Does Tailoring Gamified Educational Systems Matter? The Impact on Students' Flow Experience

Wilk Oliveira
University of São Paulo
wilk.oliveira@usp.br

Armando M. Toda
University of São Paulo
armando.toda@usp.br

Paula T. Palomino
University of São Paulo
paulatpalomino@usp.br

Lei Shi
University of Liverpool
lei.shi@liverpool.ac.uk

Seiji Isotani
University of São Paulo
sisotani@icmc.usp.br

Ig Ibert Bittencourt
Federal University of Alagoas
ig.ibert@ic.ufal.br

Julita Vassileva
University of Saskatchewan
jiv@cs.usask.ca

Abstract

Recent research has shown that using gamification can prone to impact negatively on the motivation of students in educational systems. One of the reasons is that people are motivated or demotivated by different gamification elements according to their "gamer type". Thus, one of the main challenges in this field is to tailor gamified educational systems based on the students' "gamer type" and investigate if this kind of system presents better results than the counter-tailored gamified educational systems. This paper aims to investigate the effects of a tailored gamified educational system based on gamer type in terms of students' flow experience. We conducted an experiment with 121 Brazilian elementary school students comparing a tailored version against a counter-tailored version of a gamified educational system in terms of students' flow experience. The main results indicate that there is no significant difference in terms of students' flow experience, surprising and contradicting recent important studies in this field.

1. Introduction

Despite the fact that educational systems play an important role in providing solutions to educational problems (*e.g.*, students' evasion, disengagement and demotivation), recent studies have showed contradictory results regarding the efficacy of these systems, such as students' retention in the virtual course *e.g.,* [33, 37, 5]. This makes room for more research in this field, specially, the design of better educational systems in terms of students' learning outcomes.

To contribute in solving students' evasion, disengagement and demotivation in educational systems, previous research has been using gamification elements[1] along with pedagogical activities to,

for example, decrease students' frustration and demotivation, or improve students' concentration, engagement, and learning [16, 36, 37]. These studies are implementing and evaluating the gamification elements in the educational systems, raising the concept of gamified educational systems (GES).

A series of studies show that, in many cases, using gamification in an educational context (especially gamified intelligent educational systems) is not necessarily effective to improve students' motivation, engagement, and learning gains [21, 33, 32]. Therefore, there is a need for more in-depth research, especially, in order to identify when and how the use of gamification is really effective to improve the students' satisfaction with the system and to propose solutions to provide an effective design of gamification.

Recent studies have highlighted the hypothesis that this situation occurs because students that use GESs have different behaviors, expectations, and needs, and are encouraged in different ways [35, 23, 30]. It is also important to consider that students have different gamer types[2], so they can be more (or less) motivated by different elements, according to their gamer type and the gamification elements implemented in the system [35, 32, 24, 23]. For example, if a student is competitive, it is more likely that they prefers to earn points and compete in missions, whilst they might be more reluctant to participate in collaborative or interactive activities. Thus, depending on the approach used in these systems, the final results can also decrease the students' motivation, engagement, flow experience and so on when a system presents gamification elements that are not part of the preference of some students [28, 35, 33].

Moreover, according to different secondary studies *e.g.*, [29, 13, 14], most of the GESs provide different gamification elements to students, such as points,

---

[1] Gamification is the use of game elements to engage and motivate

people to achieve their goals, outside of a game context Deterding *et al.* [12].

[2] The term gamer type is used to represent the different perceptions that we have (not just players) regarding the gamification design.

HICSS

badges, trophies, ranking. However, these systems are using the same gamification elements for all students (the "one-size-fits-all" approach or monolithic group) [33], which may be harmful to the students, who need to interact with specific gamification elements that are not on their preference. This highlights the recent big challenge presented by Orji *et al.* [32], Monterrat *et al.* [24], Masthoff *et al.* [23] and others, the need to creating tailored GES based on the students' gamer type, that is, systems with gamification adapted according to the preferences and needs of each gamer type.

To target the challenge of providing tailored GESs for each student and hypothesizing that students have different perceptions depending upon their gamer types, in this paper we conducted a mixed factorial experimental study, comparing the effects of a tailored GES based on students' gamer types (implemented based on a dynamic computational approach capable of tailoring GES according to the students' gamer types) with a counter-tailored GES in terms of students' flow experience [6].

The flow experience was chosen because it is an important construct directly related to students' motivation and engagement, especially in this type of environment [9, 31]. Thus, the flow experience is an experience totally related to the performance of the students in educational systems and highly related to the students' own learning [6, 7, 8]. In order to compare the implemented GES with gamer type-based tailoring, our experiment was conducted with 121 elementary students in Brazil, comparing the tailored versus counter-tailored version of the systems with respect to flow experience achieved by students when using both systems.

Our results indicate that there is no significant difference in terms of students' flow experience in the tailored and counter-tailored version of the system, contradicting the expectations of important recent theoretical studies *e.g.*, [24, 35, 32], emphasizing that this is still a new area and that the conduction of new experimental studies is highly important for the maturation of the area.

