Automating Gamification Personalization to the User and Beyond

Luiz Rodrigues, Armando M. Toda, Wilk Oliveira, Paula T. Palomino, Julita Vassileva, Seiji Isotani

***Abstract*—Personalized gamification explores user models to tailor gamified designs to mitigate cases wherein the one-size-fits-all approach ineffectively improves learning outcomes. The tailoring process should simultaneously consider user and contextual characteristics (e.g., activity to be done and geographic location), which leads to several occasions to tailor. Consequently, tools for automating gamification personalization are needed. The problems that emerge are that which of those characteristics are relevant and how to do such tailoring are open questions, and that the required automating tools are lacking. We tackled these problems in two steps. First, we conducted an exploratory study, collecting participants' opinions on the game elements they consider the most useful for different learning activity types (LAT) via survey. Then, we modeled opinions through Conditional Decision Trees to address the aforementioned tailoring process. Second, as a product from the first step, we implemented a recommender system that suggests personalized gamification designs (which game elements to use), addressing the problem of automating gamification personalization. Our findings i) present empirical evidence that LAT, geographic locations, and other user characteristics affect users' preferences, ii) enable defining gamification designs tailored to user and contextual features simultaneously, and iii) provide technological aid for those interested in designing personalized gamification. The main implications are that demographics, game-related characteristics, geographic location, and LAT to be done, as well as the interaction between different kinds of information (user and contextual characteristics), should be considered in defining gamification designs and that personalizing gamification designs can be improved with aid from our recommender system.**

***Index Terms*—Gamified Learning; Personalization; Educational System; Recommender Systems; Context-aware.**

I. Introduction

To improve learning technologies ability to engage and motivate users, practitioners and researchers have used gamification: the use of game elements in non-gaming contexts [1], [2]. Overall results from these applications are positive, showing improvements in learning outcomes such as academic achievement, conceptual and application-oriented knowledge, and motivation to learn [3]. However, there are situations in which gamification is ineffective in impacting learning outcomes, or even negative [4]. Often, those happen due to poorly designed gamification [5], such as assuming that the same choices will work for all users, the one-size-fits-all approach [6]. To overcome such failures, researchers started to investigate personalized gamification [7].

Personalized gamification concerns exploring knowledge about the users to enable providers (e.g., instructors or the system itself) to offer game elements tailored to those users [8]. For instance, a case would be a system changing from game elements set A to game elements set B when users are females because the latter is tailored to these users. The premise for personalizing gamification emerged from discussions that people with different demographic characteristics and cultural background have distinct preferences [9], behaviors [10], and are motivated differently [11]. Consequently, those might experience and respond to the same conditions in distinct ways [12], [13]. The common practice for gamification is selecting which game elements to add to the system from a list of available elements [14], [15]. Accordingly, researchers invested in providing recommendations indicating which game elements suit better users of different groups to provide personalized gamification, predominantly based on their preferences (e.g., [16], [17]).

Despite personalized gamification is commonly build upon user preference, it is mainly personalized to users' profiles [18], [19]. However, the application context is relevant for gamification's success as well [14], [20], and gamification designs should be aligned to it [6]. Furthermore, multiple factors (e.g., users' demographics [21]–[23] and the system's context [5]) moderate users' experience, either positive or negatively. Although, tailoring approaches often consider a single one, reflecting current gaps in the field of personalized gamification [7], [18], [24]: the fact that i) personalization models should consider more than users' characteristics, such as encompassing the learning activities and geographic locations, and that ii) personalization methods should consider multiple aspects simultaneously, as well as their interactions.

To address these gaps, we sought to understand how to tailor gamified systems to the education domain by considering the learning activity at hand, the user's characteristics, and the geographic location simultaneously, as well as the interactions between all aspects taken into account. To achieve that goal, we performed an exploratory, survey-based research to capture users' preferences, a methodology that has been widely accepted and adopted by related research, as personalization is often based on user preference [15], [25]. As this process is concerned with understanding which aspects (i.e., among learning activity at hand, user's characteristics, and the geographic location) affect user preference, as well as the most suitable game elements for each aspects combination, we sought to answer research question 1 (**RQ1**): *Does users'*

L. Rodrigues, A. M. Toda, W. Oliveira, P. T. Palomino, and S. Isotani are with the Laboratory of Applied Computing to Education and Advanced Social Technology, Institute of Mathematics and Computer Science, University of São Paulo, São Carlos, Brazil. E-mail: {lalrodrigues, armando.toda, wilk.oliveira, paulatpalomino}@usp.br, sisotani@icmc.usp.br. J. Vassileva is with the Multi-User Adaptive Distributed Mobile and Ubiquitous Computing (MADMUC) Lab, Department of Computer Science, University of Saskatchewan, Saskatoon, Canada. E-mail: jiv@cs.usask.ca.

preferences differ depending on (a) their characteristics, (b) geographic location, and (c) the type of the learning activity to be performed?; and **RQ2**: *What is the most useful game elements set, from users' preferences, according to their characteristics, geographic location, and Learning Activity Type (LAT)*[1]*?*

RQ1 informs RQ2 as it reveals which aspects should be considered when defining the most useful game elements set. That is, the combination of game elements (e.g., points, badges, and leaderboards) users prefer the most, an interpretation based on personalization being commonly built upon user preference [7]. Consequently, the challenge that emerges is that interactions from multiple characteristics lead to several combinations. For instance, five binary characteristics would lead to 25 combinations (i.e., 25 recommendations); the number of recommendations for five three-valued characteristics would exponentially increase. Thereby, providing a way to automate such recommendations becomes imperative, which corroborates another challenge of personalized gamification: automating the personalization process [24]. Thus, our **RQ3**: *How to automate gamification personalization?* To answer RQ3, we implement a Recommender System (RS) for personalized gamification [26] based on RQ1's and RQ2's answers. Our RS informs the most useful set of game elements, according to users' preferences, given an input of user's characteristics and their geographic location along with the LAT to be performed. Hence, it enables automating gamification personalization to multiple factors.

Thus, our contributions are threefold. First, evidence from users' preferences that can be used to inform researchers and practitioners on how to tailor Gamified Educational Systems (GES) to LAT, geographic location, and user characteristics. Second, an RS to automate gamification personalization, which performs recommendations by considering multiple aspects simultaneously (i.e., user characteristics, geographic location, and LAT), enabling the implementation of gamification designs more aligned to their preferences. Third, demonstrating which user characteristics impacted their preferences, along with the degree of each one's influence; thus, one might decide which user characteristic to prioritize, take into account, and/or pay more attention as, for instance, moderators of gamification's effectiveness.

II. Literature Review

This section provides background information on the topics covered by this article, reasons about the literature to justify research choices, and highlights the contribution our study provides to existing literature compared to similar works.

A. Game elements

There are many definitions and categorizations of game elements. In the scope of this article, we consider game elements similar to the definition adopted by [15]: the building blocks impacting users' experience with the system, which are characteristic to gameful systems [2], following the vocabulary used more often by similar research [18].

Given the numerous game elements available, it has been common practice for each study to self-select which set of those elements to use. Based on a literature review, [15] presented 59 general elements. In [20], the authors reviewed the literature to select 12 common game elements, without considering any content game element [1] due to the generic nature of their research. In both studies, game elements were selected with no consideration for the domain application, according to their purposes. Differently, [16] explored an element set created from gamification on education literature [27], which is composed by eight options.

