Personalization Improves Gamification: Evidence from a Mixed-methods Study

LUIZ RODRIGUES, University of São Paulo, Brazil
PAULA T. PALOMINO, University of São Paulo, Brazil and University of Waterloo, Canada
ARMANDO M. TODA, University of São Paulo, Brazil
ANA C. T. KLOCK, Federal University of Rio Grande do Sul, Brazil and Tampere University, Finland
WILK OLIVEIRA, University of São Paulo, Brazil and Tampere University, Finland
ANDERSON P. AVILA-SANTOS, University of São Paulo, Brazil
ISABELA GASPARINI, Santa Catarina State University, Brazil
SEIJI ISOTANI, University of São Paulo, Brazil

Personalization of gamification is an alternative to overcome the shortcomings of the one-size-fits-all approach, but the few empirical studies analyzing its effects do not provide conclusive results. While many user and contextual information affect gamified experiences, prior personalized gamification research focused on a single user characteristic/dimension. Therefore, we hypothesize if a multidimensional approach for personalized gamification, considering multiple (user and contextual) information, can improve user motivation when compared to the traditional implementation of gamification. In this paper, we test that hypothesis through a mixed-methods sequential explanatory study. First, 26 participants completed two assessments using one of the two gamification designs and self-reported their motivations through the Situational Motivation Scale. Then, we conducted semi-structured interviews to understand learners' subjective experiences during these assessments. As result, the students using the personalized design were more motivated than those using the one-size-fits-all approach regarding intrinsic motivation and identified regulation. Furthermore, we found the personalized design featured game elements suitable to users' preferences, being perceived as motivating and need-supporting. Thus, informing i) practitioners on the use of a strategy for personalizing gamified educational systems that is likely to improve students' motivations, compared to OSFA gamification, and ii) researchers on the potential of multidimensional personalization to improve single-dimension strategies. For transparency, dataset and analysis procedures are available at https://osf.io/grzhp/.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; • **Applied computing** → *Education.*

Additional Key Words and Phrases: Gamification in education; gameful; tailoring; adaptation; self-determination theory

ACM Reference Format:
Luiz Rodrigues, Paula T. Palomino, Armando M. Toda, Ana C. T. Klock, Wilk Oliveira, Anderson P. Avila-Santos, Isabela Gasparini, and Seiji Isotani. 2021. Personalization Improves Gamification: Evidence from a Mixed-methods Study. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 287 (September 2021), 24 pages. https://doi.org/10.1145/3474714

Authors' addresses: Luiz Rodrigues, lalrodrigues@usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590; Paula T. Palomino, paulatpalomino@usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590, University of Waterloo, Canada; Armando M. Toda, armando.toda@usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590; Ana C. T. Klock, actklock@inf.ufrgs.br, Federal University of Rio Grande do Sul, Brazil, Tampere University, Finland; Wilk Oliveira, wilk.oliveira@usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590, Tampere University, Finland; Anderson P. Avila-Santos, anderson.avila@usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590; Isabela Gasparini, isabela.gasparini@udesc.br, Santa Catarina State University, Brazil; Seiji Isotani, sisotani@icmc.usp.br, University of São Paulo, Av. Trab. São Carlense, 400, São Carlos, São Paulo, Brazil, 13566-590.

1 INTRODUCTION

Motivation is important for positive learning experiences and academic success, since it predicts learning gains and enhances performance [27, 57]; while its lack is associated with lower performance and high drop-out rates [55, 78]. Moreover, students' lack of motivation is among the major issues faced by instructors [41, 53, 68]. Gamification can aid with that issue as it promotes game-like experiences aiming to improve users' motivations [18, 30]. Overall, gamification has a positive impact on psychological learning outcomes (e.g., intrinsic motivation), compared to non-gamified interventions [62]. However, there are cases in which it is associated with the opposed: undermined intrinsic motivation [71]. Many scholars attribute such cases to bad designs [44, 69], often arguing that providing the same game elements to everyone (i.e., the one-size-fits-all - OSFA - approach) is unlikely to make gamification work in its full potential since each user has different expectations, needs, and preferences [31, 49, 73]. Thus, the need for an approach to avoid the limitations of OSFA gamification, which might harm students' learning experiences while expecting to improve them.

Tailoring gamification might be the approach to mitigate OSFA gamification's drawbacks. It provides different game elements to different users/contexts through *customization* (i.e., users freely choose the game elements they prefer the most) or *personalization*[1] (i.e., the system or designer adapts the game elements according to each user/context) [72]. While there is some evidence customization improves OSFA gamification in terms of, for instance, user performance, customization requires the users/learners to indicate their game elements preferences before using the gamified system [42, 74]. Differently, personalization of gamification does not have such requirement because it is accomplished through predefined rules applied by the system/designer [72]. However, despite much research on personalization of gamification has been published, most empirical studies compare it to random or counter-tailored gamification [60]. Practitioners will not design gamified systems based on such approaches because of the uncertain effect of the former and the probable negative impact of the latter. Therefore, such comparisons do not provide evidence that advances our understanding of whether personalization is beneficial in practice. In contrast, a few studies compared personalized to standard, OSFA gamification empirically (e.g., [46, 48]), but those are unclear on whether the former improves the latter, especially in the educational domain.

A possible reason for the inconclusive findings is that such research (e.g., [46, 48]) applied personalization strategies that consider a single user dimension for the personalization. Whereas, a dual personalization approach resulted in positive motivational outcomes in [67], providing a valuable contribution on the potential of considering more than one personalization criterion. From such context, we expect multidimensional personalization[2] to approximate the positive effects of customization (e.g., [42, 66]) because users are not reduced to a single dimension; while avoiding the burden of asking their game element preferences [2, 54]. However, the only study (to our best knowledge) applying multidimensional personalization of gamification with users did not compare it to the OSFA approach [67]. Thus, the need for empirical research to advance our understanding of the potential of personalized gamification in educational contexts, compared to what is likely to be used in practice (OSFA rather than random or counter-tailored), along with the promising of multidimensional personalization. Therefore, **the goal of this paper was to understand the effect of gamification personalized to multiple dimensions, compared to the OSFA approach, on users' motivations in assessment learning tasks**.

To achieve that goal, we conducted a mixed-methods sequential explanatory study [15]. Software Engineering students (N = 26) were randomly assigned into experimental (i.e., multidimensional personalized gamification) or control (i.e., OSFA gamification) groups to complete two classroom assessments, and feedback was collected through

[1] Or *static adaptation* according to Hallifax et al. [25]
[2] That is, personalizing to multiple dimensions (criteria).

standard human-computer interaction (HCI) methods: the Situational Motivation Scale (SIMS) [22] and semi-structured interviews [6]. The findings indicate a positive effect of the multidimensional personalization of gamification on intrinsic motivation and identified regulation compared to OSFA gamification and suggest the personalized gamification designs featured game elements suitable to users' preferences, a possible reason for students perceiving it as motivating and need-supporting, and for the positive quantitative results. Therefore, contributing initial promising evidence on the effectiveness of a personalization approach that, compared to the standard, OSFA gamification, is perceived as more motivating by the students. Thus, our contribution informs: i) designers of educational systems on how to gamify them towards enhancing students' experiences, which will likely improve their learning due to its relationship with motivation and ii) researchers of gamification with indication that multidimensional personalization can improve the OSFA approach, expanding findings from [67] that are similar but compared to random instead of OSFA gamification, hence suggesting previous personalization strategies did not work compared to the OSFA approach because they considered a single personalization criterion.

2 RELATED WORKS

This section provides background information on gamification in education, and customization and personalization of gamification.

2.1 Gamification in Education

Education is the domain with the most gamification research [35]. Meta-analyses of gamification's effect within the educational domain show its effectiveness, compared to no gamification, while demonstrating that gamification might affect different types of learning outcomes (e.g., psychological and behavioral) and that its effect depends on several moderators [4, 62]. For instance, gamification has been used in education to improve students' grades, satisfaction, motivation, and lecture attendance, among other goals [33]. Accordingly, understanding how gamification works is important to enable a proper evaluation of whether it is working as expected.

While the Theory of Gamified Learning [37] has been considered suitable to understand gamification applied to educational scenarios [62], it does not account for psychological states, which play an essential role in gamification [35]. Differently, the Gamification Science framework [38] does, which is built around four main constructs: *predictors* (game elements); *criteria* (the distal outcome to be affected; e.g., students' knowledge retention); *mediators* (users' psychological states, e.g., intrinsic motivation; and behaviors, e.g., completing quizzes); and *moderators* (independent factors that might increase/decrease predictors' or mediators' effects) [38]. Based on those, the framework suggests the following:

- Predictors affect Psychological states;
- Psychological States affect both Behaviors and Distal Outcomes;
- Behaviors affect Distal Outcomes;
- Moderators change the effect of all previous connections.

Accordingly, regardless of the behavior/distal outcome targeted, gamification must first affect users' psychological state. This is aligned to claims that low performance (distal outcome) and drop-out rates (behavior) are related to students' lack of motivation (psychological state) [55, 78]. For instance, consider using gamification to improve class attendance and, consequently, students' grades. Failing to improve grades (distal) might be because gamification did not motivate (psychological state) them to attend class (behavior) or because they felt motivated but instruction was poor.

Importantly, not considering the psychological state could lead to a misinterpretation that gamification failed, whilst it worked as expected and the problem was not related to it [38]. Such context shows the importance of motivation, and other psychological states, on gamification in education research as it is on the starting point of gamification functioning as well as among the most important factors for learning [27, 55, 57, 78]; a possible explanation for motivation being the most studied construct in gamification research [81]. Thus, considering motivation in gamification in education studies is pertinent because of both its value to learning and its role in preventing misinterpretations of whether it worked.

2.2 Customized Gamification

Customized gamification occurs when users are in charge of choosing the game elements the system presents [72]. Research on that approach has implemented it by allowing users to select the game elements they want from a predefined list (also referred to as *bottom-up* gamification) [74] or to create their gamification (i.e., participants writing their idea of a gamification design) [66]. Experimental studies comparing customized and OSFA gamification suggest positive effects of customization on behavioral outcomes (e.g., number of tasks solved) [42, 74]. While customization benefits have been discussed in terms of the freedom of choice it provides, it did not affect participants' feelings of autonomy, enjoyment, competence, and pressure [66]. Consequently, this raises the question of whether the performance improvements found are related to the customized gamification design or due to the effort participants put into creating them.