2. Related works

This section outlines studies that are related to tailored GES (TGES) and flow experience in GESs. We chose to retrieve and analyze studies considering the following criteria: *a)* published in the last 10 years; and *b)* that explicit addressed the use of TGES. We did not limited our search to works with empirical evidence due to the fact that most of these studies are theoretical.

Sajjadi *et al.* [39] proposed a model for dynamic adaptation where various aspects were used to provide a more personalized gaming experience and to maintain the engagement of the players in the game. The model allowed to adopt an educational game based on different aspects in order to provide a student's engagement and flow experience [39]. The study, however, did not provide concerns about the use of a player model and the relationship between the user's gamer type and their game elements preferences.

Monterrat *et al.* [26] presented a generic and adaptive gamification system that can be plugged on various learning environments. In their proposal, the system adaptation happens when a new user registers to the learning environment. The adaptation happens only after the students have been using the system for a while. This study used the initial player model proposed by Ferro [16], that considers only a few number of player types. At the same time, they do not consider more recent and robust player models, such as *BrainHex* [28] or *HEXAD* [41].

Monterrat *et al.* [25] provided a generic architecture that can be used to adapt the gamification in different kinds of systems, independent from the pedagogical design. This study does not present concerns in terms of the choice of player profiles, or provide information about the architecture evaluation in terms of students' engagement or learning when they are using the personalized system based on the architecture proposed.

Gil *et al.* [18] conducted a study based on the users' experience in an educational environment with the objective of adapting and validating a personalization mechanic based on the players' profiles, incorporating the known mechanics of gamification in several learning activities. The authors developed the game mechanics based on the functionalities of an educational system and investigated the effectiveness of learning based on the proposed mechanics as well as the relationships between mechanics and their types of assumed players. In summary, the study of Gil *et al.* [18] had preliminary validated gamification mechanics and player types in the educational system and identified the gamification mechanics that were corresponded to the students' inferred player types.

In the field of TGESs, most of the studies were conducted in the last few years, and these recent studies are adapted from previous studies in another area. The empirical/experimental studies conducted in the field of education, in general, are focused only on the implementation or evaluation, not covering modern player types, such as the *BrainHex* or *HEXAD* player models, neither did they provide empirical evaluations of the proposed solution. Most of the studies also did not provide an evaluation either, in

terms of student's learning aspects, *e.g.*, user's flow experience. According to the results of our systematic literature review [15], to the best of our knowledge, our study is the first empirical/experimental study to compare a tailored with a counter-tailored version of a gamified educational system in terms of student's flow experience, considering a real educational system. Our results can serve as a basis for the conduction of new studies in this domain and open space for new experimental studies.

3. Experiment

This experiment aims to compare a tailored gamified educational system with a counter-tailored gamified educational system in terms of student's flow experience. The experiment is classified as comparative experiment with one independent variable and with ten possible values or "levels" (based on [43] and [42]), as further explained. This experiment was approved by the University of Saskatchewan Behavioral Research Ethics Board (Project BEH#16-142).

3.1. Method

In our experiment, we used the Goal/Question/Metric (GQM) method [3, 43], that is an approach to software metrics. This method assumes that for an organization to measure in an accurate way it must [43]:

1. Specify the goals for itself and its projects;

2. Trace those goals to the data that is intended to define those goals operationally; and

3. Provide a framework for interpreting the data regarding the goals that were established.

We applied the GQM method in our experiment design, which resulted in the specification of a measurement model targeting a set of issues and rules for the interpretation of the measured data [3, 43]. Before explaining the design of the experiment, we present below the process we are relying to tailor GES in this work.

3.2. Process to tailor the gamified educational systems

The process was proposed based on Orji's guidelines [35, 33] which are a systematic structure providing the most suitable PT strategies for each *BrainHex* gamer type and the game mechanics associated with each PT strategy addressed. In other words, these are the

persuasion strategies which better motivate each gamer type and its game mechanics.

The tailoring process provides a registration system to the participant, and this system collects personal data (according to the general specifications of the system *i.e.* students age, gender, and level). The system needs to provide a way to identify the student's gamer type (*e.g.*, using the *BrainHex* questionnaire in the students first access in the system). After the student's gamer type identification, the implemented system based on our architectural design will be capable of providing a tailored gamification system (graphical interface), using as basis the student' gamer type. The general architecture is composed of five different steps and presented in detail by [30]. Figure 1 synthesizes this process:
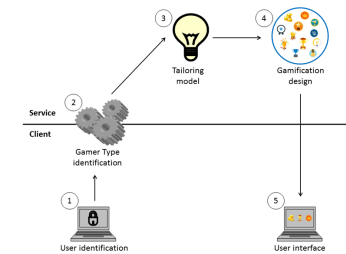