Given that our research focuses on a specific domain, education, this article differs from [15], [20] by exploring a taxonomy [28] containing the most common game elements (N = 21) from GES. This taxonomy was created through a rigorous, systematic process, and was validated by 19 experts in the field of gamification and games. Differently, [16] relied on a simpler, reduced game elements set, which was created based on a literature review. Furthermore, by selecting an expert-validated taxonomy, we ensure the game elements available are well defined, avoid using elements with the same purpose but different names, and prevent possible bias from the selection process. Additionally, the selected taxonomy also provides guidance on how the elements are expected to affect users [29], another advantage to those using it [15].

B. Personalized Gamification

Given personalization's importance to information systems, it should be deployed to enhance these systems' relevance to users [6]. Within the scope of gamified systems, a common practice to achieve personalization has been to tailor the gamification design (set of game elements) to specific user's characteristics [7]. In other words, gamified systems have been personalized by performing static adaptations on the game elements it features, based on pre-defined characteristics (i.e., behavioral profile), to tailor the gamification designs [18].

Recent literature reviews [7], [18] found that information used to drive personalization are, predominantly, users' player/gamer types (e.g., HEXAD user types) [19], followed by personality [30]. Nevertheless, it has been shown that other user characteristics, such as gaming habits (e.g., weekly playing time) [31] and gender [32], also impact their preference, as well as the relationship between user demographics (i.e., age and gender) and player types suggest the impact of those aspects [19]. Despite that, these aspects have been rarely explored in methods for tailoring gamification designs in education [33]. This research addresses this need by introducing an approach that exploits demographic and gaming habits as information used to drive the gamified designs' tailoring.

Furthermore, the user is not the only factor to be considered when defining gamification designs. A factor that has been often discussed as relevant for gamification effectiveness [3], [14], [34], which is rarely considered by tailoring methods, is the application context (e.g., geographic location). Specifically in the context of educational systems, an aspect researchers

[1]In the scope of this study, a LAT is defined based on its main expected outcome (see Section II for further details)

have recently argued as relevant, and recommended to consider when tailoring gamified systems, is the learning activity [18], [35]. This is related to the recommendation that gamified designs should match the task [6] and, given that tasks of educational systems are almost ever learning activities, personalizing the gamified designs to these activities should be accomplished.

Despite that, to the best of our knowledge, there are only two approaches for personalizing gamified designs based on learning activities [17], [25]. In [17], learning activities are considered based on their main expected objective, similar to this article. In [25], the learning activities are activities from Moodle (e.g., forum and quizzes). Hence, while recommendations from [17] can be extended to any learning activity (linked to their objective), those from [25] are limited to a specific set of Moodle activities. In addition, both works consider one user characteristic, personality trait and player type, respectively. Thus, they provide valuable contributions in terms of exploring learning activities, as well as presenting recommendations that consider the interaction between those and a user characteristic (e.g., player type X, learning activity Y).

However, these studies fall into the category of methods that rely on the most often researched user characteristic, a single user characteristic is considered in each one, and the guideline from [25] cannot be generalized to any learning activity. Therefore, the main advances of this article compared to those works are: i) considering multiple user characteristics rarely explored simultaneously, ii) taking the context into account via learning activities and users' geographic location, and iii) providing recommendations that consider the interaction between all of those aspects that are relevant for users.

C. *Learning Tasks*

To generally describe a task, one might rely on its desirable outcomes, behavioral requirements, and/or complexity [6], [36]. Similarly, from the human-computer interaction perspective, a task refers to the activities required to achieve a specific goal [37]. Consequently, given the context of our study, a learning task refers to a set of activities that aim at some educational outcome. From this definition, it is possible to note that numerous tasks might be found in GES, which makes it infeasible to develop a specific personalization approach for each one. An alternative to that limitation is categorizing the activities, which can substantially reduce their quantity; consequently, enabling the recommendation of gamification designs to each category.

To overcome the numerous learning tasks and categorize them, we opted to rely on the revision of Bloom's taxonomy of educational objectives [38]. This approach contributes to the learning process by matching the educational activities' gamification designs to a cognitive taxonomy [17]. Although there are other options available, the revision of Bloom's taxonomy is a widely cited, well-accepted taxonomy, similar to its original version [39]. It acts as a framework that can be used to classify what is expected from an educational activity (outcome), as well as its complexity [38]. The revised version is composed of two dimensions: knowledge (concerned with what is to be learned; e.g., the subject of matter) and cognitive process (concerned with actions associated with learning; e.g., how to learn) [38].

In the scope of this research, we consider the second dimension, similar to related work [17]. By categorizing learning activities based on the cognitive domain of such a taxonomy, we avoid having the gamification focused on the activity itself (e.g., completing a quiz or answering a forum) and allow it to be aligned with the activity's expected learning outcome. Thereby, addressing the recommendation that gamification should match the task [6]. Moreover, as many GES feature tasks of varied subjects, the second dimension choice makes the approach subject-independent, focusing the gamification designs' tailoring on the activities' particular objectives while allowing it to be used regardless of the system's educational topic.

The structure of the cognitive process dimension is split into six categories: remember, understand, apply, analyze, evaluate, and create. Here, we consider each dimension a different LAT, wherein their complexity increases following the order in which they were introduced (i.e., remember is the less complex and create is the most complex). Hereafter, we refer to those as LAT1 to LAT6, also following the introduced order. Furthermore, although an activity might fit in more than one LAT, our approach considers every activity will have a predominant, main objective to be achieved. Hence, the personalization process should be based on that main goal. It is worth noting that those LAT might be split again, however, we opted to work with the high-level abstraction given that the similarities within these sub-categories might be even higher. Thus, this article contributes a proposal that is based on the six high-level types of cognitive processes established in [38], that aids in tailoring gamification designs to different LAT, according to their predominant goal.

D. *Recommender Systems for Personalized Gamification*

An RS can be seen as a technique, or software tool, able to recommend items to users [40]. Such systems are especially valuable for cases in which several options are available, alleviating the burden of human selection by providing recommendations, often based on what other people recommend. Common applications of such systems are e-commerce, movies, and music. Recently, the use of RS has been suggested for personalized gamification [26], which corroborates to our research in terms of, for instance, reducing the burden of selecting the most suitable game elements for several combinations of user characteristics, geographic location, and LAT. Next, we provide a brief overview of RS for personalized gamification following the framework by [26].

RS have three main elements: inputs, outputs, and process. Inputs concerns all the aspects that are received by the RS to be taken into account before doing the recommendations. There are four main types of input: user profile (e.g., demographics, personality, behavioral profile), items (e.g., game elements), transactions (e.g., the relationship between users and items; using or preferring a game element), and context (e.g., geographic location, activities to be done). Outputs are ratings related to the choices that the RS made from the input received. For instance, if items are game elements, the output would be the rating of each one. The process is the core part of the RS, concerning the method through which it will perform the recommendations. There also are four main recommendation methods. Content-based recommenders are based on knowledge of the application, such as data log, or empirical and theoretical information. The collaborative filtering method exclusively depends on data collected implicitly or explicitly from interactions with a system. Context-aware recommenders are those that explore information of the context to make their choices. Lastly, hybrid recommenders aim at using two or more of the previous approaches together.