Customization's advantage likely emerges from providing individualized gamification designs. That is, all of someone's characteristics and preferences, as well as the context in which gamification will be used, are considered when selecting the game elements. On the other hand, the need for selecting game elements might become a burden. The effort for a one-time selection might be acceptable. However, discussions that gamification must be aligned to the task [25, 43, 56] suggest that, ideally, user selection would have to be done for each task. Hence, implementing customization might end up leading to substantial efforts from users, and designers/developers when users create their gamification.

2.3 Personalized Gamification

Personalization of gamification is when designers or the system itself chose game elements based on user information [72]. Predominantly, studies have personalized gamification by capturing user information and using it as an input to the gamified system, which provides a tailored design accordingly. For such approach, researchers often rely on preference-based recommendations. For instance, [73] and [24] show insights on the most suitable game elements depending on one's HEXAD and BRAINHEX types (i.e. motivational profiles), personality traits, gender, and age. Common to most recommendations is that the game elements are to be defined based on a single criterion [32], whereas there is empirical evidence demonstrating multiple factors affect users' preferences and, therefore, gamification success [19, 26, 31].

Furthermore, the effectiveness of personalization recommendations, when applied with users, are unclear [60]. Research comparing personalized gamification to random, counter-tailored, or no game elements are mostly positive (e.g., [1, 40, 67]). Similarly, some studies that compared personalized and OSFA gamification are also positive, while they failed to properly isolate the analyzed conditions (i.e., featuring both personalized learning and gamification for the experimental group [16, 28]). Nevertheless, most findings from comparing personalization and OSFA gamification are inconclusive. Mora et al. compared gamification personalized to students' HEXAD user types to the OSFA approach in the context of undergraduate learning. Their findings suggested a positive but statistically non-significant effect of personalization on psychological and behavioral outcomes [46]. Oliveira et al. also compared personalized and OSFA experimentally, tailoring the game elements according to participants' player types [48]. While findings were inconclusive as well, the negligible effect size suggested a null effect. In contrast, Hajarian et al. described positive (e.g., time-on-system and number of page views) outcomes when comparing personalized gamification to the OSFA design [23]. Besides the different context (i.e., social networks rather than educational), their personalization strategy was built from interaction preferences and implemented in the same system, whereas other studies [46, 48] implemented recommendations from surveys in real systems.

From that context, it is evident that there is a lack of empirical studies comparing personalized and OSFA gamification. Nevertheless, those studies have different characteristics that can be used as insights to improve personalization strategies. First, they differ in context (i.e., learning activities [46, 48] and social networks [23]), which might be a reason for the contradictory findings. Second, personalizing based on interaction instead of survey preferences might be another one [2, 54]. Consequently, Hajarian et al. [23] personalized based on data from the same context and activity, while others [46, 48] implemented recommendations from general preferences. Another point is that interaction data capture users' behaviors based on several of their dimensions, if not all, similar to what customization enables. The other empirical studies, however, personalized to a single dimension [46, 48]. Therefore, raising the question of whether personalizing gamification to multiple dimensions of a specific context would improve gamification, contrary to approaches based on a single dimension.

To personalize gamification to multiple dimensions, one needs recommendations that consider two or more information simultaneously, which are scarce [32]. Oyibo and colleagues analyzed how users' culture, age, and gender affect their persuasiveness to social influence's constructs (e.g., rewards and competition) in a series of studies [50–52]. In common, their recommendations do not consider any situational dimension. Bovermann and Bastiens suggest the most suitable game elements considering users' HEXAD user type and the learning activity [7]. However, their recommendations are limited to common Moodle activities. Differently, Baldeon et al. provide recommendations to personalize gamification based on users' characteristics (i.e., HEXAD user type, personality traits, and learning style) and the teaching activity (e.g., brainstorm and group discussion) [5]. Rodrigues et al. provide a recommender system that suggests game elements based on users' information (e.g., gender, preferred game genre and playing style, weekly playing time, and geographic location and experience with gamification), as well as the cognitive process to be worked in the learning activity [59]. A limitation common to all of those is that neither has been applied with users to evaluate its effects compared to the OSFA approach. Hence, there are some limited options to aid in a multidimensional personalization of gamification considering user and contextual/situational information. Mainly, they vary in dimensionality (i.e., the number of criteria considered) and generalization of the situational factor (i.e., whether it guides on predefined [5, 7] or any [59] task).

2.4 Summary and Hypothesis

Despite empirical evidence suggest customized gamification's benefits, it might require unfeasible efforts from users and designers/developers to be implemented. Differently, personalized gamification takes that burden away from users but demand predefined strategies on how to tailor gamification to them. Despite much research on personalized gamification, most compare it to random and/or counter-tailored approaches and few compare it to the OSFA [60]. That is problematic because to contribute to understanding personalization's effects we need evidence from comparing it to standard gamification designs. However, there is inconclusive evidence on whether personalization leads to improvements compared to the OSFA approach from the few studies available [25, 60]. Perhaps, the fact that prior research in the educational domain [46, 48] personalized gamification to a single dimension, when comparing to the OSFA approach, might be the reason for the inconclusive findings. In that line, gamification personalized to two dimensions resulted in positive results when applied with learners, but compared to random, not OSFA gamification [67]. Compared to

Table 1. Comparison between this study and main related work (either compared to OSFA gamification or applied multidimensional personalization). For a complete comparison of empirical research on personalized gamification, see [25, 59].

Work	Compared to OSFA gamification?	Multidimensional personalization?
[46]	Yes	No
[48]	Yes	No
[67]	No	Yes
This one	Yes	Yes

those, our study mainly differs by evaluating the effects of gamification personalized to multiple dimensions compared to the OSFA approach (see Table 1). Therefore, we focused on personalization aiming to achieve results similar to those of customization while preventing the burden on users and developers/designers. In doing so, we implement the recommendations from Rodrigues et al. [59] because it has the highest dimensionality (i.e., eight factors/criteria), the more generic contextual factor (i.e., task's cognitive process, which is independent of the system, e.g., Moodle, and teaching activity, e.g., discussion), and does not require users completing long questionnaires (e.g., to assess personality traits or learning style). Thus, testing the hypothesis that **a multidimensional personalization of gamification, considering user and contextual information, improves the one-size-fits-all approach in terms of learners' motivations.**

3 METHOD

The goal of this study was to understand the effect of gamification personalized to multiple dimensions, compared to the OSFA approach, on users' motivations in assessment learning tasks. Therefore, we conducted a mixed-methods sequential explanatory study [15]. In the first phase, we compared OSFA and personalized gamification through a 2x2 mixed factorial experiment. We manipulated *gamification design* (between-subject) to create two versions of the system where students would complete the assessments. Those versions featured either an OSFA or the personalized gamification design. Participants engaged in two sessions that differed by the assessment *discipline* (within-subject): Programming Techniques and Object-Oriented Analysis and Design. Thus, we were able to compare the gamification designs based on two applications. In the second phase, we conducted semi-structured interviews to understand participants' motivations to use and engage with the gamified system. Figure 1 presents an overview of the study, which was reviewed by the institution's ethical board. It shows a flow chart demonstrating the steps followed during the preparation and execution of the experiment, indicating the time interval in which they happened as well as descriptions of particular aspects of those steps.

3.1 Context and Participants

Sampling was made by convenience according to the willingness of the instructor to apply gameful interventions in their lessons. Accordingly, this study was conducted with second-period students of the undergraduate Software Engineering course of SENAI Londrina, a small, private institution from Brazil. Despite being conducted in a natural context (i.e., the assessment activities were part of the disciplines' programs), participation in this experiment was voluntary. Twenty-six out of 27 learners agreed to participate in this experiment (all men, with an average - M - of 21.92 years and a standard deviation - SD - of 3.77), a sample size similar to those seen in overall HCI studies [11] and CHI papers [10]. Twenty-three subjects attended the Programming Techniques discipline during the first day, and 22

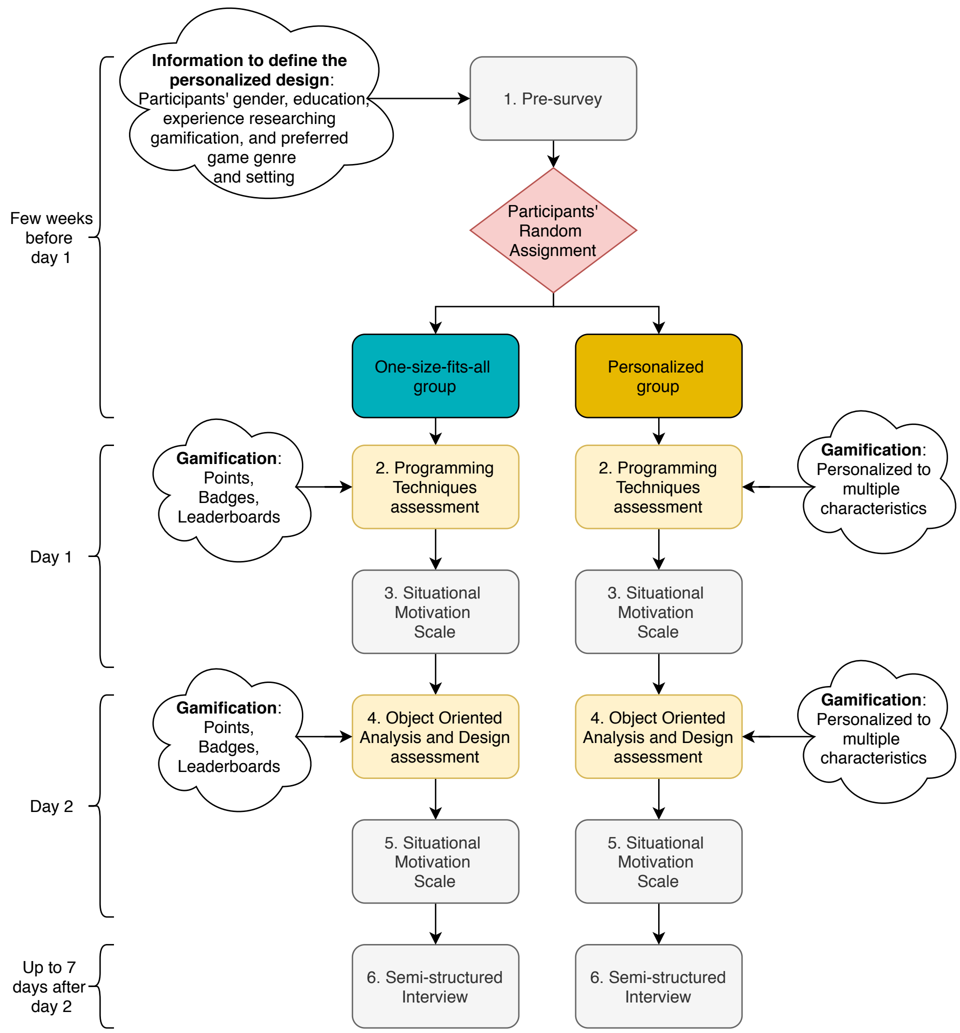

Fig. 1. Study methodological overview. The figure shows each step of the experiment, along with particular characteristics of some of those steps as well as the time interval in which they happened.

attended the Objected-Oriented Analysis and Design discipline during the second day. Of all participants, 4 agreed to participate in the interviews (Age - M: 23.75; SD: 4.65; all had been randomly assigned to the personalized gamification design). The same instructor (male, 31 years old, holds an MsC. in Computer Science, 6 years of teaching experience) taught both disciplines. All participants provided written consent through an online form since classes were held online due to the COVID-19 pandemic.

