2 depicts the response from 0 to 5 (Rating) of the judges in each question during the evaluation stages (Question).

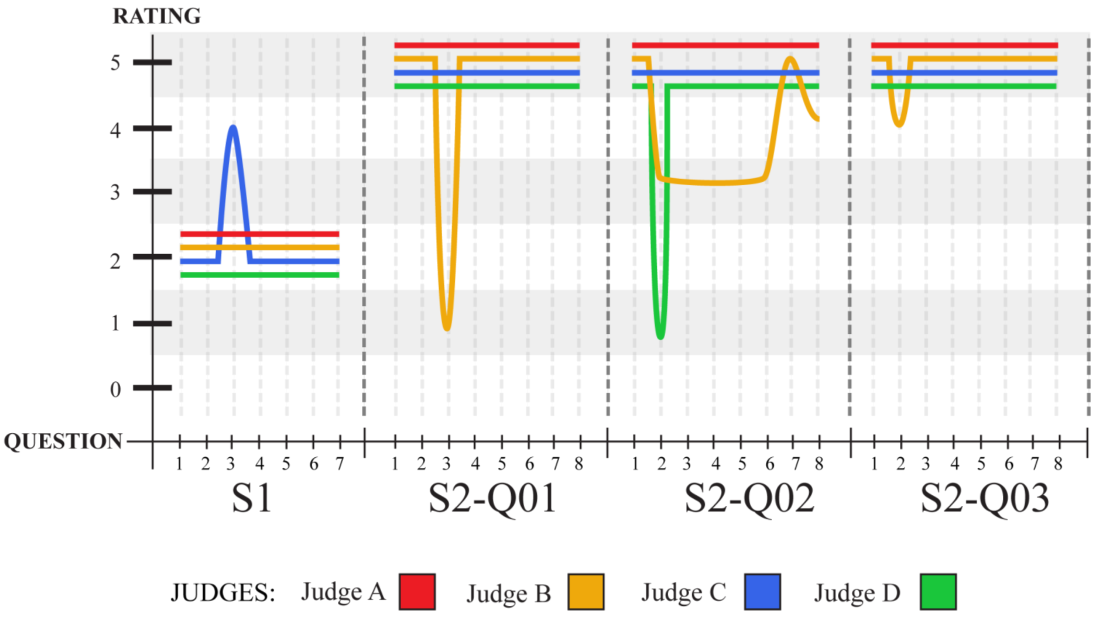

Figure 2 – Qualitative illustration of the agreement threshold between judges

Figure 3 quantitatively represents the judges' evaluation according to the average percentage threshold in each stage. The red lines represent the confidence interval between judges' response in each question assessed by the evaluation.

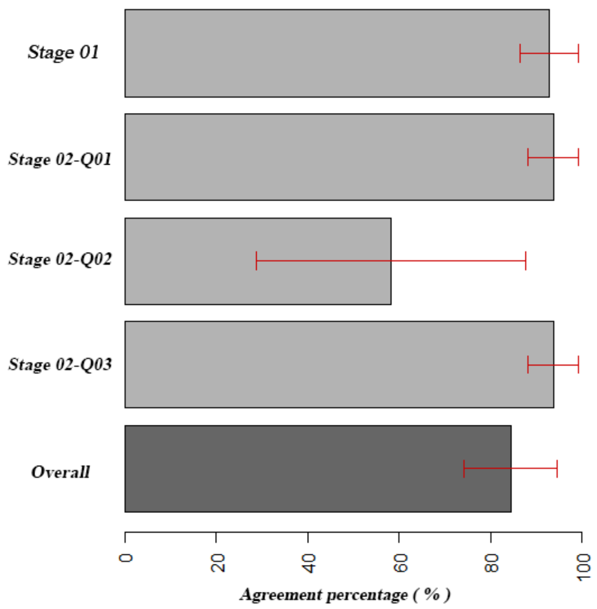

Figure 3 – Agreement among the judges

Overall, the assessment with the specialist judges in neuropsychology indicated that the activities have the IC as the main executive function requested thorough the game levels, with an average percentage of 92,9% [86,6%; 99,1%]. Moreover, the evaluation also indicated that the activities in Apollo & Rosetta mobile are adequate to their target-audience, children from Elementary School, with an average percentage of 93,8% [88,3%; 99,2%]. Besides that, it indicated that the current activities' tutorials could interfere within a wider game application aimed for cognitive stimulation; thus, it should have its expositive structure revised or reworked in order to improve the player's learning with a more interactive approach, showing an average percentage of 58,3% [28,9%; 87,8%]. Finally, it was also indicated that there is coherence between the activity and the proposed objective of the game in its mobile version, with an average percentage of 93,8% [88,3%; 99,2%]. These ratings were depicted in Figure 2 and Figure 3, which illustrated the results shown in Table 2 and 3.