Generally, there is a lack of technological support for gamifying educational environments [41]. Accordingly, the literature on personalized gamification lacks concrete RS implementations, demonstrated by recent literature reviews finding only four studies that relate to RS or other forms of automating gamification personalization [24]. Among those, one is the framework proposal itself [26], whereas the remaining are theoretical/conceptual models with no concrete implementations available for third-parties use [12], [42], [43]. Differently, we present and provide an RS for personalizing gamification, which was built upon findings from the study of this article. Hence, we advance the literature with a free, hybrid RS as it uses both contextual as well as empirical information from users' preferences.

E. Summary

Table I summarizes and demonstrates the points in which this study differs from related works based on our previous discussion. As shown, most studies focus on user characteristics, few consider the task to be done, and none but this one takes into account geographic locations. Additionally, the few works that consider information from the user and the task provide recommendations based on two factors (one from each kind). On the other hand, our approach was developed considering nine aspects, of which eight were found to be significant (see Section IV) and, therefore, are considered in the product from our research (see Section IV-D). This final product is another key difference. Whereas previous research only provides conceptual/visual guidelines, this study contributes with technological aid for the design of personalized gamified systems. This also differs from research on recommender systems for gamification [12], [42], [43] as those provide no concrete implementations from their proposals.

TABLE I
SUMMARY COMPARISON OF RELATED WORKS.

	Recommends game elements based on				
Study	User	Task	GL	N Factors	Product
[15]	X				Conceptual
[20]	X				Conceptual
[16]	X				Conceptual
[19]	X				Conceptual
[31]	X				Conceptual
[32]	X				Conceptual
[25]	X	X		2	Conceptual
[17]	X	X		2	Conceptual
This	X	X	X	8	Technology

GL = Geographic location.

III. STUDY

The goal of this research is to understand how to tailor GES to LAT, geographic location, and users' characteristic. To achieve that goal, we performed a survey-based research asking participants to indicate their preferred game elements for each LAT. Up to date, this methodology is the most used by similar works [7], [24] and has been widely accepted given the number of related research following it [15], [25]. Therefore, we considered it the most adequate approach to adopt. This study also follows an exploratory approach, which aims to understand possible relations between the observable variables to create possible research guidance [44]. Based on that, this section presents an overview of this study development process, as well as further describes the material and methods followed.

A. Overview

In developing this research, three factors had to be defined: what domain, how to interpret the tasks, and which user characteristic to consider. **First**, we opted for the education domain, which is the one gamification research has focused the most [14] and, both positive [3] and negative [4] outcomes have been found, showing the need for further research. **Second**, given the domain, users will perform learning activities when using the gamified systems. As one might create numerous of those activities, our approach considers activities types based on the revised Bloom's taxonomy [38], an established, well-accepted taxonomy within the educational context. **Third**, we chose to focus on users' demographic characteristics and gaming habits and preferences, deepening into aspects that have been discussed as relevant factors [21], [31] but received less attention from the academic community compared to the most used ones [7], [24].

Then, to achieve the desired understanding, we developed Conditional Decision Trees (CDT) [45], which takes into account the interactions between all input variables to provide recommendations on the most suitable game elements given an input set. During data collection, we operationalized gamification designs as the top three game elements participants prefer the most, provided game elements (N = 21) extracted from an expert-validated taxonomy [28], and operationalized LAT as the six cognitive process types defined in [38]. Note that we chose this top-three design to match the number of elements of the most used gamification design (PBL - points, badges, and leaderboards) [46] because the number of game elements might affect gamification's effect [47].

B. Procedure

The following five steps were performed to develop our approach for tailoring gamified designs to LAT and users.

1) **Survey development**: defining the survey design and sections and the game elements and LAT to consider;

2) **Data collection**: disclosing the survey online, through Amazon's Mechanical Turk[2] (MTurk), to collect participants opinions.
3) **Data analysis**: running analyses to identify which characteristics impact users' preferences.
4) **Users preferences analysis**: investigating our findings to identify how to tailor educational systems' gamification designs to users, geographic location, and LAT.
5) **RS design**: developing a free, ready-to-use resource, based on our findings, to aid those who want to tailor their educational systems' gamified designs.

C. Survey

The survey was developed online[3] and can be viewed in the appendix. Its design was defined in four steps. First, two researchers brainstormed and developed an initial version. Second, three other researchers revised it and provided feedback on how to improve it. Following, the survey was improved accordingly and, lastly, we ran a pilot study with 50 participants.

The final version has four sections: consent form, demographics, gaming background, and preferences. In the **consent form**, all respondents were informed to be participating in a research and agreed all information provided would be used to research ends only. The **demographics** and **gaming background** captured participants' gender, age, living country, highest level of education, and MTurk identifier to avoid repeated completions, and for how many years the participants researched/worked with gamification (0 for those who did not), how much time (in hours) they spend with games per week, and their preferred game genre and playing setting, respectively. Lastly, in the **preferences** section, participants ranked the top three game elements they prefer the most when performing each of the six LAT. To aid users, this section described each game element along with examples.

The 21 game elements available were: Acknowledgment; Chance; Competition; Cooperation; Economy; Imposed Choice; Level; Narrative; Novelty; Objectives; Point; Progression; Puzzles; Rarity; Renovation; Reputation; Sensation; Social Pressure; Stats; Storytelling; Time Pressure. Further descriptions of these elements can be seen in [28]. The LAT are those introduced in Section II (remember, understand, apply, analyze, evaluate, and create). For further information about each one, see [38]. Thus, the last section had seven items - one for each LAT - and a repeated item to assess participant attention/consistency (see next section).

Each of those seven items had three sub-items, allowing the participant to select the rank-one, -two, and -three game elements, in which the 21 game elements were possible answers. Nevertheless, the same game element could not be selected twice within the same item. That is, each participant's top three should be composed of three different game elements. A sample question was *Indicate the three gamification elements you consider will help you the most when performing an activity you need to REMEMBER something (e.g., remember what the '+' symbol means in arithmetic operations).*, whereas other items of the same section differed only in the LAT (e.g., understand instead of remember) and the example at the end of the item. All items had basic mathematical examples due to the generality of the topic.

Additionally, we highlight that this top-three survey design was adopted due to the number of game elements (21) and LAT (six), which would lead to a questionnaire with 126 items if subjects should, similar to related work [15], [16], provide a rating for each gamification element through a Likert-scale. That is, participants would answer to 21 items six times; one time per LAT. Thus, we opted for one item per LAT, featuring three options each, to reduce effort, tiredness, and time spent in completing the survey, aiming to improve answers' reliability. Lastly, note that the survey sections' order was fixed (the same as previously introduced) but, within each section, the items' order was randomized.

D. Data Collection and Filtering

We recruited participants through crowdsourcing (MTurk). We made this choice to increase our sample size, similarly to related research [15], [20], an approach that has been recommended in the literature [48], [49] to improve external reliability [50]. No participant restriction was enforced to avoid selection biases and everyone who completed the survey received a fixed remuneration.