3.2 Gamified Assessments

We worked with two classroom assessments, composed of 30 multiple-choice items each. Specifically, each assessment can be seen as a test composed of 30 multiple-choice items. Notice that there are two forms of assessment, this experiment relies on assessment *for* learning, as opposed to assessment *of* learning [34]. Because students completed the assessments towards the end of the semester and before the final exam (which is an assessment of learning), this study's assessments aimed to make students' remember concepts introduced during the semester to complete its items. Hence, as aligned with the instructor, this would help the students to recall important information about the course. Accordingly, learners had to work on the remembering dimension of the cognitive process, according to Bloom's taxonomy [36], which was further used as one of the personalization criteria of the gamification designs (see Section 3.3).

To ensure the assessment was exclusively within that dimension, all items were developed by the class instructor and one researcher, while being validated by an independent instructor. For instance, one item reads as follows: *If 'p' is a pointer, the 'p' and '*p' commands access, respectively: a) a memory address and the value stored in the address; b) a value and the memory address where the value is stored; c) a memory address and the name of the variable to which 'p' is pointing; d) a value and the name of the variable that 'p' is pointing to.* The full version of both original (i.e., Brazilian Portuguese) and English versions of these assessments are available as supplementary material.

These assessments were implemented in the gamified system Eagle-Edu[3] since our participants had experience with it because other instructors used the system in their classes. Eagle-Edu was also suitable for this study because it allows creating multiple-choice activities, such as the ones from the proposed assessments. Additionally, this system enables instructors to choose which game elements (e.g., objectives, progress, storytelling, points, badges, leaderboards, and time pressure) they want to enable or disable, assisting in creating multiple gamification designs. In practice, each assessment translated into 10 quizzes of 3-multiple-choice items, providing rewards when the learner completed each quiz. Table 2 describes Eagle-Edu's implementation of the game elements used in this study according expert-validated definitions [70]. Next, we explain the rationale for selecting game elements.

Table 2. Description of Eagle-Edu's implementation of the game elements used in this study, along with the indication of which of them were available in each gamification design. For further description of gamification design definitions, see Section 3.3

Game element	Description	OSFA	P1	P2
Acknowledgment	Badges for achieving predefined goals (e.g., completing a quiz with no error).	X	X	
Chance	Surprise item, such as virtual goods, that appears due to randomness and the user might collect.			X
Competition	Leaderboard showing users sorted by the points they have earned; highlights the first and last two users.	X		X
Objective	Skills tree demonstrating course's topics/skills to guide users towards short-term goals (i.e., completing each topic at a time)		X	
Points	Numerical feedback for achieving predefined goals (e.g., completing 10 quizzes).	X		
Progression	Progress bar indicating one's progress within quizzes and on topics/skills.		X	
Time Pressure	Decreasing timer showing the time left to improve in the leaderboard.			X

OSFA = One-size-fits-all condition; PN = Version N of the personalized gamification condition

[3]eagle-edu.com.br/

3.3 Experimental Conditions

When designing the gamification versions (OSFA and personalized), we ensured they featured the same number of game elements. We made this choice to increase the study's internal validity based on discussions that the number of game elements might affect gamification's success [39]. In the **OSFA condition**, we used Points, Badges, and Leaderboards (PBL). Although PBL have been criticized, a recent meta-analysis indicates that this combination has positive effects in gamified educational settings [4], while being among the most used ones [21, 68, 77, 81]. Therefore, we considered it a suitable baseline to represent an OSFA gamification approach.

For the **personalized condition**, we relied on recommendations for a multidimensional personalization of gamification [59]. Specifically, the recommendations are given by three decision trees that receive user and contextual information as input and output ratings for different game elements, with each tree indicating one game element. Together, they recommend the top-three most suitable game elements to a user performing some type of learning activity according to the highest rating of each tree. While each tree outputs an independent recommendation to form the top-three, they all receive the same information about the user - preferred game genre (Action, Adventure, RPG, Strategy or Other), preferred playing setting (Multiplayer or Singleplayer), weekly playing time (in hours), gender (Female, Male, Other), educational background (High School, Technical Education, Undergraduate, MsC, Ph.D., Other), and experience researching gamification (Yes or No) - and about the context in which the gamification will be used: country and learning activity type (remembering, understanding, applying, analyzing, evaluating, creating).

With such information, we analyzed the trees to define the personalized gamification design for each participant of the personalized condition. Note that some of the trees' inputs were fixed due to the characteristics of this experiment: the assessments exclusively focused on remembering activities and all participants were Brazilians with no experience in researching gamification. Then, when analyzing the trees with those fixed inputs, we found that the factor affecting the recommendations was a participant's preferred game genre, with the other factors only impacting the recommendations for people from other countries, with experience researching gamification or when completing other learning activity types. Therefore, we derived a simplified personalization algorithm (Algorithm 1) from such analyses, which is shown next and was executed to define the personalized gamification design of each participant of the personalized condition.

ALGORITHM 1: Personalization algorithm derived from [59] given that for all participants: country is Brazil; experience researching gamification is No; and learning activity type is Remembering.

```
Data: preferred game genre
Result: List of game elements forming the personalized gamification design
/* Some elements might differ from the first recommendation from [59] to avoid repeated
   elements and ensure there are three game elements in all gamification designs        */
if preferred game genre = Action then
    /* The second element would originally be Competition. It was replaced by the second
       option - Chance - to comply with the need for three game elements in all conditions */
    return [Competition, Chance, Time Pressure];
else
    /* The third element would originally be Objective. It was replaced by the second option
      - Progression - to comply with the need for three game elements in all conditions  */
    return [Acknowledgment, Objective, Progression];
end
```

Lastly, we call attention to the following aspects. First, in cases where the same element was recommended by different decision trees, we selected the following most suitable one to ensure all gamification designs had three different game elements (Algorithm 1 states this situation). Second, because the decision trees have dozens of nodes and leafs, we do not present them here due to the lack of space to present them in detail and because a partial presentation could mislead their recommendations. Instead, we provide them as supplementary materials and refer to [59] for details. Finally, we highlight one can see images of both system versions in the supplementary material.

3.4 Instruments

For the personalization of gamification, participants completed a pre-survey, which captured information to design the personalized version (i.e., preferred game genre and playing setting, weekly playing time, gender, previous experience researching gamification, education background). For the quantitative data collection, we used the SIMS [22] since it captures participant's motivations to engage with an activity based on four constructs (i.e., intrinsic motivation, identified and external regulations, and amotivation) aligned to the self-determination theory [61]. Furthermore, the SIMS has been empirically validated in the educational context and in several languages, including that of this study's participants [20]. For our data, reliability (i.e. consistency among items of the same construct) was good for intrinsic motivation (0.94), identified regulation (0.82), and amotivation (0.90), but questionable for external regulation (0.51), according to Cronbach's alpha. Additionally, note that despite suitable, we did not use assessments' outcomes as a dependent variable because previous knowledge would play a major role as covariate and because we did not plan ahead of the experiment execution to examine such variable, consequently, we did not capture any measure of previous knowledge. For the qualitative data collection, we conducted semi-structured interviews [6], wherein the interviewer relied on the following questions as the initial source of discussions:

(1) Could you introduce yourself and tell me a little bit about you and your hobbies?
(2) What do you think about games, either digital or analogical?
(3) Overall, what do you think about going to college?
(4) What do you think about the disciplines' assessment activities?
(5) What did you think about the assessment activities you made last week?
   (a) How would you compare those activities to other assessment activities you did before?
(6) What did you think about doing the assessment activity in a system unlike those you regularly use?
   (a) How would you compare that system to others used in classes?
   (b) Would you compare the experience of doing the assessment activities in that system to what other experience?
   (c) After you began the activity, what were your reasons to keep doing it?
(7) What did you think about the system you made the assessment activities?
   (a) Did you note elements uncommon to other systems that might be used to perform similar activities?
   (b) Overall, would you compare that system to another one?
   (c) What did you think about the game elements available in the system?
   (d) How would you compare the experience of doing the assessment activity with and without game elements?
   (e) What changes do you suggest?

Generally, the goal was to understand participants' subjective experiences while performing the assessment activity in terms of how the gamified system affected their motivations. We sought to elicit answers about those aspects especially with items 6-7. However, we began the interview with items 1-5 to capture an overview of who the participants were and their view and overall motivations to go to college and perform assessment activities. The item's generality (e.g., what do you think about) was to reduce bias, leaving for the participant the option of mentioning specific aspects (e.g., *I liked the competition*) early in the interview if relevant to them. This often happened, avoiding the need to explicitly ask subsequent questions (e.g., item 7.c). Interviews (M: 45 minutes) were all conducted by the same person (male, Ph.D. student, 26 years old) using Google Meet. The codebook and quotes supporting the qualitative results are available in the supplementary material. Interviews' transcriptions are not available due to sensitive information.