Finally, in order to evaluate overall agreement among the 4 judges, the Kappa coefficient (Fleiss e Cohen, 1973) was calculated, although it was not possible to perform the calculation separately in the second stage due to low sample size and low response variability. For Kappa coefficient interpretation, it was utilized the scale proposed by Landis and Koch (1977), according to the description in Table 1. The coefficient indicated k=0,659 (P value=0,000), which reveals a substantial agreement among raters. Additionally, the average percentage of agreement shown previously, in Table 2, indicated that the judges showed an overall percentual agreement around 84,4% [74,3%; 94,5%] of the evaluations, demonstrating a high percentage of agreement regarding the previously stated questions.

Final considerations

The present work proposed and applied a combination of techniques aimed for the evaluation of a mobile game geared towards generating executive functions stimulation. In order to do that, it was performed a single session evaluation with four judges. The inclusion criteria were at least 2 years of experience on the EF field for the subjects. As the results show, the statistical Kappa overall index (k=0,659 – p value=0,000) among the four judges indicated a substantial interrater agreement on questions related to the predominant EF on each activity: comprehension, adequacy to the target audience and coherence between the activities with the objectives. Besides, an average percentage of the evaluation between the judges on each question indicated general agreement of 84.4% [74.3%; 94.5%]. Hence, the Apollo & Rosetta mobile game shows substantial agreement on the predominant EF component being stimulated by the 7 activities present in the game, which is the IC.

It should be highlighted that among the topics covered in the questionnaires, the lowest agreement rate (58.3%) among the judges was given in the question regarding the clarity of the instructions and descriptions of the activities, in which one of the judges raised a possible need of greater "modelling" in the presentation of task instructions for younger children in Elementary School I. This point can be further evaluated in an assessment with children from the target audience for further contribution to this topic, which could indicate different approaches to the modifications in the activities' tutorials.

As regards future researching work, this game is scheduled for a pilot study in late 2019 on a random controlled trial with a few children from Elementary School. This future investigation is programmed for a 25 session with the subjects in the school environment, supported by pre and post neuropsychological tests aimed for sustaining the point, or disagreeing over it, whether this version of Apollo & Rosetta for mobile games is able to stimulate EF, thus contributing to the current discussion in the field. Additionally, future studies on the clinical applicability of this game in different perspectives, such as neurodevelopmental disorders, could also contribute and would be helpful to mental health and to the development of education methodologies for children.

Acknowledgements

We thank the National Council for Scientific and Technological Development – CNPq/Brazil (http://www.cnpq.br) and National Council for the Improvement of Higher Education – CAPES (http://www.capes.gov.br/) for providing financial support for this study. Finally, we would like to thank Feevale University for embracing this research.

References

Baba, Y. & Tschang, F., Product development in japanese tv game software: The case of an innovative game. International Journal of Innovation Management, 05(04), 2001, 487–515. DOI: 10.1142/S1363919601000464

BARBOSA, D. et al, 2018. Adaptation of an educational Exergame to mobile platforms: A development process. Communications in Computer and Information Science (PRINT), v. 1, p. 287-298. DOI: 10.1007/978-3-319-95522-3_24

Bernier, A. et al., From external regulation to self-regulation: Early parenting precursors of young children's executive functioning. Child development, 81(1), 2010, 326-339. DOI: 10.1111/j.1467-8624.2009.01397.x

Buelow, M.; et al., A. The influence of video games on executive functions in college students. Computers in Human Behavior, 45, 2015, 228-234. DOI: 10.1016/j.chb.2014.12.029

Bull, R. et al., A comparison of performance on the Towers of London and Hanoi in young children. Journal of Child Psychology and Psychiatry, 45(4), 2004, 743-754. DOI: 10.1111/j.1469-7610.2004.00268.x

Bull, R. et al., Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. Developmental neuropsychology, 33(3), 2008, 205-228. DOI: 10.1080/87565640801982312

Carlson, S. et al., Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. Journal of experimental child psychology, 87(4), 2004, 299-319. DOI: 10.1016/j.jecp.2004.01.002

