Evaluation of the Use of Leaderboards in a Gamified Application for Academic Learning

Josinaldo Dias Macena[1]
Aislânia Alves de Araújo[2], **Jucelio Soares dos Santos**[3]

[1]Universidade Estadual da Paraíba (UEPB)

[2]Faculdade de Filosofia, Ciências e Letras de Cajazeiras (FAFIC)

[3]Universidade Federal de Campina Grande (UFCG)

`{josinaldodiassb, aislanialves}@gmail.com,`

`jucelio@copin.ufcg.edu.br`

***Abstract.** In this paper, we analyze the interference of the leaderboards in the number of correct answers in the problem-solving activities present in a gamified application. Also, we checked the usability level of the application with/without leaderboards. In order to achieve these objectives, we applied a study in a test scenario and verified the effect of the leaderboards on the number of correct answers. Then, we applied a survey with the participants to analyze the usability level of the application. As a result, although the level of satisfaction of the participants who used the application with the leaderboards was higher, this difference was not significant. Also, the leaderboards did not influence the number of correct answers to problem-solving.*
***Keywords:** gamification; leaderboards; usability.*

***Resumo.** Neste artigo, analisamos a interferência dos leaderboards no número de acertos nas atividades de resolução de problemas presentes em um aplicativo gamificado. Além disso, verificamos o nível de usabilidade do aplicativo com/sem leaderboards. Para atingir esses objetivos, aplicamos um estudo em um cenário de teste e verificamos o efeito dos leaderboards sobre o número de acertos. Em seguida, aplicamos uma pesquisa com os participantes para analisar o nível de usabilidade do aplicativo. Desta forma, embora o nível de satisfação dos participantes que utilizaram a aplicação com leaderboards tenha sido superior, esta diferença não foi significativa. Além disso, os leaderboards não influenciaram o número de respostas corretas para a resolução de problemas.*
***Palavras-chave:** gamificação; leaderboards; usabilidade.*

1. Introduction

One of the significant challenges in higher education is to attract student attention in the classroom. This scenario brings difficulties, but also the possibility of innovating in the way of teaching and interacting with students. This new time requires a connection between the teacher and the student, creating new environments that help not only in learning but also in sharing/building new possibilities for knowledge acquisition [Borghetti et al. 2017].

Improving the way of implementing new teaching practices and strategies and looking for new teaching means comes from the need to instill curiosity and increase students' interest. Thus, the teaching-learning process benefits from the use of technologies in the classroom as the student improves his skills, adjusting the curricular component to his needs, opening the possibilities for the teacher to guide/develop/expand the students' skills gradually [Soares et al. 2015].

Given this scenario, we assume that the use of playful strategies, such as games, can motivate and encourage specific behaviors. For example, changes in the behavior of individuals so that they can quickly develop their activities in the face of the situations they are in, to support the academic learning process, minimizing low performance, failures and even evasion. Within the context of playful strategies is gamification that uses game devices not only for fun but for learning and interaction. Gamification uses elements and techniques of games in scenarios not related to games, to solve problems, or encourage engagement for a pre-selected audience [Vianna et al. 2013].

In gamification, mechanics are used to keeping the individual motivated; one of these mechanics is the use of leaderboards, which is formed by a ranking system. That is, its purpose is to induce the subject to strive in a specific task. Because it is through the scores that he can observe his performance through punctuated tables—but always using methods of encouragement so that the student does not feel inferior in the middle of the competition [Aires and Leite 2017].

The leaderboards are a component of a game that consists of a screen that shows people in order and agreement with certain fundamentals. Leaderboards are one of the most used elements in gamification [Hamari et al. 2014]. In research developed by [Mekler et al. 2013] found that the use of the scoring technique, levels and leaderboards is essential for players to have good progress in the games. However, leaderboards have the power to keep players more motivated than scoring.

Based on this principle, there is a lack of gamified platforms that provide the basis for learning any academic discipline [Pereira 2018]. Due to this lack, we identified through this research the need to address appropriate gamified tools for academic learning. Thus, we developed the "Camaleão", which is a gamified environment with features that draw attention in the practice of exercises. The environment aims to enhance student learning by automatically providing feedback to students and teachers. The creation of personalized questions by the teacher is allowed, and students will have access to a daily, weekly and general classification that will allow motivation in a competitive way among them, leaving the teaching motivating.