3.5 Procedure

The study was conducted on consecutive days of the last week of the semester, following six steps. First, learners completed the pre-survey a few weeks before the experiment. Second, they completed the assessment activity of the Programming Techniques discipline during class-time[4] (around one hour). Third, students completed the SIMS right after finishing the first assessment. Fourth, on the second day of execution, participants completed the second assessment activity (Object-Oriented Analysis and Design discipline), also lasting around one hour of class-time. Fifth, they completed the SIMS again, right after the fourth step. The sixth step was participating in a semi-structured interview to talk about their experiences with the gamified system during the assessment activities. Despite all learners were invited, only four of them completed this last step.

3.6 Data Analysis

For the quantitative analysis, given the experimental design (2x2 mixed factorial) and the study goal, Mixed ANOVAs were applied. As the first recommended step of the data analysis [79], we tested the assumption of residuals' (normal) distribution. Since there were violations, we analyzed our data with robust methods (i.e., 20% trimming) [11, 80]. We conducted the robust Mixed ANOVAs and main effect analyses (effect of one independent variable) with the *WRS2* R package [45]. Additionally, we calculated effect sizes through the explanatory measure of Effect Size (ES), which is a robust location measure that handles non-normal data as well as groups with unequal variances; values of 0.1, 0.3, and 0.5 correspond to small, moderate, and large effects, as suggested by the package authors [45]. We executed that process for each motivation regulation type, considering a 0.05 alpha level. As recommended [3, 79], we do not correct p-values because each test concerns a different planned comparison (i.e., a different motivation/regulation), similar to prior research [1, 74]. The dataset and data analysis procedure are available in the supplementary material.

The qualitative analysis aimed to understand learners' experiences in terms of how the gamification influenced their motivations compared to their motivations to perform traditional, non-gamified assessment activities; and, consequently, further explain the quantitative results. Thereby, we adopted the *discourse analysis* analytical framework [14] to enable the understanding of implicit and hidden meanings in participants' answers. We performed a thematic analysis on the semi-structured interviews' transcriptions. Following the thematic analysis procedure [9], one author got familiarized with the data and generated initial codes. Together with other two authors, they searched, reviewed, defined, and named themes to produce a report. During these steps, we adopted an *interpretivist semi-structured approach*[5] that, besides being aligned with our data collection method (i.e., semi-structured interviews), is one of the qualitative approaches most common in HCI research [6].

[4]Due to the COVID-19 pandemic, classes were online - streamed through a synchronous meeting service - despite the course was originally face-to-face.
[5]Interpretivist refers to assuming a subjective view of reality; semi-structured concerns the fact themes will be covered to different extents, depending on the best line of inquiry [6].

Table 3. Descriptive statistics about participants' motivations overall or in either the Programming Techniques (PT) or the Object-Oriented Analysis and Design (OOAD) disciplines. Data are shown as Mean (Standard Deviation), which were collected in a seven-point Likert scale.

			Motivation/Regulation			
	Design	N	Intrinsic	Identified	External	Amotivation
Overall	One-size-fits-all	25	4.20 (1.40)	4.74 (1.23)	**4.65 (1.15)**	**3.35 (1.65)**
Overall	Personalized	20	**5.39 (1.46)**	**5.79 (1.16)**	4.49 (1.26)	2.65 (1.66)
PT	One-size-fits-all	12	4.48 (1.21)	5.00 (1.15)	4.56 (1.23)	**3.08 (1.57)**
PT	Personalized	11	**5.66 (1.06)**	**6.14 (0.68)**	**4.59 (1.43)**	2.64 (1.48)
OOAD	One-size-fits-all	13	3.94 (1.55)	4.50 (1.31)	**4.73 (1.12)**	**3.60 (1.74)**
OOAD	Personalized	09	**5.06 (1.86)**	**5.36 (1.50)**	4.36 (1.08)	2.67 (1.95)

N = Number of participants/responses.
The larger values, when comparing gamification designs, are **bolded**.

For coding, we adopted a mixture of inductive and deductive schemes. We used inductive coding especially in the first iterations of the analysis to understand learners' subject experiences. Although rare in HCI research [6], we used deductive coding in the latter steps to relate low-level codes and themes to motivation- and learning-related theories (e.g., Self-determination Theory, SDT). Hence, the codebook was developed throughout the analysis to allow the identification of emergent themes from interviewees' subjective experiences. To support results' validity, we quote participants' answers, whereas we rely on triangulation to support reliability because the multiple independent coders approach is inappropriate to validate rich subjective analyses [6]. Therefore, we employ methodological and theoretical triangulation. That is, relating qualitative to quantitative data and interpreting qualitative results in terms of multiple theoretical lenses, respectively.

4 RESULTS

Table 3 shows descriptive statistics of the quantitative results. It demonstrates the average and standard deviation, as well as the number of participants (N), for each motivation/regulation in each discipline and overall. Figure 2 presents further descriptive information through boxplots, comparing conditions in both disciplines for each motivation/regulation. Table 4 introduces the results of the statistical tests. Those reveal the main effect of design was significant for intrinsic motivation and identified regulation but nonsignificant for external regulation and amotivation. Differently, the effect of discipline, as well as that of its interaction with design, were nonsignificant for all motivations/regulations. We only conducted further analyses for the significant differences, which reveal large significant effects of the design factor for both intrinsic motivation, $F(1, 21.04) = 8.491$; $p = 0.00829$; ES = 0.64; ES 95% Confidence Interval (CI) [Lower CI; Upper CI] = [0.22; 0.98], and identified regulation, $F(1, 21.92) = 8.4093$; $p = 0.00833$; ES = 0.62; ES 95% CI = [0.21; 0.94]. Figure 3 shows boxplots comparing both design conditions, in terms of intrinsic motivation and identified regulation, regardless of the disciplines.

The qualitative results revealed two main themes, which concern learners' subjective experiences with either traditional (non-gamified) or the experiment's assessment activities. (See Tables 5 and 6 for quotes supporting all tags.) Concerning a *common assessment activity*, interviewees' perceived them as (subthemes): *pressing*, *unsatisfactory*, and *satisfactory*. Interviewees consider those activities as pressing because they might lead to bad results, one needs to perform them to actually learn, and they require significant preparation time. Also, the interviewees showed

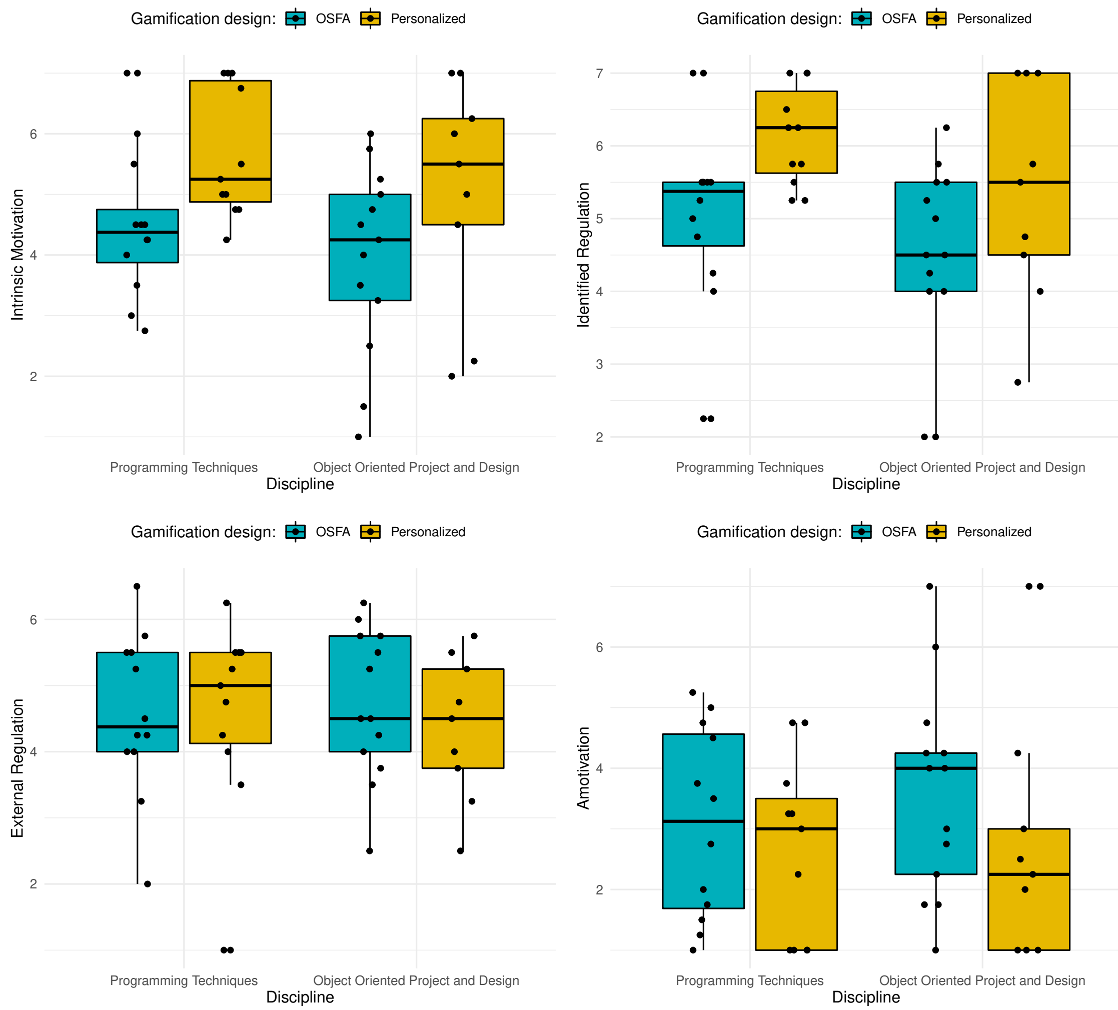

Fig. 2. Boxtplots comparing participants' motivations in both conditions (one-size-fits-all, OSFA, and personalized) for each discipline in which the experiment was executed. Intrinsic motivation, identified regulation, external regulation, and amotivation are shown on top-left, top-right, bottom-left, and bottom-right, respectively.

dissatisfaction because such activities are not the best evaluation method and obligatory. Differently, they felt satisfaction in terms of improving knowledge, having the chance for self-assessment, and gaining grades.