Figure 1. General view of process

1. **User identification:** in the first step, the system must provide options for the students to create an account and provide basic information, such as login, password, email, and others (according to the default system specifications). These data are important to associate the student with their gamer type. In the case of pre-existing systems (edited by the proposed structure), the basic information previously obtained can be used;

2. **Gamer type identification:** the gamer type identification consists of a semi-automatic process, based on the *BrainHex* player model. In the same moment of the user identification, the system should provide the *BrainHex* questionnaire to the students. After the student completes the questionnaire, the system processes the student' answers to identify students' gamer types, and to provide tailored interface according to the student' gamer type. In the case of pre-existing systems (edited by the proposed architectural design), the system can provide the *BrainHex* questionnaire after the student login in the system, and associate their personal data with their gamer type;

3. **Tailoring process:** the tailoring process is the main step of the process. The tailoring model is responsible for the identification of the student's gamer type, to make the game design elements available on the system, and to associate algorithmically the "correct" game design elements to each student's gamer type, based on the structure previously defined. So, in summary, in this step, the algorithm adapts gamification, selecting the most suitable elements (available in the system) for each gamer type considered in the approach;

4. **Gamification design:** the gamification design is the system model that is responsible for generating the graphical interface with the best game design elements associated with the student's gamer type, based on the tailoring model process before its execution;

5. **User's interface:** the user's interface represents the system output and is responsible for providing to the students the game elements associated to their player type, generated by the gamification design model.

3.2.1. Implementation The tailoring process and architectural design for gamified intelligent educational system based on gamer types was implemented in a GES called *MeuTutor* [36]. The system was chosen after a comparative analysis among nine different gamified intelligent educational system. *MeuTutor* was of interest to this study because it was considered more geographically accessible, as well it implements the nine most used gamification elements in gamified intelligent educational system (Points, Levels/Stages, Badges, Leaderboards, Prizes and Rewards, Progress bars, Storyline, and Feedback), as identified by Nah *et al.* [29], avoiding some validity threats. The system allowed the students study some topic through videos tutorials and hypertext, as well as answer different questions about the topic that was studied.

The complete process to tailoring the system compared in this experiment follows the proposed process in this study and presented in the subsection 3.2. It starts with the students' gamer type identification, continues with the selection of gamification elements, and finishes with different versions of the system, with the specific gamification elements to each gamer type. So, after logging into the system and answering the *BrainHex* questionnaire, our algorithm identifies the student' gamer type, selects the most suitable gamification elements for the identified profile, and

personalize the user interface only with the selected gamification elements (according to the proposed process).

3.3. Goal of the experiment

The main goal of this experiment is to evaluate the effectiveness of gamified intelligent educational system tailored according to the students' gamer types, in comparison with the counter-tailored version of the system in terms of students' flow experience.

3.4. Research Question and Hypothesis

RQ. Is the tailored system based on the students' gamer types more effective to lead the students to the flow state during the activities than the counter-tailored system?

According to the research question presented, we defined seven hypothesis in order to test our research question for each gamer type:

$H_{1 \cdot 0 \cdot 0}$: The Achiever students' flow experience in the tailored system is less or equal than the Achiever students' flow experience in the counter-tailored system.

$H_{1 \cdot 0 \cdot 1}$: The Achiever students' flow experience in the tailored system is greater than the achiever students' flow experience in the counter-tailored system.

The other hypotheses follow the same format, however, doing the tests for the other gamer types. The formal definition of the hypothesis has the objective to organize the hypothesis in a mathematical formulation. Following, Table 1 organizes our formal definition of the hypothesis.

Null Hypothesis
H_0: F (TS (Gt)) $\leq$ F (cTS (Gt))
Alternative Hypothesis
H_1: F (TS (Gt)) $>$ E (cTS (Gt))
Key: Gt: Gamer type; F: Flow experience; TS: Tailored System; cTS: Counter-tailored System.

Table 1. Formal Definition of the Hypothesis

3.5. Factors and response variables, experimental unities and participants

According to the hypothesis previously defined, our experiment has one factor: the gamified system with two levels (tailored and counter-tailored versions). The response variables are the students' flow experience. We defined a metric to measure these preferences in partnership with two academic professionals with expertise in Statistical and Behavioral Science. This

metric is an average between 1 and 7 according to the students' answers regarding their flow experience.

Our experimental unities are our sample composed of elementary students (organized according to their gamer types) interacting with gamified systems. We collected data from 121 Brazilian middle school students, in which 52 were males and 69 were females, aged between 11 and 13 years. Table 2 depicts our sample.

Gamer type	Sample	Male	Female
Achiever	48	18	30
Conqueror	28	15	12
Daredevil	6	5	1
Mastermind	3	0	3
Seeker	16	5	11
Socializer	9	5	4
Survivor	12	4	8
Total	**121**	**52**	**69**

Table 2. Sample of experiment

The study matches different criteria for an adequate sample size. According to Bentler [2] there must be a minimum ratio of 5 respondents per 1 construct in the model. Hair [19] suggest the same rule for factor analyses. Only the number of Masterminds (only three students) is not a number considered adequate for some statistical tests and therefore will be disregarded in the statistical tests from our study hereafter. In addition, although to increase the confidence of the results, a larger sample is ideal, considering the scenario (real classroom), the number of students becomes representative in relation to the number of students who normally participate in this type of study.