Chen, P., Krauss, A., Interrater agreement. The sage encyclopedia of social science research methods, 2, 2004, 511-513. DOI: 10.4135/9781412950589.n444

Christopher, M. et al., Predicting word reading and comprehension with executive function and speed measures across development: A latent variable analysis. Journal of Experimental Psychology: General, 141(3), 2012, 470. DOI: 10.1037/a0027375

Conklin, H. et al., Working memory performance in typically developing children and adolescents: Behavioral evidence of protracted frontal lobe development. Developmental neuropsychology, 31(1), 2007, 103-128. DOI: 10.1080/87565640709336889

de Jong., Effects of training working memory in adolescents with a below average IQ. Workshop on Enhancing Executive Functions in Education in Nijmegen, May 20, 2014 Nijmegen, NL.

Diamond, A., & Lee, K., Interventions shown to aid Executive Function development in children 4 to 12 years old. Science, 333(6045), 2011, 959-964. DOI: 10.1126/science.1204529

Diamond, A., Executive functions. Annual review of psychology, 64. 2013, 135-168. DOI: 10.1146/annurev-psych-113011-143750

Diamond, A; Ling, D., Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. Developmental cognitive neuroscience, 18. 2016, 34-48. DOI: 10.1016/j.dcn.2015.11.005

Fleiss, J. L.; Cohen, J.M., The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. Educational and psychological measurement, 33(3), 1973, 613-619. DOI: 10.1177/001316447303300309

Holmes, J.; et al., Adaptive training leads to sustained enhancement of poor working memory in children. Developmental science, 12(4), 2009, 1-7. DOI: 10.1111/j.1467-7687.2009.00848.x

Hughes, C. & Ensor, R., Executive function and theory of mind: Predictive relations from ages 2 to 4. Developmental psychology, 43(6), 2007, 1447. DOI: 10.1037/0012-1649.43.6.1447

Baranowski et al., Games for health for children—Current status and needed research. Games for Health Journal, 5(1), 2016, 1-12. DOI: 10.1089/g4h.2015.0026

Jurado, M., Rosselli, M., The elusive nature of executive functions: a review of our current understanding. Neuropsychology review, 17(3), 2007, 213-233. DOI: 10.1007/s11065-007-9040-z

Landis, J.R., Koch, G.G., The measurement of observer agreement for categorical data. Biometrics, 2000, 159–174.

Likert, R., A technique for the measurement of attitudes. Archives of Psychology. New York: Columbia University Press, 1932.

Mansur-Alves, M; Saldanha-Silva, R., Does Working Memory Training Promote Changes in Fluid Intelligence? Trends in Psychology, 25 (2), 2017,787-807. DOI: 10.9788/TP2017.2-19En

Monette, S. et al. The role of the executive functions in school achievement at the end of Grade 1. Journal of experimental child psychology, 109(2), 2011, 158-173. DOI: 10.1016/j.jecp.2011.01.008

Mossmann, J. B. et al, 2017. Evaluation of the Usability and Playability of an Exergame for Executive Functions Stimulation and Its Development Process. Lecture Notes in Computer Science, v. 10275, p. 164. DOI: 10.1007/978-3-319-58472-0_14

Mossmann, J., in press. Exergames Como Mediadores Da Estimulação De Componentes Das Funções Executivas Em Crianças Do Ensino Fundamental I. Tese de Doutorado em Informática na Educação. Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Informática na Educação – PPGIE, UFRGS, Brasil.

Müller, U. et al., The effect of labeling on preschool children's performance in the Dimensional Change Card Sort Task. Cognitive Development, 23(3), 2008, 395-408. DOI: 10.1016/j.cogdev.2008.06.001

Posner, K. et al., Measuring interrater reliability among multiple raters: an example of methods for nominal data. Statistics in medicine, 9(9), 1990, 1103-1115.

R Core Team., R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2018, URL <https://www.R-project.org/>.

Toll, S. et al., Executive functions as predictors of math learning disabilities. Journal of Learning Disabilities, 44(6), 2011, 521-532. DOI: 10.1177/0022219410387302

Tschang, T., Videogames as interactive experiential products and their manner of development. International Journal of Innovation Management, 9(1), 2005, 103-131. DOI: 10.1142/S1363919605001198

Welsh et al., The development of cognitive skills and gains in academic school readiness for children from low-income families. Journal of Educational Psychology, 102(1), 2010, 43-53. DOI: 10.1037/a0016738

Copyright Disclaimer