Concerning the *experiment's assessment activities*, interviewees' perceive them as (subthemes) *need-supporting*, *enjoyable*, and *unsatisfactory* at some extent, while also noting its *gamification*[6]. The need-supporting perception comes from interviewees feeling autonomous, competent, and related. The enjoyable subtheme emerged because the activity was something free of risk, relaxing, and novel. Also, the activity allowed self-assessment. On the other hand, interviewees felt dissatisfaction regarding some system bugs. Furthermore, they noted the gamification, which

[6]We named the subtheme as *personalized gamification* because all interviewees were from the personalized condition.

Table 4. Results of the robust two-way ANOVAs for different Intrinsic Motivation (IM), Identified Regulation (IR), External Regulation (ER), and Amotivation (AM) as Dependent Variables (DV) and gamification design (one-size-fits-all or personalized) and discipline (Programming Techniques and Object-Oriented Analysis and Design) as factors.

	Factor					
	Design		Discipline		Design:Discipline	
DV	F(df1, df2)	P-value	F(df1, df2)	P-value	F(df1, df2)	P-value
IM	5.2533(1, 11.3193)	**0.0420**	0.2657(1, 9.8951)	0.6175	0.0160(1, 11.3193)	0.9015
IR	6.3288(1, 9.7345)	**0.0312**	0.9237(1, 8.3351)	0.3635	0.0887(1, 9.7345)	0.7721
ER	0.0302(1, 11.5548)	0.8650	0.2175(1, 12.8690)	0.6487	0.5773(1, 11.5548)	0.4626
AM	1.8524(1, 12.7829)	0.1970	0.0217(1, 12.8194)	0.8852	0.2223(1, 12.7829)	0.6453

Significant p-values (< 0.05) in **bold**.

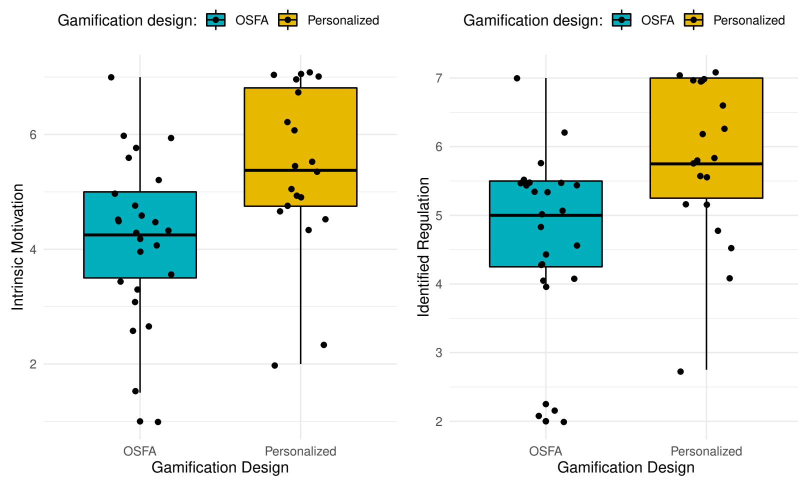

Fig. 3. Boxplots' comparing participants' overall intrinsic motivation (left) and identified regulation (right) among experimental conditions (i.e., one-size-fits-all, OSFA, and personalized).

they considered better than not having it and suitable to their preferences while noting some missing game elements. Additionally, interviewees perceived the gamification as motivating because it made them want to perform well and motivated social comparison, as well as supported basic psychological needs. Figure 4 shows the thematic map linking themes and subthemes, with each subtheme's tag shown within its box and different colors for distinct kinds of experiences.

5 DISCUSSION

This section discusses our results, reasons for why they were positive, why they differ from previous research, findings' implications, and study limitations.

Table 5. Quotes supporting the *Common assessment activity* theme found in the thematic analysis of semi-structured interviews. I*N* refers to Interviewee N (1-4).

Tag	Subtheme	Quote
Demands significant preparing	Pressing	*Because sometimes there is a test that has a lot of burden to study for that specific subject*, I1
Necessary to actually learn	Pressing	*I do [assessment activities] mainly because I have to learn*, I4
Might lead to bad results	Pressing	*So I want to do well [in the assessment] to prove that I really absorbed that [knowledge]*, I3
Is not the best method	Unsatisfactory	*... sometimes it doesn't test all of your content retention and everything*, I1
Obligatory	Unsatisfactory	*It is an obligation. Total obligation. In the real sense of the word, is to do it because otherwise you will not pass*, I1
Improve knowledge	Satisfactory	*But generally, when there is a test I feel good because I try to study hard and dedicate myself to doing well in the test, and also learning the content*, I2
Self-assessment	Satisfactory	*I feel good doing the activities..even because I have to do it to be graded, but I feel good doing because I will be able to follow my evolution*, I2
Gain grade	Satisfactory	*Assessment activity when you think, you will think about the return, right, the grade that you can take and take advantage of it*, I3

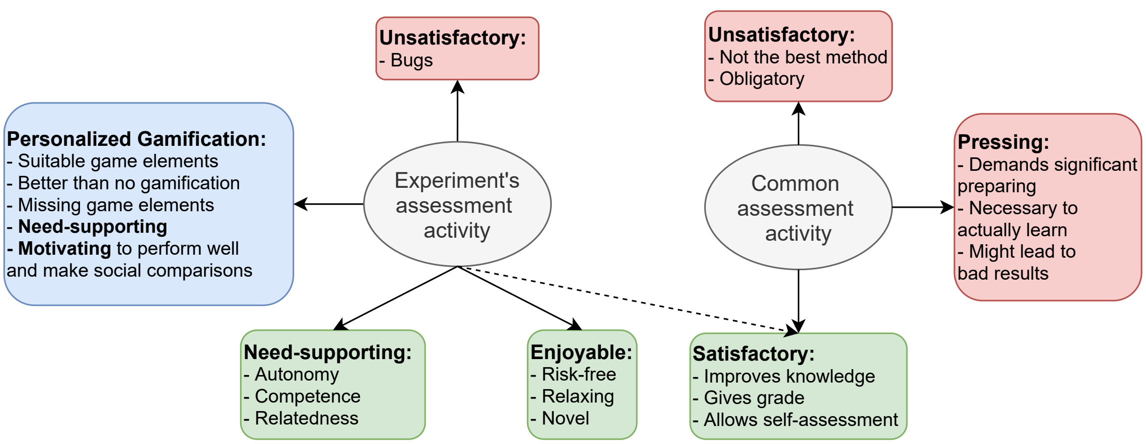

Fig. 4. Thematic map from the analysis of the semi-structured interviews showing learners' subjective experiences with assessment activities. Ovals represent the main themes, rectangles represent the subthemes, bolded text indicates subthemes' names, and regular text indicates tags. Red and green indicate negative and positive experiences, respectively. Solid connections from X to Y means X is/has Y and all of its tags; the dashed line means the connection is partial.

5.1 Overall Findings

Our quantitative findings show a large, significant difference in the perception of participants who used the personalized gamification design, compared to reports of those who used the OSFA gamification, in terms of intrinsic motivation and identified regulation. This significant difference supports the hypothesis that offering users a gamification design

Table 6. Quotes supporting *Experiment's assessment activity* theme found in the thematic analysis of semi-structured interviews. I*N* refers to Interviewee N (1-4).

Tag	Subtheme	Quote
Allows self-assessment	Satisfactory	*[...] I could see that, for example, there was something I didn't learn well*, I2
Autonomy	Need-supporting	*I felt it was a little [...] out of class [...] because I think it was not an obligation*, I3
Competence	Need-supporting	*I think doing the activity was much more satisfying [...] because I have real-time feedback*, I3
Relatedness	Need-supporting	*[...] the little rank is just to make fun of each other when doing the activity*, I1
Bugs	Unsatisfactory	*Then an error goes to the screen and it broke.*, I4
Risk-free	Enjoyable	*[...] when you take it and even when you miss a question like that, [...] it feels lighter, I would tell you*, I1
Relaxing	Enjoyable	*[the activity was like] a form of relaxation and occupying part of a gap between something at work and another*, I1
Novel	Enjoyable	*So I think that compared to other activities of the day, I think it was something different [...] maybe because I don't do it often, I think it was a really cool thing to do ....*, I3
Suitable game elements	Personalized gamification	*If I were to talk about the game aspect, this [competition - the weekly leaderboard] would be the type of game that I like, interestingly*, I3
Better than no gamification	Personalized gamification	*with game element is much better*, I2; *[with gamification] It didn't get boring, right? you don't get tired*, I4
Missing game elements	Personalized gamification	*Eagle would be good if you have a progress bar*, I2; *It would be nice to show the medal maybe on the profile or display next to the name*, I3
To perform well	Motivating	*when I was there doing a question, I was very careful 'damn it, I have to get this one right [...]*, I1
To make social comparisons	Motivating	*I can compete with that ranking, with the people who also performed the activity*, I3

personalized to multiple characteristics (e.g., information of both users and task at hand - learning activity) is more effective than providing a single, general design for all. The magnitude of that difference might be interpreted from two perspectives. First, gamification applied to education is associated with overall small effects on motivational learning outcomes when compared to no gamification [62]. Second, the hinge-point for effect sizes of general educational interventions is a *moderate* effect (Cohen's d > 0.4 [13]) according to [29]. Our findings suggest a large effect size above the average in both perspectives. Therefore, indicating the multidimensional personalization of gamification employed in this study represented an improvement over the OSFA approach that is above both the common gamification effects on motivation as well as the threshold for educational interventions.

Additionally, we found that completing the task with personalized gamification mitigated negative perceptions of common assessment activities while motivating and supporting the learners. From semi-structured interviews with participants that used the personalized gamification design, we found they acknowledge assessment activities are valuable to the learning process but perceived them as pressing and unsatisfactory. The effectiveness of the personalized gamification is suggested because participants considered the game elements available to them suitable to their preferences, as well as mentioned the personalized gamification supported basic psychological needs. Also, when using the personalized gamification, participants felt the activity was less pressing and more relaxing, compared to common assessments (see Tables 5 and 6), possibly because the gamification mitigated some of those negative feelings (e.g., of pressure), supporting basic psychological needs. Furthermore, the personalized gamification motivated them to perform well in the assessments as well as to make social comparisons, indicating its educational value.