In this paper, we analyze the interference of the leaderboards in the number of correct answers in the problem-solving activities present in the application "Camaleão". Also, we checked the usability level of the application "Camaleão" with/without leaderboards. We want to answer the following research questions: **(RQ1)** Did the participants in the experimental group who interfered with the use of Leaderboards through the application "Camaleão" achieve better performance in the number of correct answers when compared to the control group? **(RQ2)** Is the usability level of the "Camaleão" application higher when compared to its version without Leaderboards?

We organized this paper as follows. In Section 2, we present the methodology

applied to conduct in this paper. In section 3, we offer the results and their discussions. Finally, in Section 4, we present the conclusions and suggestions for future work.

2. Method

In this section, we present the platform "Camaleão", individually the study participants, the procedures that are adopted, and how the data is analyzed.

2.1. Camaleão

The user has access to the application through a browser present on a computer, tablet, or smartphone. Through the login screen, the user can register in the application, or if he/she has access, enter through the email and password registered in the registration form.

The "Camaleão" presents three users. The administrator approves the registration request and responds to the contact request. The user/teacher edits information, creates, manages, and configures the room, adds a question, and visualizes the ranking (student performance). Student edits information, manages rooms, answers questions, and visualizes ranking.

Figure 1 shows the screen when the user has access to the application. This figure shows the classes in which the user participates, as well as the classes in which he administers. That is, any user can create their class, regardless of whether they are students or teachers. The difference is that when wanting to participate in a class, it is necessary to put the access code. For the teacher (class administrator), the application generates the code automatically. Through this code, the teacher can invite students to join the class. When the student enters this code, he will automatically participate in the class and have access to all of his available activities, along with the rankings of that class.
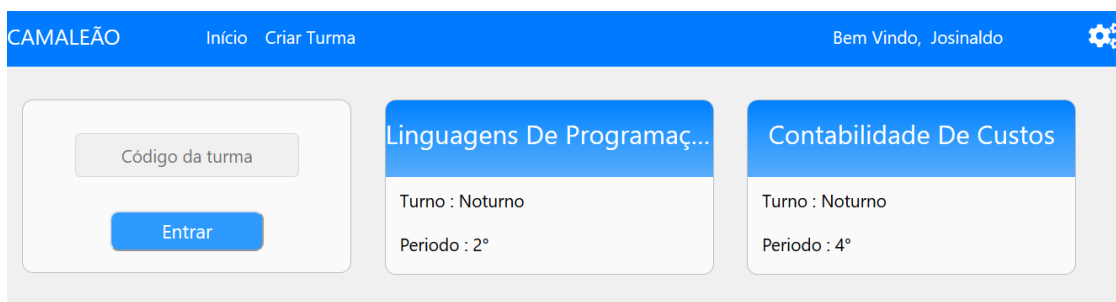

Figure 1. User Home

Figure 2(a) represents the view of the teacher with his class, where he has access to the participation code, as well as the option to edit the name, shift, and period of the class. Also, the teacher has the option of thoroughly listing students with names and emails who are currently part of their class. The teacher also has access to the visualization of the rankings with the complete list of all students and their scores in the daily, weekly and general rankings, together with all the activities created by him, as well as the option to create new activities, edit and view the old activities. Figure 2(b) represents the student's view of the class he participates in, where he only has access to the visualization of the rankings and all the activities developed by the teacher, and each activity has the "reply" button to start the activity response if the teacher has made it available.

(a) Class in the teacher's view

(b) Class in the student's view

Figure 2. Class in the user's view

Figure 3 represents the creation of new activity by the teacher, where automatically after its creation, it will be available for editing, viewing, and students' response if it is configured as available.
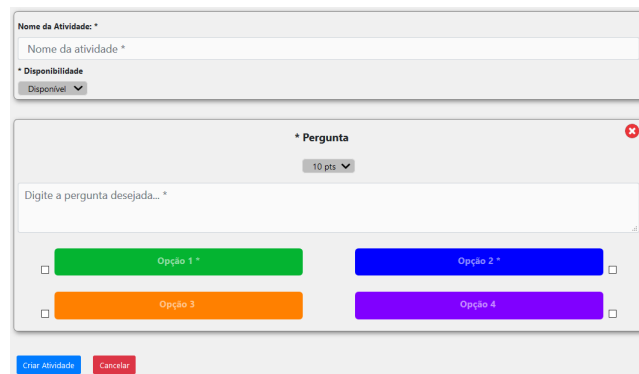
Figure 3. New activity creation page

Figures 4(a) and 4(b) show the behavior of the application when the user hits or misses a specific question. On the hit, the application displays the message "Congratulations," along with the total points earned. In error, the application presents the case in which the participant answers the question incorrectly, automatically receiving feedback with the correct answer where it will not be punished with a loss of points, this being a new incentive factor so that the student does not come discouraged before his colleagues.