3.6. Data collection, metrics definition, and analysis

The experiment was conducted in a controlled environment (a laboratory with 20 computers with Windows 7 installed and free access to the internet). In the *first step*, the students responded to the *BrainHex* questionnaire to identify their gamer type. The data were stored using the pseudonym and association with questionnaire data and data logs. Data were stored in a password-protected computer system and to be available only for the investigators. A secure Cabinet was used to store the data for the long term.

The *BrainHex* [28] model acknowledges that the Gamer Types are not mutually exclusive and therefore, scores from each type are summed to find the player's dominant gamer type (primary type) and sub-types [32]. It includes 28 questions about the respondent game experiences and preferences in order to classify participants into their dominant gamer types [28, 33, 22].

In the *second step*, students were automatically divided in seven different groups according to their gamer types, inside each gamer type group; each group of students was randomly divided into two different groups, one group to use first the tailored system and other to use first the counter-tailored system (static version). In each usage section, the students used the system for at least 30 minutes, where they studied the subject Basic Operations (Mathematics) and then answer different questions on the subject.

In the *third step*, the students of the "experimental group" used the tailored version[3] for their gamer type and in the sequence responded to scale to measure flow experience. At the same time, the "control group" used the counter-tailored version (one size fits-all approach) and also responded the scale. Next, the groups were inverted, so the experimental group of the first execution became the control group and the control group became the experimental group.

In *the fourth* step, the students' answers were organized in tables (.csv files) and separated according to each gamer type and version of system (tailored and counter-tailored version), like the flow experience of the students in the tailored and counter-tailored version.

In the *fifth step*, the answers were organized into each constructed flow, according to the original scale factorial structure validated by [20]. Then, we calculated the individual student's flow experience. Finally, in the *sixth step*, we calculated the metrics defined (see next section) in order to compare the difference in terms flow experience for the seven different gamer types in the tailored and in the counter-tailored version of the system.

The metrics were extracted through the analysis of data previously collected, using different statistical software tools and the application of different statistical tests. In order to identify the students' flow state, we used the Dispositional Flow Scale-2 (DFS-2) [20]. The DFS-2 scale consists of structures based on the nine flow dimensions defined by [6], and empirically validated for the gamification domain by Hamari *et al.* [20]. Thus, the structures included in the DFS-2 are the following: time (transformation) (T), challenge-skill balance (CSB), merging of action and awareness (MMA), clear goals (G), feedback (F), concentration (C), control (CTRL), loss of self-consciousness (LSC), and autotelic experience (A). Each of the nine structures contains four items. Together, the structures form a 36-item scale for measuring the flow experience. The DFS-2 relies on self-reported data. Therefore, similarly

---

[3]Personalized version according to the student' gamer type.

to previous DFS-2 studies, the items were measured on a Likert scale ranging from strong disagreement (1 on the Likert scale) to strong agreement (5 on the Likert scale) with the statement [20].

After organizing our data in tables, we calculated the individual and general mean for each flow experience structure following the original DFS-2 factorial structure (challenge/skill balance: 1, 10, 19 and 28; merging actions – awareness: 2, 11, 20 and 29; clear goals: 3, 12, 21, 30; feedback: 4, 13, 22 and 31; concentration: 5, 14, 23 and 32; control: 6, 15, 24 and 33; loss of self – consciousness: 7, 16, 25 and 34; time: 8, 17, 26 and 35; autotelic experience [4]: 9, 18, 27 and 36).

In order to investigate the gamer type's flow experience differences in each version of the implemented system, we calculate each flow experience structure individually (a mean of the 36 questions). We calculate the median, standard deviation, average variation, maximum shared variation, average shared variation and normality of the data. Finally, based on the normality of data, we calculated the variance analyses between the mean of each flow experience design in the tailored and in the counter-tailored version of the system for each gamer type. The data of the preferences metric are normal; hence we decided to apply the *t-test* to verify the hypothesis. The application of the test, considering a $p \leq 0.05$ level of significance with a 95% confidence level. We also calculated the "effect size" for each gamer type in the tailored and counter-tailored version of the system. The data of the preferences metric are normal; hence we decided to apply the Cohen-test (with a 95% confidence level) to verify the effect size. The complete data set used in this study can be accessed in the supplementary materials available in the final of this paper.

4. Results

This section aims to present our study results. We organized the results in subsections according to each conducted analysis.