Summarizing those findings, we have the following takeaways:

- Multidimensional personalization improved students' experiences with the gamified educational system, compared to the OSFA approach, in a degree above the hinge-point for educational interventions and OSFA gamification when compared to no gamification;
- The personalized design overcame the OSFA approach, in terms of autonomous motivation, by providing game elements suitable to learners' preferences that supported their needs and mitigated drawbacks from regular assessment activities.

5.2 Why Personalization Worked?

We discuss our findings in terms of why learners using the personalized gamification felt more motivated than those using the OSFA design from two perspectives. First, *preference satisfaction*. Previous research has advocated that one size does not fit all [47]. For instance, there is evidence showing different users have distinct preferences [24, 49, 73] and that gamification's effect varies from user to user [58, 75]. Hence, supporting the argument regards the suitability of the game elements offered to them. Second, *psychological needs support*. That is, offering users the right game elements supported their basic psychological needs. Consequently, improving their autonomous motivation compared to participants that received the OSFA design. Accordingly, the non-significant effect on participants' external regulation might be because the game elements were suitable to their preferences, instead of being perceived as external drivers. Note that while the identified regulation belongs to the extrinsic perspective, it is the closest regulation to intrinsic motivation, which together are seen as autonomous motivation [61, 76]. In contrast, the non-significant effect on amotivation might be attributed to the importance of the experimental assessment activities, which probably led to low amotivation levels for all.

5.3 Why Previous Personalization Strategies did not Work?

Regarding why previous empirical studies comparing personalized and OSFA designs failed to provide conclusive results, unlike this study, we discuss three perspectives.

First, the degree of personalization. While most prior research used a single personalization criterion [46, 48], the literature demonstrates that users' motivations/experiences differ in many factors (e.g., user types [24, 73], task familiarity/knowledge [59, 63], and gender [31], among others - see [32, 60] for reviews). Accordingly, a study that applied a dual personalization suggests it overcame the single dimension personalization - with random gamification as the baseline [67]. Similarly, this study employed a multidimensional approach but considered eight values as input for the personalization of gamification. Hence, the multidimensional strategy might be the explanation for this study finding a large, positive effect of personalization of gamification when compared to the OSFA approach, unlike prior research. Note that despite Mora et al. discussed their non-significant results might be a product of low statistical power [46], their sample size was larger than the one of this study, although more unbalanced. Then, if they had found an effect of similar magnitude, statistical significance would likely be found as well.

Second, considering the context. Whereas research has highlighted contextual information affects gamification's success [24], arguing factors such as the task to be done should be considered by personalization strategies [25, 56], most of the related research only focused on user information [46, 48]. The somewhat exception is Hajarian et al. that, differently, found that the personalized gamification overcame the OSFA design [23]. In that study, the personalization is driven by user information as well (their likes about the game elements), but such preferences were captured in one context, doing one activity (i.e., using a social network to arrange dates) and, then, the authors used those insights to personalize gamification for the same context and task. In our study, we used a personalization strategy focused on the educational context [59], unlike those used by research reporting inconclusive findings. Additionally, guidance from that strategy considers the type of the task/activity to be done, which was a remembering learning task in the case of this study. Therefore, it might be that considering contextual information (the domain and the task in our case) is the reason for this study and that of Hajarian et al. [23] finding positive effects from the personalization of gamification.

Third, approximating to what customization provides. The indications that customization of gamification is effective when compared to the OSFA approach (e.g., [42, 66]) support the rationales discussed previously, at least to some extent. That is because multiple information from the user and the context are considered, simultaneously, while one is customizing their gamification design. For instance, one's user type, gender, gaming preferences and habits, etc., as well as the domain and task, are taken into accounting when they are selecting the game elements they want to be available. Consequently, the more information the personalization strategy receives, the more its outcomes are expected to approximate to those of customization; similarly for considering contextual factors. Based on that relation, the lack of considering multiple factors simultaneously, including contextual ones, might be the reason for previous studies finding positive effects for customization of gamification but not for its personalization.

5.4 Summary of Implications

Based on our findings, we derive two main implications:

(1) *Multidimensional personalization of gamification improves students' motivations in assessment activities when compared to the OSFA approach.* Because motivation, especially autonomous, is among the most important factors for learning [27, 55, 57, 76, 78], this practical implication provides valuable information for designers of gamified educational systems. Then, practitioners can follow the strategy used in our experiment to personalize gamified educational systems, which is likely to improve students' experiences and, according to the relationship between autonomous motivation and learning, improve students' overall learning. Consequently, future research needs to ground and expand the understanding of the effects of gamification personalized to multiple information. Therefore, we call for future research to run similar experiments with larger, varied samples and in other contexts (e.g., with other learning activity types and subjects). Those would increase the understanding of how the multidimensional personalization's benefits we found generalize.

(2) *If one wants to personalize, they cannot oversimplify it.* Because much research on personalization of gamification has been conducted [32], but with little evidence on its effectiveness compared to the OSFA approach [60], this implication is of special value to researchers. In this study, we hypothesized that the unexpected results of prior research comparing personalized and OSFA gamification were due to personalizing to a single user characteristic. While our positive findings might be due to the multidimensional personalization, considering the task to be done is a factor as well. In that context, we also discussed how the increased complexity of the multidimensional strategy approximated it to some of the benefits of customization of gamification. Accordingly, we need further research to better understand how such strategies compare to each other and what are the main reasons for the differences. Thus, we call for future research to experimentally compare tailoring strategies,

such as single-dimension personalization, multidimensional personalization, and customization, as well as to investigate which are the most relevant criteria for their effectiveness when applied to users.

5.5 Limitations and Recommendations to Future Research

Some limitations need to be considered when interpreting our findings. First, about our sample. Subjects of our sample had already used the system before but with different game elements, and that might have affected their experiences during this study. To mitigate that risk, we created a new account for each user, aiming to ensure they would have a baseline for when the study began, and highlighted they should complete the motivation measure based on their experiences during the specific assessment activities. Furthermore, twenty-six learners from the same period participated in this study; a total of 45 data points considering both experiment days. This sample is similar to those seen in overall HCI studies [11] and CHI papers [10]. While that does not set off the small sample size limitations, it shows conducting large-scale studies in this field is challenging and expensive and reflects that small-sampled studies can also contribute to the literature. Note that our study was conducted in an ecological context (i.e., classroom assessment activities), which increases its validity and further supports the difficulty of recruiting large samples.

Also, all participants were males. Overall, around 20% of first-year students of computing-related courses are females [64, 65] - similarly, the rate is 14% in Brazil [17] - and such courses are known to have a high drop-out rate, which was 26% for first-years in Brazil in 2019 [8]. Therefore, while the convenience sampling limits our findings' generalization towards an overall learning domain, the impact of not having female subjects in our study is mitigated by the fact that the distribution is similar to that of similar courses. Thus, we advocate that the context above, along with the increased ecological validity by working within a real learning context, leads to a positive trade-off compared to an increased external validity due to a large sample recruited through, for instance, a crowdsourcing platform in which participants would complete a possibly meaningless task. To cope with generalization limitations, we call for replications with larger and varied samples to further ground our findings.

Second, about the study design. Because all participants were from the same class, most participants possibly discussed about having game elements that were not available to others. Those differences were unavoidable due to the personalization of gamification and participants were aware they would exist. While that might have affected their experiences, our findings suggest personalization worked overall, regardless of some participants wanting more game elements. Additionally, subjects did not know in which condition they were, preventing the hypothesis guessing threat. Moreover, our findings are based on two measurements, which were performed in subsequent days, and suggest that personalized gamification's effectiveness did not change from one day to another. However, we cannot ensure it would continue to work in the long run. Because the novelty effect is often considered to influence the effectiveness of OSFA gamification, it might also affect personalized gamification, especially considering the novelty of the approach was suggested by our qualitative findings. On one hand, students complete assessment activities like this experiment's ones on few occasions (e.g., once or twice per term), similar to our approach, which reduces the impact of that limitation. On the other hand, research assessing personalized gamification's effect based on more applications remains needed. For instance, applying it every semester during a class's full stay in college (e.g., 4 years) given that assessments such as the ones explored in this study are not frequent.

Third, about the instruments. To the multidimensional personalization, we followed the recommendations from [59]. Despite build from statistical analyses of survey data, similar to other recommendations [32], the authors acknowledge the recommendations demand empirical validation, such as what we did in this article, because it is based on user preferences. While this might be seen as a limitation, the only way to validate those recommendations is through research such as this one, which yielded encouraging results towards the recommendations' validity. Whereas we selected Eagle-Edu by convenience, its flexibility is valuable for this kind of research. The system is in its beta version, however, and a few bugs likely affected participants' experiences. Because all subjects used the same system, those bugs likely did not differ among experimental versions, consequently, having little impact on our findings. The external regulation construct from our quantitative results showed a *questionable* reliability. This might be related to some limitations the SIMS have in many of its versions [20, 22]. While that problem might not be so pertinent to HCI studies, our quantitative data analysis helped to handle it by controlling for between participants variation [11].

Lastly, about the data and its analysis. First, our dataset might suffer from hypothesis-guessing because some participants' answers were, for instance, all sevens and ones (maximum and minimum choices). While such patterns might be perceived as unreliable, the goal of Likert scales is to capture people's self-reports of their experiences while removing their ambiguities. Accordingly, one cannot ensure whether someone has experienced that motivation level or if they carelessly selected that option due to, e.g., hypothesis-guessing. Therefore, we chose not to systematically inspect or remove data based on those patterns following similar studies [2, 40, 46, 66, 67, 74] while we complemented our dataset with qualitative data from semi-structured interviews to enrich our findings [15]. Second, our sample size is limited to 45 data points, which is associated with low statistical power and often large confidence intervals. As discussed in [12], while analyzing small samples might lead to overestimation of effect sizes, such situations often happen along with p-hacking and publication bias. To cope with that thread, we followed recommendations from the same study by adhering to data and material transparency, which encourages replication and clear reporting - including effect sizes' confidence intervals. We also limited our analyses to the four constructs SIMS measures and limited further testing to significant effects. Additionally, we highlight gamification studies have not suffered from publication bias [62]. Thus, mitigating the limitations from our sample size on conclusion validity by following guidelines to address the replication crisis. Third, a single author coded the semi-structured interviews. Having multiple independent coders to ensure validity is inadequate for our *interpretivist semi-structured approach*, while relying on multiple coders to achieve complementary views would be suitable. To mitigate that limitation, other researchers worked during subsequent analysis steps (e.g., gathering and reviewing themes) and reviewed the codebook to ensure findings were supported by the interviews' data. Lastly, while having four interviewees is within HCI common practice, none of them was from the OSFA condition, which limits our qualitative findings and points to the need for more mixed-methods and qualitative studies to examine users' subjective experiences with personalized gamification that despite elementary for HCI research, are missing in the personalized gamification field [60].