(a) Question screen answered correctly

(b) Question screen answered incorrectly

Figure 4. Behavior of the "Camaleão" in a question answered by the user

Figure 5 shows how the rankings are represented on the screen, showing the three types of rankings, daily, weekly, and general. The application calculates the points according to the values added for each question, that is, if the student gets the question right he gets the points assigned to it, and if he does not get it right he does not get the score, but neither is he punished with losing points when answering each question, the points earned by the student in the activity are added. As the application displays an incentive for those who remain in the last positions, not all colleagues who are in front of them, for this, we have implemented a function that hides a certain number of users, both above and below, showing only the first five in all situations. Two more positions above the student's current position and two positions below it, the application hides the rest of the ranking participants.


Figure 5. Rankings page

2.2. Participants

We applied the "Camaleão" in the classroom to a class of 34 students linked to the curricular component "Cost Accounting". This component is part of the Administration course at the State University of Paraíba, Campus VII, Patos, Paraíba, Brazil. We categorize the choice of this class as available, that is, proximity to the teacher of this component, but it could be applied to any other context according to the experimental conditions. Everyone involved in the research signed a deliberative term to volunteer for the study.

We divided the participants into two groups (experimental and control). Both groups solved the same questions and had access to all features of the tool, except that the experimental group had interference in accessing the leaderboards. For better management, we divide the two groups at different times. For a better division of the groups and to prevent skilled subjects from interfering in the results, we separated the individuals based on the performance in the discipline. To verify that there was no difference in this division, we plotted the ggplot in Figure 6 in order to visually observe the performance of the two groups. Although the standard deviation is slightly different, the average performance of the two groups is equivalent.

It is interesting to verify that the two groups are homogeneous to their skills based on their respective grades. For this, we use the test to verify that the groups are equal. As the data do not follow a normal distribution, we used the Wilcoxon test (Tab. 1) for independent samples with 95 % significance. So when performing the test, if the p-value is less than 0.05, then we can refute the null hypothesis.
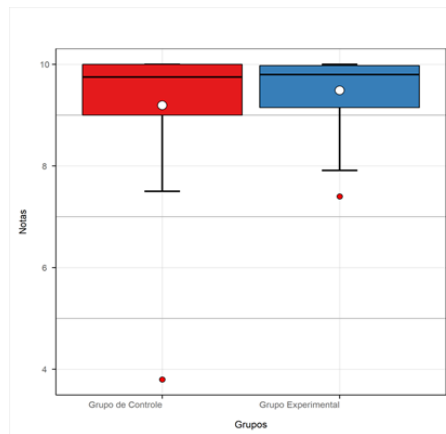
Figure 6. Boxplot of group performance in the curricular component

Table 1. Performance between groups in the Curricular Component

Null hypothesis	Control Group		Experimental Group		P-value
	Average	**Standard Deviation**	**Average**	**Standard Deviation**	
There is significance in the difference between groups	9.19	1.52	9.49	0.72	0.0419

According to the Wilcoxon test result, we accept the null hypothesis, indicating that both groups are equal, that is, homogeneous to the investigated skill.

2.3. Procedures

After dividing the groups, we performed the study only once for each group. All participants had to answer up to ten multiple-choice questions in up to thirty minutes. We collected the number of correct answers from both groups; then, all participants assessed the usability of the application through a Survey.

For the Survey, we used the usability technique through the metrics based on the USE model by [Davis 1989] and [Lund 2001]: Utility, Ease of use, Ease of learning, and Satisfaction. We adopted the Likert scale, in which students answered a total of sixteen questions, giving them value on a 5-point scale, where: 1 = Not appropriate, 2 = Not very appropriate, 3 = Moderately appropriate, 4 = Very appropriate, 5 = Completely appropriate.

2.4. Data analysis

Based on the previous steps, we separate and analyze the data obtained in the performance search for each analyzed scenario and compare them. Based on the results obtained, we contest or accept the null hypothesis proposed in the paper.

We represent the data obtained in the Survey using descriptive statistical analysis in its interpretation, and we apply the proportion test to verify if the participants who used the application "Camaleão" and without leaderboards evaluate whether its usability obeys the questioned characteristics.