4.1. Flow experience comparison per game types and general

In this analysis, we conducted a comparison of how the flow experience occurred in each gamer type. The Table 3 presents the complete result of our analysis for each gamer type using the tailored and counter-tailored versions of the system (M = mean; SD = standard deviation). The last column of the table

---

[4]Autotelic experience, in the flow theory [6] is the result of an activity or situation that produces its own intrinsic motivation, rewards, or incentives, specifically without any outside goals or rewards.

shows the *t-value* and the *p-value*, which compares the average of each gamer type in the tailored and in the counter-tailored version of the system.

GT	CT		T		t-test	
	M	**SD**	**M**	**SD**	*t*	*p*
Ach.	3.528	0.729	3.553	0.700	-0.170	0.865
Con.	3.607	0.742	3.625	0.726	-0.095	0.925
Dar.	3.256	0.494	3.156	0.610	-0.284	0.783
See.	3.906	0.580	3.662	0.643	-1.052	0.303
Soc.	3.640	0.533	3.634	0.368	-0.036	0.972
Sur.	3.670	0.680	3.721	0.699	0.164	0.871
Gen.	3.585	0.693	3.580	0.666	0.050	0.960

Key: GT: gamer types; CT: counter-tailored; T: tailored; M: mean; Ach.: Achiever; Con.: Conqueror; Dar: Daredevil; See.: Seeker; Soc.: Socializer; Sur.: Survivor; Gen.: General

Table 3. Flow experience comparison per gamer types and general

Analyzing the individual results obtained for each gamer type, it is possible to perceive that in general, there was no significant difference in relation to the students' flow experience in the tailored and in the counter-tailored systems. Also, after calculating the effect size (using the Cohen-test), we identified that the difference was considerate "negligible" for all gamer types. This means that there was no significant difference in student flow experience in any of the analyzed scenarios, confirming our previously presented results.

However, it is very important to discuss this result and some specific points related to each gamer type, since the results may indicate, for example, that the personalization needs be more specific for each gamer type and this needs to be considered in future studies. Therefore, we will analyze in detail the results of each gamer type, trying to understand how each of the results can influence the design of adaptive gamified systems.

The Achiever students' flow experience was higher in the tailored versions of the system. So, we rejected our null hypothesis (the Achiever students' flow experience in the tailored system was greater than the achiever students' flow experience in the counter-tailored system). In addiction, analyzing some specific flow experience constructs, the tailored version presented better results (*e.g.*, "control" and "clear goals", in which the tailored version was better than the counter-tailored version), however, in others flow experience constructs, the perception was better in the counter-tailored system (*e.g.*, "challenge-skill balance" and "time transformation").

For the Conquerors, the flow experience was better in the tailored version, in comparison with the

counter-tailored systems, and we also rejected our null hypothesis. However, in similar fashion with the Achievers, the some flow experience constructs, for instance "time transformation" was better in the tailored version of the system. For the Conquerors, the "autoletic" experience was better in the tailored system.

Different from Achievers and Conquerors, the Daredevil's flow experience was better in the counter-tailored system. Therefore, for the Daredevil students, we don't rejected our null hypothesis. On some of the other flow experience constructs (*e.g.* "autoletic" experience and "time transformation"), the Daredevils had a better experience in the counter-tailored version. However, although in general the flow experience has been larger in the counter-tailored system, in many dimensions, for instance "challenge-skills balance", "clear-goals", and "feedback", the Daredevils had a better experience in the tailored version of the system, similar with the results obtained by [34].

The Masterminds were the only ones that presented positive difference in all of the flow experience constructs in the counter-tailored version. For this gamer type, the flow experience was ever better in the counter-tailored version of the system. In this case, we also do not rejected our null hypothesis flow experience. However, it is important to note that in the case of this gamer type, statistical tests were not applied due to the small number of participants with this gamer type.

The Seekers also presented a difference in most of the flow experience constructs. However, different then the Masterminds, the Seekers presented a better flow experience in the counter-tailored system. Therefore, the Seekers' flow experience was better in the tailored version in general, comparing to the counter-tailored system. So, for this case, we no reject our null hypothesis for the students' flow experience.

The Socializer students presented different levels in relation to each flow experience dimension. In other words, for some flow experience constructs, the experience of the Socializer was better in the counter-tailored system, and, in other cases, their experience was better in the tailored version of the system. In general, the flow experience of the Socializer students' was better in the counter-tailored system, so we no reject our null hypothesis.

The experience was similar for the Survivor students. As such, for some flow experience constructs the tailored system was better than the counter-tailored system, while in other flow experience constructs the counter-tailored system was better than the tailored system. The general flow experience was better in the tailored system, so, we reject our null hypothesis.

The comparison considering all gamer types using the tailored and the counter-tailored system, also show no significant difference in terms of student's flow experience. Finally, it is important to note that, despite the user flow experience, for some gamer types to have been higher in the tailored system and for others to have been higher in the counter-tailored system, the statistical test conducted in the experiment indicated that in all cases the statistical difference was not significant ($p \geq .05$). This result indicates that the personalization of the GES may not be relevant to the students.