6 CONCLUSION

Personalization of gamification has attracted researchers' and practitioners' interest as the literature highlights the limitations of the OSFA - one-size-fits-all - approach. However, there is a lack of empirical evidence on how those approaches compare and the few studies comparing them personalized gamification to a single user characteristic. This study faces that gap with a mixed-methods experimental study comparing learners' motivations when performing two assessment activities, on different days, with either the OSFA approach or a gamification design personalized to multiple user and contextual information. The quantitative results showed learners who used the personalized design felt higher levels of intrinsic motivation and identified regulation compared to those who used the OSFA alternative; differences in the external regulation and amotivation were nonsignificant. Additionally, the results suggest personalization's effect did not change from the first to the second assessment. Accordingly, the qualitative findings revealed the students who used the personalized gamification considered that the game elements were suitable to their preferences and that they motivated them to perform well and make social comparisons, besides supporting their basic psychological needs.

From those findings, we derive two main implications. First, to the design of gamified educational systems. Our empirical evidence suggests that gamifying educational systems with the strategy for multidimensional personalization we used can improve students' autonomous motivation (intrinsic and identified) compared to using the standard, OSFA approach. Thus, given that increased motivation is related to learning gains, applying multidimensional personalization is likely to enhance students' learning more than OSFA gamification. Second, to research on personalized gamification. Our findings provide promising evidence that multidimensional personalization can improve OSFA gamification, a result that has not been found by other studies personalizing gamification using a single dimension. Thereby, contributing indication that personalizing to a single criterion might explain why related research found inconclusive results. Thus, suggesting the need for more complex personalization strategies such as the one used in this paper, and providing an initial empirical validation of that strategy. Nevertheless, such findings must be interpreted with caution mainly because of the study's limited sample size and the experiment being limited to remembering learning activities. Therefore, the need for replications, which can rely on our open materials for planning, execution, and data analysis.

ACKNOWLEDGMENTS

This research was partially funded by the Brazilian National Council for Scientific and Technological Development (CNPq - 141859/2019-9; 308395/2020-4), Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES - Finance code 001), São Paulo Research Foundation (FAPESP - 2018/15917-0; 2018/07688-1; 2020/02801-4), and Santa Catarina State Research and Innovation Support Foundation (FAPESC - T.O. No.: 2017TR1755). Additionally, the authors are thankful to SENAI Londrina, Brazil, for enabling this research.

REFERENCES

[1] Maximilian Altmeyer, Pascal Lessel, Subhashini Jantwal, Linda Muller, Florian Daiber, and Antonio Krüger. 2021. Potential and effects of personalizing gameful fitness applications using behavior change intentions and Hexad user types. *User Model User-Adap Inter* (2021). https://doi.org/10.1007/s11257-021-09288-6

[2] Maximilian Altmeyer, Gustavo F Tondello, Antonio Krüger, and Lennart E Nacke. 2020. HexArcade: Predicting Hexad User Types By Using Gameful Applications. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 219–230.

[3] Richard A Armstrong. 2014. When to use the B onferroni correction. *Ophthalmic and Physiological Optics* 34, 5 (2014), 502–508.

[4] Shurui Bai, Khe Foon Hew, and Biyun Huang. 2020. Is gamification "bullshit"? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review* (2020), 100322.

[5] Johan Baldeón, Inmaculada Rodríguez, and Anna Puig. 2016. LEGA: A LEarner-centered GAmification Design Framework. In *Proceedings of the XVII International Conference on Human Computer Interaction* (Salamanca, Spain) *(Interacción '16)*. ACM, New York, NY, USA, Article 45, 8 pages. https://doi.org/10.1145/2998626.2998673

[6] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. Qualitative HCI research: Going behind the scenes. *Synthesis lectures on human-centered informatics* 9, 1 (2016), 1–115.

[7] Klaudia Bovermann and Theo J Bastiaens. 2020. Towards a motivational design? Connecting gamification user types and online learning activities. *Research and Practice in Technology Enhanced Learning* 15, 1 (2020), 1–18. https://doi.org/10.1186/s41039-019-0121-4

[8] Brasscom. 2019. *Estudo Brasscom – Formação Educacional e Empregabilidade em TIC (Brasscom Study - Educational Training and Employability in ICT)*. https://brasscom.org.br/estudo-brasscom-formacao-educacional-e-empregabilidade-em-tic-achados-e-recomendacoes/

[9] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[10] Kelly Caine. 2016. Local standards for sample size at CHI. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 981–992.

[11] Paul Cairns. 2019. *Doing better statistics in human-computer interaction*. Cambridge University Press.

[12] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a replication crisis in empirical computer science. *Commun. ACM* 63, 8 (2020), 70–79.

[13] Sarah Cohen, Werner Nutt, and Yehoshua Sagic. 2007. Deciding equivalances among conjunctive aggregate queries. *J. ACM* 54, 2, Article 5 (April 2007), 50 pages. https://doi.org/10.1145/1219092.1219093

[14] Adrian Coyle. 1995. Discourse analysis. *Research Methods in Psychology. London* (1995).

[15] John W Creswell and J David Creswell. 2017. *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage publications.

[16] Lamya F Daghestani, Lamiaa F Ibrahim, Reem S Al-Towirgi, and Hesham A Salman. 2020. Adapting gamified learning systems using educational data mining techniques. *Computer Applications in Engineering Education* 28, 3 (2020), 568–589.

[17] Sociedade Brasileira de Computação (SBC Brazilian Computing Society). 2019. *Higher Computing Education - Statistics 2019*. https://www.sbc.org.br/documentos-da-sbc/summary/133-estatisticas/1324-educacao-superior-em-computacao-estatisticas-2019

[18] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. ACM, 9–15. https://doi.org/10.1145/2181037.2181040

[19] Christo Dichev and Darina Dicheva. 2017. Gamifying education: what is known, what is believed and what remains uncertain: a critical review. *International journal of educational technology in higher education* 14, 1 (2017), 9.

[20] Vítor Gamboa, Sandra T Valadas, and Olímpio Paixão. 2013. Validação da versão portuguesa da Situational Motivation Scale (SIMS) em contextos académicos. In *Atas do XII Congresso Galego-Português de Psicopedagogia, 11-13 de setembro de 2013*. Universidade do Minho, 4868–4882.

[21] MN Gari, GS Walia, and AD Radermacher. 2018. Gamification in computer science education: A systematic literature review. In *American Society for Engineering Education*.

[22] Frédéric Guay, Robert J Vallerand, and Céline Blanchard. 2000. On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and emotion* 24, 3 (2000), 175–213.

[23] Mohammad Hajarian, Azam Bastanfard, Javad Mohammadzadeh, and Madjid Khalilian. 2019. A personalized gamification method for increasing user engagement in social networks. *Social Network Analysis and Mining* 9, 1 (2019), 47.

[24] Stuart Hallifax, Serna Audrey, Marty Jean-Charles, Lavoué Guillaume, and Lavoué Elise. 2019. Factors to Consider for Tailored Gamification. In *CHI Play*. ACM, Barcelona, Spain, 559–572. https://hal.archives-ouvertes.fr/hal-02185647

[25] Stuart Hallifax, Audrey Serna, Jean-Charles Marty, and Élise Lavoué. 2019. Adaptive Gamification in Education: A Literature Review of Current Trends and Developments. In *Transforming Learning with Meaningful Technologies*, Maren Scheffel, Julien Broisin, Viktoria Pammer-Schindler, Andri Ioannou, and Jan Schneider (Eds.). Springer International Publishing, Cham, 294–307. https://doi.org/10.1007/978-3-030-29736-7_22

[26] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work?-A Literature Review of Empirical Studies on Gamification.. In *HICSS*, Vol. 14. 3025–3034.

[27] Michael D. Hanus and Jesse Fox. 2015. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education* 80 (2015), 152 – 161. https://doi.org/10.1016/j.compedu.2014.08.019

[28] Muhammad Awais Hassan, Ume Habiba, Fiaz Majeed, and Muhammad Shoaib. 2019. Adaptive gamification in e-learning based on students' learning styles. *Interactive Learning Environments* (2019), 1–21.

[29] John Hattie. 2012. *Visible learning for teachers: Maximizing impact on learning*. Routledge.

[30] Karl M Kapp. 2012. *The gamification of learning and instruction*. Wiley San Francisco.

[31] Ana Carolina Tomé Klock, Isabela Gasparini, and Marcelo Soares Pimenta. 2019. Designing, Developing and Evaluating Gamification: An Overview and Conceptual Approach. In *Data Analytics Approaches in Educational Games and Gamification Systems*. Springer, 227–246.

[32] Ana Carolina Tomé Klock, Isabela Gasparini, Marcelo Soares Pimenta, and Juho Hamari. 2020. Tailored gamification: A review of literature. *International Journal of Human-Computer Studies* (2020), 102495.

[33] Ana Carolina Tomé Klock, Aline Nunes Ogawa, Isabela Gasparini, and Marcelo Soares Pimenta. 2018. Does gamification matter?: a systematic mapping about the evaluation of gamification in educational environments. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. ACM, 2006–2012.

[34] Selay Arkn Kocadere and Şeyma Çağlar. 2015. The design and implementation of a gamified assessment. *Journal of e-Learning and Knowledge Society* 11, 3 (2015).

[35] Jonna Koivisto and Juho Hamari. 2019. The rise of motivational information systems: A review of gamification research. *International Journal of Information Management* 45 (2019), 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

[36] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218. https://doi.org/10.1207/s15430421tip4104_2

[37] Richard N Landers. 2014. Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming* 45, 6 (2014), 752–768.

[38] Richard N Landers, Elena M Auer, Andrew B Collmus, and Michael B Armstrong. 2018. Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming* 49, 3 (2018), 315–337.