The proportion test is an estimate of data analysis involving only two categories, such as success or failure. With that, we compared the proportion of students with the hypothetical proportion specified in the user evaluation questionnaire. We performed the proportion test using the Action Stat version 3.6 tool, which uses the proportion of a single sample to generate data with an indication of success. We transformed the data that were on a Likert scale into dichotomous. That is, we assigned 0 to indicators below the value 4, and we assigned 1 to indicators 4 or 5, disagreeing or agreeing that the tool meets the quality requirement.

We feed the data into the application and use a proportion test with reasonable approximation and with continuity correction with a significance level of 95%, null hypothesis of 50% and an alternative hypothesis "greater than", we must accept that 50% or more of students rated the "Camaleão" as meeting usability and quality requirements. In the results obtained, if the p-value is more significant than 0.05 of relevance, it means that the null hypothesis cannot be denied, that is, there is a significant difference and that it was successful. Thus, we analyze the modalities covered in the Survey.

3. Results and discussions

In this section, we present and discuss the results achieved by applying the studies.

3.1. Interference level in using leaderboards

We plot in Figure 7 the graph that represents the performance (number of correct answers) of both groups during the experiment.

Based on the analysis of the graph in Figure 7, we noticed that the control group got more questions right when compared to the experimental group, which we did not expect according to the literature.

As the data do not follow a normal distribution, we verified the significance in this difference using the Wilcoxon test for independent samples with 95% significance and with the following null hypothesis: the experimental group answered fewer questions than the control group. That is if the leaderboards present on the instrument can positively impact the students' successes. Table 2 shows the result of this test.
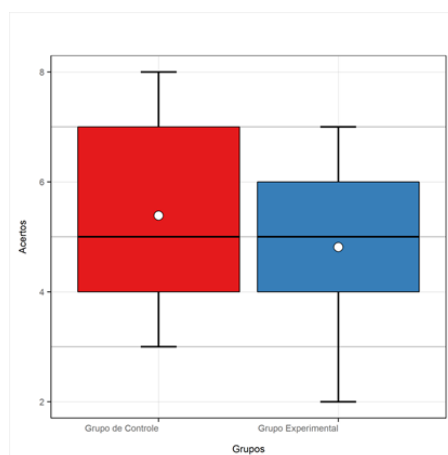

Figure 7. Boxplot performance between groups in Experiment

Table 2. Performance between groups in Experiment

Null hypothesis	Control Group		Experimental Group		P-value
	Average	**Standard Deviation**	**Average**	**Standard Deviation**	
The experimental group hit less than the control group	5.39	1.54	4.8	1.52	0.1799

According to the Wilcoxon test, we were unable to refute the null hypothesis, giving strong indications that the leaderboards do not influence the number of questions settled during the experiment. Even if the control group got more questions right than the experimental group, this difference is not significant. Regarding the involvement and engagement in the activity, both groups obtained active participation, showing interest in using the gamified platform, being captivated by the simplicity and effectiveness of the "Camaleão". Next, we present the level of usability of the application.

3.2. Application usability level

In this subsection, we present/discuss the results obtained with the application of a Research focused on the usability of the "Camaleão" application with/without leaderboards. We present the results in Table 3.

In Usefulness, the version of the "Camaleão" with leaderboards is more useful and practical when compared to its version without leaderboards. However, this difference is not significant; that is, we do not have strong indications that the participants had a particular preference for the use of the leaderboards. Regarding functionality, the participants evaluated the version of the "Camaleão" better with leaderboards, stimulating more than its version without leaderboards; however, there was no difference in its significance. About time, we also did not find a significant difference between the versions of the "Camaleão." The participants of the group that used the application without leaderboards got a feeling that they took longer to answer the questions; perhaps the entertainment present in the leaderboards tries to show that the participant's experience in answering questions with competition criteria takes less time.

In Ease of use, participants assessed that there is not much difference in the use of the versions of "Camaleão", but felt more motivated to use their version with leaderboards more successfully.

In Ease of learning, there is no significant difference in the Ease of learning between the versions of "Camaleão". The version with leaderboards proved to be more to use and remember when compared to its version without leaderboards. However, we do not rule out the hypothesis that the leaderboards may have engaged the participants more in these criteria. While, the criterion of quick ability to use the application, the participants who used the version without leaderboards demonstrated a better ability, even if it is not significant.