Nevertheless, similar studies [15], [20] have employed additional items to survey's long sections to assess whether participants are paying attention and providing consistent answers. Then, based on those specific items' answers, researchers filter participants according to some assertion threshold (e.g., discarding those who failed in more than one item [20]). In this study, we adopted a similar approach. On the *preferences* section, we added a repeated question for one LAT, which allowed us to assess whether the participant was consistent their answer (i.e., did they select the same top-three game elements in both items?). Participants' remuneration was not conditional to consistently answering, neither participants were warned about the repeated item, aiming to improve the reliability assessment.

Following related work, we adopted a tolerance for inconsistent completions. Hence, we removed all participants that provided consistent answers in less than two out of the three game elements. For instance, one selected Acknowledgment, Chance, and Competition and, then, in the repeated question, selected Acknowledgment, Cooperation, and Economy. This participant would be discarded by selecting two different game elements for the same question. In total, 1018 individuals have completed the survey, from which 657 answers were discarded based on our criteria. Thus, the final dataset contains 361 consistent answers. The description of these reliable, valid answers is shown in Table II.

Overall, our sample is composed of adults (51.5% males, 47.4% females, and 1.1% others) with 32 years on average ($\pm$11) and undergraduate or higher degrees (65.4%). Hence, we might expect our sample to feature responsible people with good educational background. Furthermore, despite the

[2]https://www.mturk.com/
[3]Online survey: http://bit.ly/2JWxwqs

large majority never researched gamification (91%), there is an interesting variation in their preferred playing setting (Singleplayer: 59%; Multiplayer: 41%) as well as game genre (20% for the most preferred genre: Role Playing Game), with an overall playing time of 12 hours (±13) per week. Thereby, we might expect participants to be familiar with games and their elements.

TABLE II
DATASET DESCRIPTION.

Value	N(%)	Value	N(%)
Gender		*Preferred playing setting*	
Female	186 (0.515)	Singleplayer	214(0.593)
Male	171 (0.474)	Multiplayer	147(0.407)
Other Gender	4 (0.011)	*Researched gamification*	
Country		No	329 (0.911)
United States	259 (0.717)	Yes	32 (0.089)
India	22 (0.061)	*Age*	
United Kingdom	20 (0.055)	Mean	32.615
Canada	18 (0.050)	SD	11.299
Brazil	16 (0.044)	Min.	18.000
Others	26 (0.072)	25%	24.000
Highest education level		50%	29.000
Undergraduate	161 (0.446)	75%	39.000
High School	81 (0.224)	Max.	75.000
MsC	63 (0.175)	*Weekly playing time (hours)*	
Technical education	30 (0.083)	Mean	12.874
Other Education	14 (0.039)	SD	13.782
Ph.D	12 (0.033)	Min.	0.000
Preferred game genre		25%	4.000
Role Playing Game	75 (0.208)	50%	10.000
Adventure	61 (0.169)	75%	20.000
Action	60 (0.166)	Max.	112.000
Strategy	50 (0.139)		
Other Genre	115(0.319)		

Others, shown as: Country (Count) = Italy (6), Germany (5), Spain (3), Australia (2), Netherlands (1), Albania (1), France (1), Ireland (1), Poland (1), Turkey (1), Austria (1), Nigeria (1), Belize (1), Jamaica (1).

E. Data Analysis

For data analyses, we decided to work with decision trees, algorithms that determine an output based on the interaction between elements from an input set [51]. Besides handling interactions, which is key for our objective, a decision tree provides other three positive points that led us to choose it. First, it allows visualizing the rules followed to determine the output. Therefore, we can comprehensively discuss and understand how game elements are selected, given an input set (user data and LAT). Second, it demonstrates which aspects are more or less important, as the main ones are in the tree's top, and vice-versa. It also ignores unnecessary inputs, excluding from the tree those that do not contribute. That is, it works as a feature selection method. Hence, providing insights on which aspects influence users' preferences, as well as which are most influencing ones from those we studied. Third, the algorithm itself determines how each characteristic will be split (e.g., should age be split in 18-28, 29-39 or 18-23, 24-29, 30-39?), removing human bias that are likely to be inserted in this process.

Nevertheless, decision trees might be implemented in varied ways. The standard classification approach is optimized towards predicting a single output class (i.e., one of the output's values) given an input set [52]. Then, for our aim of creating an RS (RQ3), this approach would yield limited performance, in terms of the RS's ratings, because of the focus on a single, definitive output. Additionally, to cope with our survey design (top-three selections), we would need to create tree decision trees. An alternative is multi-label classification [53], wherein the output has multiple values (e.g., A-B instead of just A or just B). While this alternative would lead to a single, multi-label decision tree, it also has limitation concerning the RS ratings. Besides, it would limit our answer to RQ1 because we would not be able to distinguish how trees' rules change as the importance of users' preferences change from first to second to third selection. Moreover, there are algorithms that learn to rank [54], wherein the input is a list of items that can be ordered (i.e, they have a rating that is used as the sorting criterion), but creating three trees remains necessary.

Apart from those, there are CDT, an alternative based on a statistical approach [45]. Consequently, CDT allow working with frequency tables as outputs, such as in a Chi-square test. That is, a distribution-based output. Hence, differently from standard singe output classification, the algorithm is optimized to such output format that gives a rate for each game element. Consequently, the output has ratings for our RS (RQ3), while we can also identify the most (and least) useful game elements (RQ2) based on highest (and lowest) values. Additionally, because CDT are tree-based algorithms, we can still identify which factors are relevant and how choices are made (RQ1) from their visual and feature selection properties. Therefore, we use CDT because they i) are designed to work with distribution-based outputs, ii) conduct internal feature selection, and iii) provide visual interpretations of their rules.

Furthermore, as CDT follow a statistical approach, they optimize the model evaluation and validation process [45]. In creating a model, for each factor (i.e., a feature of the input set), the algorithm analyses whether splitting it has a significant impact on the output (i.e., $p \leq$ alpha). For instance, an example would be checking whether including gender as a factor would change the distribution of participants' preferences for each game element. For our analysis, because the output is a distribution (i.e., a frequency table of game elements' selections), CDT use a test like a Chi-square. Hence, the algorithm itself performs the feature selection process, based on whether a factor significantly changes the output. Note that the algorithm only creates a node after analyzing all factors and selecting the most discriminant one (based on statistical significance). Then, the algorithm recursively repeats that procedure until no more significant splits are found. Consequently, the higher the node in a tree, the higher its importance.

Importantly, the algorithm defines how to split a feature (e.g., United State vs all others or United States and Canada vs all others). Therefore, CDT allow an statistical, in-deep analysis of which input factors are associated with the output, as well as reveal each factor's importance [45]. Additionally, notice that all splits are statistically significant. Hence, them as well as the factors included in the trees are expected to be variables with roles that generalize from the sample to the population, according to the Null Hypothesis Significance Testing Framework [55]. Thereby, that statistical approach acts as an embedded method of model validation, similar to cross-validation. Consequently, the model is evaluation depends on the extent to which the splits are statistically significant. While that approach differs from standard classification metrics (e.g., accuracy and precision), it is aligned to our method, especially because we used distributions as the CDT's outputs. Hence, metrics such as accuracy and precision would not properly evaluate our models.