4.2. Flow experience comparison per game types and and flow experience constructs

In this analysis, we conducted a comparison of how each flow experience dimension occurred with each of the gamer types. In all, 63 analyzes were made. For reasons of space, as well as to favor the discussion in the paper, we have included the complete table with all analyzes in the supplementary materials available at the end of the paper (Appendix 1). We also included in the supplementary materials a comparison chart for each flow dimension/gamer type (Appendix 2).

Even with this in-depth analysis, surprising the expectation of different previously conducted theoretical studies *e.g.*, [1, 35, 28], it was possible to realize that among all analyzes, only the LSC dimension presented a statistically significant difference ($p = 0.033 \geq .05$). In all other cases, there was no significant difference between the tailored and counter-tailored system.

Even more surprising, in the only construct in which there was a significant difference, it was apparent that students preferred the counter-tailored version over the tailored version. This may have been because, according to some studies *e.g.*, [33, 28], Seeker gamer types often simply prefer to choose the elements they want to use. Thus, the presentation of all gamification elements may have given Seeker students a sense of choice.

5. Threats to Validity

This section aims to present concerns that must be discussed in future replications of this study and other aspects that must be considered in order to maximize the results of the evaluation performed in this section [11]. We worked to minimize the threats discussed in this section. The section was organized according to the categorization of Wohlin *et al.* [43]: Internal, External, Construct and Conclusion.

Internal: The threats in category are: *(i)* history – perhaps at the time when the experiment occurred the users' experience may have been affected, this threat was minimized through a methodology that allowed participants to participate in the experiment without

the direct interference of other students and teachers.; *(ii)* maturation – Participants used the system for more than 30 minutes before responding to the questionnaire, so they may have been tired when responding to the survey; and *(iii)* positive or negative bias – as this experiment was conducted with participants without experience in this kind of evaluation, it is likely that the participants did not have a basis for comparison [40]. To minimize this threat, the two versions of the system (tailor and counter-tailored) were evaluated by the students in different moments.

External: The participants of the experiment are representative only in the context of elementary education. In this way, we might not be able to maximize the results of this experiment into different contexts [40]. The subjects of this evaluation should be expanded to other academic settings to obtain more generic results. The player model used in our study (*i.e.*, BrainHex) maybe be unreliable and have issues with its psychometric properties [17]. Thus, we also recommend future replications of our study using others different player models.

Constructs: This experiment has many different items and different aspects, however, some constructs may not be measured by the issues [40]. To minimize these threats, we selected methodologies and instruments empirically validated and commonly used in the scientific empirical studies from the technological and educational community. The system used in this experiment has an interface design (in terms of gamification elements), so the students' experience in the system can be influenced by the system design beyond the gamification elements. At the same time, the general quality of the system can also influence the student's experience and need to be observed in futures replication for this experiment.

Conclusion: The sample size of this experiment was of 121 students. However, this was a blocking factorial experiment, and some groups may have a small sample for confirm some hypothesis. The elementary students that were participants of the experiment could have been anxious in the moment that they answered the questionnaire, and they may have answered it without the proper attention. To mitigate this, we used only instruments previously validated in different domains, including gamification domains.

6. Discussions

From the actual discussions regarding the need to improve the education quality in different aspects, one of the main recent challenges is to identify the student's individual features, needs and preferences, in order to provide an adapted education to each group of students. Based on theses challenges and the growing use of gamification elements in educational systems, identifying the better gamification elements to motivate each group of students and providing a tailored system based on the students' gamer type are also two important recent challenges of the technological and educational communities.

In our study it was possible to identify that all of the gamer types had not a significant difference in terms of flow experience between in the tailored system and in the counter-tailored system. For some gamer types, in most of the flow experience constructs, the tailored system was better than the counter-tailored system, rejecting our null hypothesis, and showing that students can be more motivated for educational systems tailored according to their gamification preferences.

However, our results, in some cases surprised the expectations of the communities, contradicting some recent studies conducted in this field (*e.g.*, Orji *et al.* [35] Orji *et al.* [33], Monterrat *et al.* [27] and others) and not reject our null hypothesis, indicating that was not a difference in terms of flow experience, and, in addiction, for some gamer types, the flow experience were high in the counter-tailored system (although statistical confirmation has not yet been possible). It is important to note that although our results are different from the presented studies, there were experiments with different configurations and further studies are needed to further deepen the results.

A series of contribution were obtained from these results. For instance, the literature presents some initial studies about systems personalization based on gamer types, but our study is among the first specifically in the field of education. Our results also open a series of discussions about the educational systems personalization, and about how to tailor GES to better motivate each student gamer type.

Besides, our results indicate that the process used to implement the tailored system is capable of tailoring gamified intelligent educational systems based on students' gamer types, and, the tailored and the counter-tailored versions of the systems had different influences on the students experience. At the same time, we also suspect is that perhaps the flow experience is a very deep engaging experience and difficult to achieve in even GES.