[39] Richard N. Landers, Kristina N. Bauer, Rachel C. Callan, and Michael B. Armstrong. 2015. *Psychological Theory and the Gamification of Learning*. Springer International Publishing, Cham, 165–186. https://doi.org/10.1007/978-3-319-10208-5_9

[40] Elise Lavoué, Baptiste Monterrat, Michel Desmarais, and Sébastien George. 2018. Adaptive gamification for learning environments. *IEEE Transactions on Learning Technologies* 12, 1 (2018), 16–28.

[41] Joey J Lee and Jessica Hammer. 2011. Gamification in education: What, how, why bother? *Academic exchange quarterly* 15, 2 (2011), 146.

[42] Pascal Lessel, Maximilian Altmeyer, Marc Müller, Christian Wolff, and Antonio Krüger. 2017. Measuring the effect of" bottom-up" gamification in a microtask setting. In *Proceedings of the 21st International Academic Mindtrek Conference*. 63–72.

[43] De Liu, Radhika Santhanam, and Jane Webster. 2017. Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS quarterly* 41, 4 (2017), 1011–1034.

[44] Kevin Loughrey and Daire Broin. 2018. Are We Having Fun Yet? Misapplying Motivation to Gamification. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 1–9. https://doi.org/10.1109/GEM.2018.8516535

[45] Patrick Mair and Rand Wilcox. 2018. Robust statistical methods using WRS2. *The WRS2 Package* (2018).

[46] A. Mora, G. F. Tondello, L. E. Nacke, and J. Arnedo-Moreno. 2018. Effect of personalized gameful design on student engagement. In *2018 IEEE Global Engineering Education Conference (EDUCON)*. 1925–1933. https://doi.org/10.1109/EDUCON.2018.8363471

[47] Lennart E Nacke and Christoph Sebastian Deterding. 2017. The maturing of gamification research. *Computers in Human Behaviour* (2017), 450–454.

[48] Wilk Oliveira, Armando Toda, Paula Toledo, Lei Shi, Julita Vassileva, Ig Ibert Bittencourt, and Seiji Isotani. 2020. Does Tailoring Gamified Educational Systems Matter? The Impact on Students' Flow Experience. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*. ScholarSpace, 1226–1235.

[49] Rita Orji, Gustavo F Tondello, and Lennart E Nacke. 2018. Personalizing persuasive strategies in gameful systems to gamification user types. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 435. https://doi.org/10.1145/3173574.3174009

[50] Kiemute Oyibo, Rita Orji, and Julita Vassileva. 2017. The influence of culture in the effect of age and gender on social influence in persuasive technology. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*. 47–52.

[51] Kiemute Oyibo, Rita Orji, and Julita Vassileva. 2017. Investigation of the Persuasiveness of Social Influence in Persuasive Technology and the Effect of Age and Gender.. In *PPT@ PERSUASIVE*. 32–44.

[52] Kiemute Oyibo, Rita Orji, and Julita Vassileva. 2017. Investigation of the social predictors of competitive behavior and the moderating effect of culture. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization*. 419–424.

[53] Paula Palomino, Armando Toda, Luiz Rodrigues, Wilk Oliveira, and Seiji Isotani. 2020. From the Lack of Engagement to Motivation: Gamification Strategies to Enhance Users Learning Experiences. In *19th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)-GranDGames BR Forum*. 1127–1130.

[54] Paula T Palomino, Armando M Toda, Wilk Oliveira, Luiz Rodrigues, and Seiji Isotani. 2019. Gamification journey: A Novel approach for classifying gamer types for gamified educational systems. *Proceedings of SBGames* (2019), 165–173.

[55] Paul R Pintrich. 2003. A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of educational Psychology* 95, 4 (2003), 667.

[56] Luiz Rodrigues, Wilk Oliveira, Armando Toda, Paula Palomino, and Seiji Isotani. 2019. Thinking Inside the Box: How to Tailor Gamified Educational Systems Based on Learning Activities Types. In *Proceedings of the Brazilian Symposium of Computers on Education*. SBC, 823–832. https://doi.org/10.5753/cbie.sbie.2019.823

[57] Luiz Rodrigues, Armando M Toda, Wilk Oliveira, Paula T Palomino, Anderson Paulo Avila-Santos, and Seiji Isotani. 2021. Gamification Works, but How and to Whom? An Experimental Study in the Context of Programming Lessons. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*. 184–190.

[58] Luiz Rodrigues, Armando M Toda, Wilk Oliveira, Paula T Palomino, and Seiji Isotani. 2020. Just beat it: Exploring the influences of competition and task-related factors in gamified learning environments. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*. SBC, 461–470.

[59] Luiz Rodrigues, Armando M. Toda, Wilk Oliveira, Paula T. Palomino, Julita Vassileva, and Seiji Isotani. 2021. Automating Gamification Personalization: To the User and Beyond. arXiv:2101.05718 [cs.HC]

[60] Luiz Rodrigues, Armando M. Toda, Paula T. Palomino, Wilk Oliveira, and Seiji Isotani. 2020. Personalized gamification: A literature review of outcomes, experiments, and approaches. In *Proceedings of the 8th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2020) (Salamanca, Spain, October 21-23, 2020)*, F. J. García-Peñalvo (Ed.).

[61] Richard M Ryan and Edward L Deci. 2017. *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Publications.

[62] Michael Sailer and Lisa Homner. 2020. The Gamification of Learning: a Meta-analysis. *Educ Psychol Rev* 32 (2020), 77–112. https://doi.org/10.1007/s10648-019-09498-w

[63] Diana R. Sanchez, Markus Langer, and Rupinder Kaur. 2020. Gamification in the classroom: Examining the impact of gamified quizzes on student learning. *Computers & Education* 144 (2020), 103666. https://doi.org/10.1016/j.compedu.2019.103666

[64] Sharon Sassler, Katherine Michelmore, and Kristin Smith. 2017. A tale of two majors: explaining the gender gap in stem employment among computer science and engineering degree holders. *Social Sciences* 6, 3 (2017), 69.

[65] Christian Schindler and Matthias Müller. 2019. Gender gap? a snapshot of a bachelor computer science course at Graz University of Technology. In *Proceedings of the 13th European Conference on Software Architecture-Volume 2*. 100–104.

[66] Marc Schubhan, Maximilian Altmeyer, Dominic Buchheit, and Pascal Lessel. 2020. Investigating User-Created Gamification in an Image Tagging Task. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[67] Hallifax Stuart, Elise Lavoué, and Audrey Serna. 2020. To tailor or not to tailor gamification? An analysis of the impact of tailored game elements on learners' behaviours and motivation. In *21th International Conference on Artificial Intelligence in Education*.

[68] Sujit Subhash and Elizabeth A Cudney. 2018. Gamified learning in higher education: A systematic review of the literature. *Computers in Human Behavior* 87 (2018), 192–206.

[69] Armando Toda, Wilk Oliveira, Lei Shi, Ig Bittencourt, Seiji Isotani, and Alexandra Cristea. 2019. Planning Gamification Strategies based on User Characteristics and DM: A Gender-based Case Study. In *12th International Conference on Educational Data Mining*. 438–443.

[70] Armando M Toda, Ana CT Klock, Wilk Oliveira, Paula T Palomino, Luiz Rodrigues, Lei Shi, Ig Bittencourt, Isabela Gasparini, Seiji Isotani, and Alexandra I Cristea. 2019. Analysing gamification elements in educational environments using an existing Gamification taxonomy. *Smart Learning Environments* 6, 1 (2019), 16. https://doi.org/10.1186/s40561-019-0106-1

[71] Armando M. Toda, Pedro H. D. Valle, and Seiji Isotani. 2018. The Dark Side of Gamification: An Overview of Negative Effects of Gamification in Education. In *Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions*, Alexandra Ioana Cristea, Ig Ibert Bittencourt, and Fernanda Lima (Eds.). Springer International Publishing, Cham, 143–156. https://doi.org/10.1007/978-3-319-97934-2_9

[72] Gustavo Fortes Tondello. 2019. *Dynamic Personalization of Gameful Interactive Systems*. Ph.D. Dissertation. University of Waterloo.

[73] Gustavo F Tondello, Alberto Mora, and Lennart E Nacke. 2017. Elements of gameful design emerging from user preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, 129–142. https://doi.org/10.1145/3116595.3116627

[74] Gustavo F Tondello and Lennart E Nacke. 2020. Validation of User Preferences and Effects of Personalized Gamification on Task Performance. *Frontiers in Computer Science* 2 (2020), 29.

[75] Rob Van Roy and Bieke Zaman. 2018. Need-supporting gamification in education: An assessment of motivational effects over time. *Computers & Education* 127 (2018), 283–297. https://doi.org/10.1016/j.compedu.2018.08.018

[76] Maarten Vansteenkiste, Eline Sierens, Bart Soenens, Koen Luyckx, and Willy Lens. 2009. Motivational profiles from a self-determination perspective: The quality of motivation matters. *Journal of educational psychology* 101, 3 (2009), 671.

[77] Marisa Venter. 2020. Gamification in STEM programming courses: State of the art. In *2020 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 859–866.

[78] Cathrin Vogel, Jana Hochberg, Sarah Hackstein, Alina Bockshecker, Theo J Bastiaens, and Ulrike Baumöl. 2018. Dropout in Distance Education and how to Prevent it. In *EdMedia+ Innovate Learning*. Association for the Advancement of Computing in Education (AACE), 1788–1799.

[79] Jan B Vornhagen, April Tyack, and Elisa D Mekler. 2020. Statistical Significance Testing at CHI PLAY: Challenges and Opportunities for More Transparency. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 4–18.

[80] Rand R Wilcox. 2011. *Introduction to robust estimation and hypothesis testing*. Academic press.

[81] Zamzami Zainuddin, Samuel Kai Wah Chu, Muhammad Shujahat, and Corinne Jacqueline Perera. 2020. The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review* (2020), 100326.

A SUPPLEMENTARY MATERIAL

To cope with transparent research, we attached at https://osf.io/grzhp/: the assessment activities (original Portuguese version and a version translated to English); the decision trees used to drive the personalization of gamification; the SIMS (Portuguese); the thematic analysis codebook with quotes for each tag; and the quantitative dataset along with the data analysis process. The interview transcriptions will not be made available due to sensitive data.

Received February 2021; revised June 2021; accepted July 2021