Thus, according to our goal, data captured via our survey (Table II) was entered as input to generate three CDT. Each tree's output was users' preferences for either their first, second, or third selected game elements. Accordingly, each tree predicts the distribution of users' preferred game elements. Note that the dataset described in Table II is in the *wide format*. That is, one row per participant, and one column for their preference on each LAT (one column for LAT1, one for column LAT2, and so on). Then, to generate trees able to distinct users' preferences from one LAT to another, we converted the dataset to the *long format*; that is, six rows per participant, a new column indicating the LAT each row corresponds to, and a single column indicating the preferred game element from each user for each LAT. We highlight that, although this increases the size of the dataset inputted to the CDT, the characteristics' distribution remains the same.

IV. Results

First, this section explores the CDT generated from our data, discussing the answers for our research questions in light of insights gained from them.

A. Conditional Decision Trees Overview

We built our first CDT - CDT1 - from participants' number one choice. That is, the game element they prefer the most for each LAT. Similarly, our second and third CDT - CDT2 and CDT3, respectively - concern the game element participants selected as the second- and third-preferred ones for each LAT. CDT1 is shown in Figure 1, where circles represent decision nodes and rectangles are leaf ones. Decision nodes function as if/else statements. For instance, the first node tests if, for a given input, the preferred game genre is equal to adventure, other genres, role playing game, or strategy (left), or equal to action (right). Based on the answer, it is decided whether one should follow to the left or right path of the tree. This procedure is iteratively repeated for each decision node until reaching a leaf node. Leaf nodes indicate the tree's output, which are the game elements' ratings (for simplicity, Figure 1 shows the game element(s) with the highest rating). Hence, for someone whose preferred the game genre is action and lives in the Netherlands or Spain, CDT1 recommends Objectives.

Note, however, that the tree in Figure 1 is a simplified version compared to the original tree generated from the R package *party*. That version has two main differences. First, it shows p values for each decision node, demonstrating they are significant splits. Second, their leaf nodes present bar plots, demonstrating to each game element's rating. Such information can be used to recommend the most preferred game element (i.e., highest rating) or to provide ratings on the most likely preferred ones (i.e., output all elements' ratings). Considering this context, we highlight Figure 1 only presents the most preferred game element due to the limited space, as the full image would not be readable within the article template. For similar reasons, CDT2 and CDT3 are not shown in the article. Nevertheless, the full images, with barpots for all leaf nodes, from all CDT we created, are available in the appendix.

B. RQ1: Characteristics that Impact User Preferences

RQ1 concerns finding which aspects, among user characteristics, geographic location, and LAT, impact users' preferences for game elements. Therefore, we analyze which of those appear in our CDT to identify the ones that influence participants' choices.

CDT1 used six of the nine (eight from Table II plus LAT) inputs: preferred game genre, LAT, gender, country, experience researching gamification, and education, which appeared in the tree in this order. Thus, for participants number one choice, those are the characteristics that impacted their preferences, with preferred game genre and education being the most and the less influencing ones. *CDT2* also used six out of the nine inputs: country, LAT, preferred game genre, gender, preferred playing setting, and weekly playing time, with the same order of relevance as presented here. Thus, for participants number two choice, those are the six characteristics that impacted their preferences, with country and weekly playing time being the most and the less influencing ones. *CDT3* used five of the nine inputs: country, preferred game genre, experience researching gamification, LAT, and education, which appeared in the tree in this order. Thus, these are the characteristics that influenced participants' preferences for their third choice, in which country was the most relevant one, as opposed to education and LAT that were both the less relevant ones.

Based on these findings, we answer **RQ1** with evidence that factors impacting users' preferences are country, LAT, preferred game genre and playing setting, gender, experience researching gamification, weekly playing time, and education. Additionally, we also found the order of importance of these characteristics for each of the three selections. This finding is summarized in Table III, which demonstrates the highest level of the tree where each characteristic appears (because one might appear multiple times and at different levels). Consequently - as the higher the level, the more the importance - allowing us to identify each one's importance.

TABLE III
Level in which each characteristic appeared in the CDT of each users' choices.

Choice	Cnt	LAT	PGG	PPS	G	ERG	WPT	Edu
First	4	2	1		3	4		5
Second	1	2	2	5	5		6	
Third	1	4	2			2		4

Cnt = country; LAT = learning activity type; PGG = preferred game genre; PPS = preferred playing setting; G = gender; ERS = experience researching gamification; WPT = weekly playing time; Edu = education.

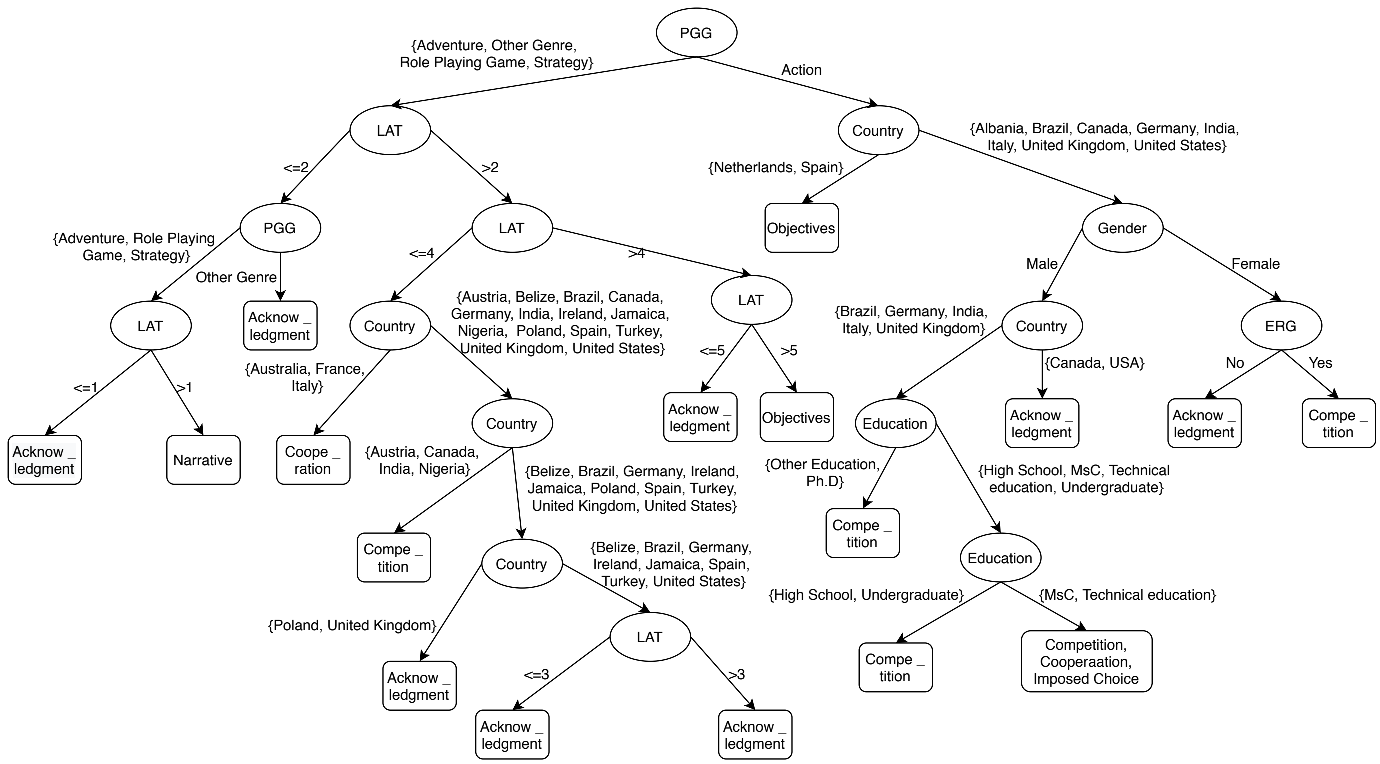

Fig. 1. Conditional decision tree for participants most preferred game element. Codes refer to preferred game genre (PGG), learning activity type (LAT), and experience researching gamification (ERG). Note that in cases of ties, leaf nodes present all game elements tied in alphabetical order.