In summary, our results show that the statistical difference between the tailored and counter-tailored versions of the system are not significant, so the study failed to prove that alternative hypothesizes are true or false. So, the result is still inconclusive indicating that this is an open research field, that needs to receive more

experimental studies, in order to investigate important things, *e.g.*, conducting studies to investigate which are the most suitable persuasive strategies and which are the most suitable gamification elements for each student' gamer type, specifically, in field of education.

7. Concluding Remarks

Gamification may produce good results in different studies, like, for instance, keeping the students motivated in the activities provided by the system or improving the students' learning aspects [36, 10, 4]. Further more, other studies show that gamification may cause the opposite effect [35, 33, 38]. One of the main hypothesis to this situation is that students have different gamer types and are motivated by different gamification elements. In order to investigate this hypothesis, we conducted an experimental study to compare the student's flow experience using a counter-tailored GES with the student's flow experience using a TGES.

The main results of this study indicate not significant difference in terms of student's flow experience between in the tailored and in the counter-tailored versions of the system for the gamer types, contradicting to some of the recent important studies conducted in this field. In some cases, the tailored system was more effective in terms of flow experience in comparison with the counter-tailored system for the majority of gamer types, confirming our expectations and the recent studies in this field. However, interestingly, the study also indicates that for other gamer types, the counter-tailored system was more effective than the tailored system, also contradicting to the results of recent studies conducted in different fields (emphasizing that always without statistical significance, contradicting the expectations of different theoretical studies).

In order to continue this research, as future works, we plan to conduct new studies, evaluating different constructs, such as students' hyper-focus, learning, and human-computer interaction of students in the tailored and counter-tailored versions of the system. We will also investigate the direct perception of each gamer type in terms preferences about gamification elements, and provide an automatic process to identify the students gamer types during their use in the system.

Acknowledgment

References

[1] R. Bartle. Hearts, clubs, diamonds, spades: Players who suit muds. *Journal of MUD research*, 1(1):19, 1996.

[2] P. M. Bentler and C.-P. Chou. Practical issues in structural modeling. *Sociological Methods & Research*, 16(1):78–117, 1987.

[3] V. R. B.-G. Caldiera and H. D. Rombach. Goal question metric paradigm. *Encyclopedia of Software Engineering*, 1:528–532, 1994.

[4] G. C. Challco, R. Mizoguchi, I. I. Bittencourt, and S. Isotani. Gamification of collaborative learning scenarios: Structuring persuasive strategies using game elements and ontologies. In *International Workshop on Social Computing in Digital Education*, pages 12–28. Springer, 2015.

[5] G. C. Challco, D. A. Moreira, I. I. Bittencourt, R. Mizoguchi, and S. Isotani. Personalization of gamification in collaborative learning contexts using ontologies. *IEEE Latin America Transactions*, 13(6):1995–2002, 2015.

[6] M. Csikszentmihalyi. *Finding flow: The psychology of engagement with everyday life*. Basic Books, 1997.

[7] M. Csikszentmihalyi. *Beyond boredom and anxiety*. Jossey-Bass, 2000.

[8] M. Csikszentmihalyi. *Applications of flow in human development and education*. Springer, 2014.

[9] M. Csikszentmihalyi. Toward a psychology of optimal experience. In *Flow and the foundations of positive psychology*, pages 209–226. Springer, 2014.

[10] S. J. de Santana, H. A. Souza, V. A. Florentin, R. Paiva, I. I. Bittencourt, and S. Isotani. A quantitative analysis of the most relevant gamification elements in an online learning environment. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 911–916. International World Wide Web Conferences Steering Committee, 2016.

[11] D. Dermeval and I. I. Bittencourt. Authoring gamified intelligent tutoring systems. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 14, 2017.

[12] S. Deterding, D. Dixon, R. Khaled, and L. Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

[13] D. Dicheva and C. Dichev. Gamification in education: Where are we in 2015? In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 1445–1454. Association for the Advancement of Computing in Education (AACE), 2015.

[14] D. Dicheva, C. Dichev, G. Agre, and G. Angelova. Gamification in education: a systematic mapping study. *Journal of Educational Technology & Society*, 18(3):75, 2015.

[15] W. O. dos Santos, I. I. Bittencourt, S. Isotani, D. Dermeval, L. B. Marques, and I. F. Silveira. Flow theory to promote learning in educational systems: Is it really relevant? *Brazilian Journal of Computers in Education*, 26(02):29, 2018.

[16] L. S. Ferro, S. P. Walz, and S. Greuter. Towards personalised, gamified systems: an investigation into game design, personality and player typologies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment: Matters of Life and Death*, page 7. ACM, 2013.

[17] G. Fortes Tondello, D. Valtchanov, A. Reetz, R. R. Wehbe, R. Orji, and L. E. Nacke. Towards a trait model of video game preferences. *International Journal of Human–Computer Interaction*, 34(8):732–748, 2018.