Table 3. Analysis of results on the opinion of professionals in relation to the USE framework

USE framework	"Camaleão" with leaderboards Success %	"Camaleão" without leaderboards Success %	P-value
Usefulness	**95.5%**	**87.5%**	-
Is "Camaleão" useful?	100%	89%	0.08
Is "Camaleão" effective?	94%	89%	0.30
Does "Camaleão" meet my needs in measuring introductory skills related to the discipline?	94%	89%	0.30
Does "Camaleão" save time?	94%	83%	0.17
Ease of Use	**87.6%**	**72.3%**	-
Is "Camaleão" easy to use?	94%	89%	0.38
The "Camaleão" can be used without written instructions?	75%	61%	0.19
The "Camaleão" can be used successfully every time?	84%	67%	0.02
Ease of Learning	**93.6%**	**90.3%**	-
Did I learn to use it quickly?	100%	94%	0.16
Is it easy to remember how you use "Camaleão"?	100%	94%	0.16
Did I quickly become adept at using "Camaleão"?	81%	83%	0.56
Satisfaction	**93.8%**	**89.5%**	-
Am I satisfied with "Camaleão"?	100%	89%	0.08
Would I recommend "Camaleão" to a friend?	100%	94%	0.16
Does "Camaleão" work the way I want?	94%	83%	0.17
Is "Camaleão" pleasant to use?	94%	94%	0.53
Is "Camaleão" fun?	94%	94%	0.53
Is "Camaleão" suitable for its users?	100%	94%	0.16

In Satisfaction, the level of satisfaction of the participants who used the version with leaderboards is higher when compared to their version without leaderboards, but this difference is not significant. The level of interest of participants who used the version with leaderboards to indicate the application to another friend is also higher. The level of pleasantness and enjoyment of the participants who used the version with/without leaderboards is the same, with no percentage differences. However, the version with leaderboards presents a higher level of feel for the way the user wants to use the application. Also, in both versions, the instrument seemed suitable for the participants.

4. Final considerations and future work

In this paper, we analyze whether the leaderboards component influences the number of correct answers by students in solving exercises. Also, we applied a survey in order to

assess the usability of the application "Camaleão" with/without leaderboard. As a result, although the level of satisfaction of the participants who used the application with the leaderboards was higher, this difference was not significant. Also, the leaderboards did not influence the number of correct answers to problem-solving. These studies brought relevant information and essential characteristics for improving the tool and the use of gamified components in future work.

Numerous researches can be continued from this, among which some stand out for future work: improving the "Camaleão" application, adding new gamified components, such as weekly challenges and missions, medals and achievements, levels, rewards when completing an activity or challenge; implement history of solved exercises and student performance in each one, feedback with all successes and errors in each question answered; Also, apply new experiments with the enhanced "Camaleão" app to get a more extended time.

References

Aires, S. and Leite, J. C. (2017). Técnicas de motivação para colaboração no moodle imd. In *VII Workshop sobre Aspectos da Interação Humano-Computador para a Web Social*.

Borghetti, R., Scherer, L., Botelho, L. L. R., Jaeschke, J., Ajala, R., and Bolzan, E. (2017). O desafio da docência no ensino superior: Uma abordagem baseada na transição de paradigmas e a na crise do profissionalismo. In *XVII Colóquio Internacional de Gestão Universitária*.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 13(3).

Hamari, J., Koivisto, J., and Sarsa, H. (2014). Does gamification work? – a literature review of empirical studies on gamification. In *47th Hawaii International Conference on System Sciences*.

Lund, A. (2001). Measuring usability with the use questionnaire. *Usability and user experience newsletter*, 8(2).

Mekler, E. D., Brühlmann, F., Opwis, K., and Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation? an empirical analysis of common gamification elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications*.

Pereira, I. B. (2018). Uma proposta para concepção de um ambiente gamificado que auxilie no aprendizado acadêmico. Trabalho de Conclusão de Curso (Bacharelado em Computação), Universidade Estadual da Paraíba, Patos.

Soares, S. J., Dias, R. F. N. C., and Calegari, L. M. (2015). O uso das tecnologias digitais de informação e comunicação no processo de ensino-aprendizagem. In *XXI Congresso Internacional ABED de Educação a Distância*.

Vianna, Y., Vianna, M., Medina, B., and Tanaka, S. (2013). Gamification, inc: Como reinventar empresas a partir de jogos. *Rio de Janeiro: MJV*.