C. RQ2: Most useful Game Elements Sets from User' Preferences

RQ2 concerns identifying the most useful game elements, given users' characteristics, the LAT they will perform, and their geographic location, according to participants' preferences. From the three CDT we generated, note that CDT1, CDT2, and CDT3 have 17, 16, and 15 terminal nodes, respectively. This means that, together, all trees provide recommendations for 48 combinations of the input set. Consequently, presenting a complete description of the recommendation for each of these combinations is unfeasible. Nevertheless, we demonstrate recommendations for specific cases to illustrate the most useful game elements for some cases, according to our findings.

First, consider the simple case wherein one wants to personalize gamification to LAT only, without considering any user characteristic. To illustrate that case, we split our dataset in six: each one containing only rows of one LAT. Then, we predict the output from each of our CDT using each sub-dataset. The results (Table IV) show there are cases (e.g., first, second, and third rows) in which the same element is recommended as second and third preferred. Although one participant could not select the same element for both cases, this corroborates the fact that the most select game element as second and third, considering the overall sample, was Objectives. Accordingly, our CDT recommend the same element as the second and third choices. With that in mind, our findings suggest that the most useful game elements set, considering LAT and no user characteristic, for LAT1 is Acknowledgment and Objectives, for LAT2 is Narrative and Objectives, and so on.

TABLE IV
RECOMMENDATIONS FOR PERSONALIZING GAMIFICATION TO LAT ONLY, WITHOUT CONSIDERING ANY USER CHARACTERISTIC, BASED ON OUR DATASET.

LAT	First	Second	Third
1	Acknowledgment	Objectives	Objectives
2	Narrative	Objectives	Objectives
3	Acknowledgment	Objectives	Objectives
4	Acknowledgment	Objectives	Acknowledgment
5	Acknowledgment	Level	Point
6	Objectives	Objectives	Progression

Among the main contributions from our approach, is its ability to handle multiple characteristics simultaneously, as well as the interaction between these characteristics. Therefore, we exemplify cases of personalizing gamification for a learning activity wherein students need to remember (LAT1) some content from long-term memory and then perform a second activity in which they need to evaluate (LAT5) others' opinions. Additionally, let us compare the most useful game elements set for Brazilian and Americans performing such activities. For simplicity, assume all students are males, never researched gamification, High School degree is their highest education level, play similar amounts of time per week (10 hours), and prefer the same game genre and playing setting: action and singleplayer, respectively[4]. In this context, the recommendations are likely to vary due to changes in LAT,

[4]This fixed combination was selected arbitrarily, aiming to simplify the illustration. Other characteristics were not mentioned as they were found not to influence user preferences (see Table III)

as well as geographic location (country), as all other relevant characteristics are the same. Finally, Table V demonstrates the recommendations for those cases.

TABLE V
RECOMMENDATIONS, DEPENDING ON LAT AND COUNTRY, FOR AN ARBITRARILY SELECTED SAMPLE: MALES, WHO NEVER RESEARCHED GAMIFICATION AND HAVE HIGH SCHOOL DEGREE AS THEIR HIGHEST EDUCATION LEVEL, PLAY 10 HOURS PER WEEK, AND PREFER PLAYING ACTION GAMES ALONE.

Combination	First	Second	Third
LAT1 - USA	Acknowledgment	Competition	Competition
LAT1 - Brazil	Competition	Competition	Time pressure
LAT5 - USA	Acknowledgment	Level	Point
LAT5 - Brazil	Competition	Level	Point

Our results (Table V) suggest the most useful game element set for LAT1 for Brazilians is Acknowledgment and Competition, whereas that for Americans is Competition and Time Pressure. For LAT5, the recommendation for Brazilians is Acknowledgment, Level, and Point, while for Americans the difference is Competition rather than Acknowledgment. Hence, highlighting the impact of contextual factors on users' preference, which differed depending on the LAT they were expecting to perform, as well as their geographic location.

In summary, we demonstrated which are the most useful game element set for specific combinations of user and contextual characteristics. In doing so, we selected the elements with the highest ratings according to our CDT. We did not show the recommendations for all combinations due to space restrictions. However, our CDT can be analyzed in their completeness - incuding each element's ratings - in the appendix for finding the recommendations.

D. RQ3: Recommender System to automate personalization

To provide technological aid that helps automating gamification personalization, consequently coping with the complexity of determining recommendations from visual inspection of CDT, we converted our tree CDT into an RS (**RQ3**). This system encapsulates all trees and simplifies the task of determining which game elements to use given a user, a LAT, and a geographic location. In summary, the RS's algorithm is as follows:

// 1. receives external information to create the trees's input
$a \leftarrow$ *user's preferred game genre*
$b \leftarrow$ *user's preferred playing setting*
$c \leftarrow$ *user's weekly playing time*
$d \leftarrow$ *user's gender*
$e \leftarrow$ *user's highest educational level*
$f \leftarrow$ *whether the user researched gamification before*
$g \leftarrow$ *user's living country*
$h \leftarrow$ *LAT to be gamified*
// 2. creates the trees's input
$input \leftarrow [a, b, c, d, e, f, g, h]$
// 3. uses the trees to get recommendations based on the input
$rec\_cdt1 \leftarrow CDT1(input)$
$rec\_cdt2 \leftarrow CDT2(input)$
$rec\_cdt3 \leftarrow CDT3(input)$
// 4. creates the output merging the recommendations from all trees
$output \leftarrow [rec\_cdt1, rec\_cdt2, rec\_cdt3]$

Specifically, the algorithm's first block receives external information in the same format as users passed to our survey (see Section III). Second, the algorithm merges such information to create the CDT's input. Third, the algorithm passes that same input to each of the CDT, which return a list with a rating for each game element (values might be zero). Lastly, the algorithm merges the rating lists from all trees into a single matrix-like output similar to Table VI. Next, we describe the characteristics of our RS and briefly present technical concerns on how we converted our CDT into a free, easy-to-use system able to automate the personalization of GES.

We characterize our RS according to the framework for RS for personalized gamification introduced in [26]. Our RS considers six user inputs. Those are their preferred game genre and playing setting, weekly playing time, gender, highest education level, and whether the user researched gamification before. In addition, the user's living country, as well as the LAT that will be gamified, must also be entered, inputs related to the context [56]. Items are the game elements users could choose in the survey, the 21 game elements from the taxonomy proposed and validated in [28]. Lastly, the transactions concern users' preferred game elements (i.e., user with characteristics X, from geographic location Y, prefers element Z for LAT W), which is defined according to our findings.