[18] B. Gil, I. Cantador, and A. Marczewski. Validating gamification mechanics and player types in an e-learning environment. In *Design for Teaching and Learning in a Networked World*, pages 568–572. Springer, 2015.

[19] J. F. Hair, W. C. Black, B. J. Babin, R. E. Anderson, R. L. Tatham, et al. *Multivariate data analysis*, volume 5. Prentice hall Upper Saddle River, NJ, 1998.

[20] J. Hamari and J. Koivisto. Measuring flow in gamification: Dispositional flow scale-2. *Computers in Human Behavior*, 40:133–143, 2014.

[21] J. Hamari, J. Koivisto, and H. Sarsa. Does gamification work?–a literature review of empirical studies on gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on System Sciences*, pages 3025–3034. IEEE, 2014.

[22] É. Lavoué, B. Monterrat, M. Desmarais, and S. George. Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies*, 2018.

[23] J. Masthoff and J. Vassileva. Tutorial on personalization for behaviour change. In *Proceedings of the 20th International Conference on Intelligent User Interfaces*, pages 439–442. ACM, 2015.

[24] B. Monterrat, M. Desmarais, E. Lavoué, and S. George. A player model for adaptive gamification in learning environments. In *International Conference on Artificial Intelligence in Education*, pages 297–306. Springer, 2015.

[25] B. Monterrat, É. Lavoué, and S. George. A framework to adapt gamification in learning environments. In *European Conference on Technology Enhanced Learning*, pages 578–579. Springer, 2014.

[26] B. Monterrat, E. Lavoué, and S. George. Motivation for learning: Adaptive gamification for web-based learning environments. In *6th International Conference on Computer Supported Education (CSEDU 2014)*, pages 117–125, 2014.

[27] B. Monterrat, É. Lavoué, and S. George. Toward an adaptive gamification system for learning environments. In *International Conference on Computer Supported Education*, pages 115–129. Springer, 2014.

[28] L. E. Nacke, C. Bateman, and R. L. Mandryk. Brainhex: A neurobiological gamer typology survey. *Entertainment computing*, 5(1):55–62, 2014.

[29] F. F.-H. Nah, Q. Zeng, V. R. Telaprolu, A. P. Ayyappa, and B. Eschenbrenner. Gamification of education: a review of literature. In *International conference on hci in business*, pages 401–409. Springer, 2014.

[30] W. Oliveira and I. I. Bittencourt. *Tailored Gamification to Educational Technologies*. Springer, 2019.

[31] W. Oliveira, A. Toda, P. Palomino, L. Rodrigues, L. Shi, and S. Isotani. Towards automatic flow experience identification in educational systems: A

theory-driven approach. In *Proceedings of the 2019 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, 2019.

[32] R. Orji. *Design for behaviour change: a model-driven approach for tailoring persuasive technologies*. PhD thesis, University of Saskatchewan, 2014.

[33] R. Orji, R. L. Mandryk, and J. Vassileva. Selecting effective strategies for tailoring persuasive health games to gamer types. In *Graphics, Animation and New Media*, 2014.

[34] R. Orji, R. L. Mandryk, and J. Vassileva. Improving the efficacy of games for change using personalization models. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 24(5):32, 2017.

[35] R. Orji, R. L. Mandryk, J. Vassileva, and K. M. Gerling. Tailoring persuasive health games to gamer type. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2467–2476. ACM, 2013.

[36] R. Paiva, A. Barbosa, E. Batista, D. Pimentel, and I. I. Bittencourt. Badges and xp: An observational study about learning. In *Frontiers in Education Conference (FIE), 2015 IEEE*, pages 1–8. IEEE, 2015.

[37] R. Paiva, I. I. Bittencourt, T. Tenório, P. Jaques, and S. Isotani. What do students do on-line? modeling students' interactions to improve their learning experience. *Computers in Human Behavior*, 64:769–781, 2016.

[38] L. Rodrigues, W. Oliveira, A. Toda, P. Palomino, and S. Isotani. Thinking inside the box: How to tailor gamified educational systems based on learning activities types. In *Proceedings of the Brazilian Symposiumon Computers in Education*, 2019.

[39] P. Sajjadi, F. Van Broeckhoven, and O. De Troyer. Dynamically adaptive educational games: A new perspective. In *International Conference on Serious Games*, pages 71–76. Springer, 2014.

[40] W. O. Santos, T. Gomes, and C. Silva. Towards to flow state identification in educational games: An empirical study. In *Brazilian Symposium on Computers in Education*, volume 28, page 927, 2017.

[41] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke. The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play*, pages 229–243. ACM, 2016.

[42] R. Van Solingen and E. Berghout. *The Goal/Question/Metric Method: a practical guide for quality improvement of software development*. McGraw-Hill, 1999.

[43] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslén. *Experimentation in software engineering*. Springer Science & Business Media, 2012.

A. Supplementary Materials

Through the link below you can access the complete materials used in this study: `http://twixar.me/ZFsn`