The method adopted for output selection characterizes our RS as a hybrid recommender [26]. The input involves two contextual characteristics, geographic location and the LAT to be performed. Accordingly, the method would be characterized as a context-aware recommender. However, the selection process also relies on empirical information from our findings, which concerns a content-based recommender. Thus, our RS is a hybrid recommender due to exploring the characteristics of two methods. Lastly, our RS outputs are the ratings for game elements, for each of the top-three recommendations, defined according to the percentual of each game element's selection for that input. Consequently, the highest percentual reflects a recommendation's accuracy, given that the element with the highest rating is recommended. For instance, if we consider the case shown in Table VI, the Acknowledgment game element rating would be roughly 0.25 for the first selection. Accordingly, the recommendation accuracy is 0.25 because 25% of the observations within those criteria selected Acknowledgment.

Table VI demonstrates an output of our RS for people who live in the United States, have no experience in researching gamification, the preferred game genre is RPG, and will complete an analyzing learning activity. It demonstrates a full output of the RS with the ratings for all game elements when considered as first-, second-, and third-preference. For participants' number one preference, the game element with the highest rating is Acknowledgment (0.246), followed by Novelty (0.116). For participants number two choice, the highest rating is for Level (0.123), followed by Novelty (0.107). For their third preferred element, Point holds the highest rating

(0.125), followed by Novelty (0.112). When using the RS for other input sets, similar outputs will be given, likely with different ratings for each game element. Hence, based on outputs as that shown in Table VI, one can assess which elements are more likely to be the preferred ones for a given situation and define their gamification design accordingly.

Note that, in a different perspective, one could select the top items from the first tree (column two: Acknowledgment, Novelty, and Level) or those with the highest average rating when considering all trees. While we hope future research explores those approaches, they would remove the implicit, importance ordering participants provided when completing the survey. Accordingly, our discussion was only based on the highest rating from each tree to respect such ordering.

TABLE VI
RATINGS OF OUR RS FOR PEOPLE WHO LIVE IN THE UNITED STATES, HAS NO EXPERIENCE IN RESEARCHING GAMIFICATION, PREFERRED GAME GENRE IS RPG, AND WILL COMPLETE AN ANALYZING LEARNING ACTIVITY. HIGHEST RATINGS OF EACH CHOICE HIGHLIGHTED IN BOLD.

Game element	First	Second	Third
Acknowledgment	**0.246**	0.069	0.086
Chance	0.043	0.032	0.036
Competition	0.050	0.060	0.056
Cooperation	0.076	0.095	0.076
Economy	0.040	0.041	0.040
Imposed Choice	0.050	0.091	0.076
Level	0.106	**0.123**	0.046
Narrative	0.003	0.009	0.010
Novelty	0.116	0.107	0.112
Objectives	0.023	0.079	0.063
Point	0.027	0.057	**0.125**
Progression	0.053	0.038	0.030
Puzzles	0.007	0.006	0.013
Rarity	0.013	0.022	0.017
Renovation	0.013	0.022	0.050
Reputation	0.003	0.019	0.033
Sensation	0.010	0.003	0.023
Social pressure	0.076	0.069	0.069
Stats	0.017	0.035	0.017
Storytelling	0.027	0.022	0.023
Time Pressure	0.000	0.000	0.000

Aiming to improve the usability of our RS, we reimplemented the CDT generated through the R package *party* [45] in Javascript. Although one could access the R objects, or try to make some external connection to R code from, for instance, a web browser, this process could be laborious and discouraging. On the other hand, Javascript can be easily run in most web browsers, as well as be easily plugged-in into a web site. Furthermore, as decision trees can be represented through a set of if/else statements, the conversion from R objects to Javascript does not require handling complex programming technical challenges. This is another advantage because the procedure of transforming our CDT into a Javascript plugin can be replicated to any other programming language.

Our RS is freely available (see the appendix), and there are two main use cases in which we believe it can be explored. First, the main case of automating gamification personalization, in which other systems use it as an external resource/tool. In this case, a gamified system can explore our RS as a plug-in that is consulted to find which game elements should be available for some occasion. To this end, the system would call the plug-in, passing the needed inputs as parameters to receive the ratings of each game element. Then, the system could, for instance, turn on those elements with the highest ratings. This procedure could be iteratively repeated, when the type of the learning activity to be performed changed, for instance. Thus, the RS would aid the system in performing dynamic adaptations [18] of its gamification design according to the user's characteristics and geographic locations as well as the tasks performed. The second case is using our RS as a standalone tool to provide recommendations for one interested in, for instance, personalizing an unplugged gamified environment [57] or to manually define their system gamification.

V. DISCUSSION

Based on participants' preferences captured though a survey, our findings provided evidence that users' preferences differ depending on their characteristics, geographic location, and the LAT to be performed (RQ1). Also, we were able to develop an RS that recommends the preferred gamification design for a LAT to be performed by a user with some specific characteristic in a defined geographic location (RQ2). The main contribution of this research is, therefore, providing a free RS for personalized gamification, built upon a state-of-the-art approach, that aids in automating the tailoring of gamification designs by suggesting which game elements to use (RQ3). This RS is based on three aspects of personalization: domain, user, and task [6], implemented as the educational domain, demographics and gaming characteristics, and LAT and geographic location, respectively. Additionally, we revealed which context and user characteristics impact their preferences, and which of those are more or less relevant, contributing to expanding and grounding knowledge from previous studies (e.g., [15], [31]).

Concerning the results on users' characteristics impacting their preferences, our findings are aligned with the literature. Previous studies have shown that, for instance, demographics [19], [32] and attributes related to users' gaming habits [31] affect user preference. We corroborate those by providing more empirical evidence that users with different characteristics have different preferences, as well as presenting which of those are more important than others. For instance, we found simple user attributes, such as gender and having researched gamification, are less relevant than gaming-related characteristics (see Table III), which is in line with previous literature suggestions [58]. Furthermore, it also has been discussed that the task to be performed influences the perceptions of gamified systems' users [6]. Following that and within the educational context, suggestions to consider learning activities within the tailoring process of educational systems have emerged [18], [35]. Our findings are aligned with those theories as well, showing that users' preferences differ depending on the LAT they expect to perform (see Figure 1 and Section IV-C). Additionally, we found geographic location to be another relevant factor, a finding consistent with recent literature suggestions [24].

Concerning the results on users' preferred gamification design for each LAT given their characteristics, we expand the literature by i) providing recommendations applicable to any task (by considering its main objective - type) and ii) exploring less studied user characteristics (i.e., demographics and gaming-related) as well as taking into account their geographic location. On one hand, besides not guiding on how to tailor to LAT and geographic location, other personalization approaches often rely on user profiles [7]. However, as shown by our findings (see Table III), demographics and gaming-related characteristics are relevant as well. On the other hand, despite the recent calls for considering learning activities when personalizing [18], [35], available approaches considering such aspects are yet limited, mainly due to considering only two characteristics (one for from user and the learning activity) [17], [25]. Although, if multiple aspects are relevant, they all should be considered, as well as their interaction [6], [22]. Our research contributes to these concerns, guiding how to personalize gamification to users (i.e., demographics and game-related) and contextual (i.e., LAT and geographic location) aspects simultaneously.

Moreover, this article advances the literature by providing an RS for personalized gamification. In [26], a framework for such RS has been proposed, however, the literature still lacks concrete implementations of these systems. On the other hand, recent research has highlighted the need for research to aid in the automation of gamification personalization [24]. This article contributes to this vein by introducing a free RS for personalized gamification that can be both plugged-in gamified systems to automate their personalization process, as well as independently used as a guide for defining personalized gamification designs. As this system is built upon the findings from this article, it implements a state-of-the-art personalization approach, which addresses a couple of literature challenges, namely the need for considering contextual factors along with user information, as well as the interaction between all relevant characteristics (see Section II). Nevertheless, note that our sample size limits our RS, as well as the unbalance in various characteristics such as participants' countries. Therefore, while we validated our models based on standard statistical practices, their recommendations must be interpreted with caution, always analyzing ratings and having their limitations in mind.

A. Implications

There are five main implications of our findings. First, demographics and gaming-related characteristics are moderators of user preference that should be prioritized differently. We showed that these characteristics do affect user preference but that each one's importance differs from one to another. Additionally, those exploring gamification effectiveness might rely on our results to define which data to capture from their samples to further assess whether these characteristics also play a role in other aspects (e.g., motivation or learning from interacting with GES).

Second, personalization approaches should be expanded beyond the user. We have shown that the game elements people prefer when expecting to perform one LAT differ from what they prefer when expecting to perform another; similarly for users who live in different countries. These findings' implication is empirical evidence that rather than just thinking on what users generally prefer, aspects of the task that will be performed and the user's geographic location should be taken into account, supporting recent literature arguments [6], [18], [24], [35].

Third, the interaction between relevant characteristics cannot be ignored. Our results demonstrated that the game elements preferred the most are likely to change when a single characteristic (e.g., country) changes. For example, we demonstrated that the recommended game elements for the same LAT will differ for Brazilian and American users, even if all other characteristics are the same. Thus, confirming the need for tailoring gamification designs not only to the user but also to the context [18], [35] as well as considering the interaction between different aspects [6], [22]. Hence, the implication is that only one side of the whole is likely not to work in full potential.

Fourth, when designing GES, two people might prefer the same game elements, but with different priorities. When surveying participants, we asked them to rank the top three game elements that would help them the most in learning activities of a specific type. Hence, gathering data able to inform not only which game elements are the most preferred on each occasion, but also the importance order of the selected elements. Thus, we imply that when relying on our findings to design GES, one should define the emphasis each game element will receive based on users' selection order (see Section IV-C) because despite different individuals might prefer the same game elements set, they might prefer those with different priorities.

Lastly, putting together our findings and analyses, one can use our RS (see the appendix) to automate gamified systems' personalization process as well as be informed on how to tailor gamification designs of educational systems. Practitioners can exploit our RS to define their systems' gamification designs, as well as researchers can apply its recommendations on their studies to assess the effectiveness of users' preferred designs. To aid those interested in using our RS, we have made it freely available for use and briefly discussed how it can be either incorporated into an existing system as well as using it as a guide. Thus, to the best of our knowledge, these findings pose a direct implication to the design and development of GES as it offers the first technological for personalization of gamification. Nevertheless, one should not disregard the RS's limitations, which are consequences from our sample size and characteristics and must be taken into account when using it.

B. Limitations

This section discusses our study limitations. Concerning the survey: It presented a description for each LAT and each game element to avoid misinterpretations and, consequently, guarantee answers reliability. However, this likely increased the complexity of understanding it as well as the time required to complete the survey, possibly contributing to tiring the participants throughout the process. To address this limitation, we adopted the rank-based design, which reduced the number of items, and added an attention question, which allowed us to discard inconsistent answers.

Concerning the sample: Some participants' attributes were highly unbalanced. For instance, 71% of the participants are from the United States, and many countries have less than 10 participants. Consequently, the external validity of recommendations for groups with a small presence in the dataset (e.g., those from countries other than the United States) is substantially affected. We addressed that limitation using CDT, which decide how to split countries (e.g., having recommendations for each country or for a group of countries; see Figure 1) based on statistical significance, which handles the different subsample sizes to some extent. Thus, despite sample size limitations do affect our findings' external validity, our recommendations were built to control for those.

Concerning recommendations' effectiveness: This was an exploratory, preference-based research, following a methodology commonly adopted by related research. Consequently, as in previous research, we cannot ensure that personalizing to users' preferences will be effective. However, given the number of game elements to be considered (21) as well as LAT (six), thousands of combinations would have to be tested in user studies, which is unfeasible. Our survey-based study addresses this limitation by presenting a valuable first step in suggesting which game elements to use for specific conditions and providing guidance for future studies to test our preference-based recommendations. Furthermore, whereas our survey-based recommendations have to be empirically tested, such recommendations would only be possible after data collection. Therefore, our RS helps to address the cold-starting problem [26] while it can be enhanced with real usage data in the future.

Nevertheless, readers must consider that the unbalanced features from our sample limits the generalization from our recommendations. To address that, we used CDT, which handle such variations through its statistical approach. Consequently, that led to the limitation of validating and evaluating our models based on statistical rather than standard machine learning approaches. Hence, while we used a well-established approach according to our goal, we call for future research to extend our contribution based on larger, more heterogeneous samples.

Concerning the RS's input: Although we selected the revision of Bloom's taxonomy due to its relevance within the education context, the lack of a systematic selection process also limits our findings in terms of how our study interprets LAT. Also, as our recommendations are based on averages, it might be that it will not work for some users. Lastly, although our RS is a ready-to-use resource, it is a plug-in in its initial version that can be further enhanced to improve, for instance, its compatibility with other systems, as well as its presentation for independent use.

VI. Final Remarks

Personalization emerged as an alternative to improve gamification effectiveness. Most studies in this field exploit user profiles to tailor the gamified designs. Hence, they ignore the fact that, besides the user, tasks and domain play a significant role in gamification's success. Additionally, studies often do not consider the interaction between multiple relevant characteristics, neither offer concrete resources to help in automating gamification personalization. To address these gaps, this article introduced a preference-based RS that suggests game elements tailored to the user (demographics and game-related characteristics) and the context (LAT - tasks - and geographic location), focused on the educational domain. This RS considers the interaction between its inputs and is freely available for anyone to use it.

Thus, our contributions are twofold. First, we provided practitioners with a ready-to-use resource able to guide them on how to design GES that are tailored to users' characteristics, as well as geographic location, according to the tasks they will perform. Second, we expanded the literature on how to tailor gamification designs to any learning activity (based on its type) by presenting recommendations that might be empirically tested in future research, providing empirical evidence on which demographics and game-related user characteristics impact their preferences, as well as which one is more important than the others, and supporting literature suggestions by showing that LAT and geographic location do affect user preference.

As future studies, we mainly recommend validating the effectiveness of our RS recommendations (e.g., ability to improve user motivation, flow, academic performance, or learning gains), compared to one-size-fits-all and other personalization methods, to identify whether personalizing to users' preferences will positively impact them as expected. Another line of future research is expanding our RS, especially because of our sample restrictions due to some unbalanced attributes, with more heterogeneous samples. Additional improvements are transforming it into a service to mitigate compatibility problems as well as the need for manually adding the code to the project. Additionally, future studies might tackle the limitation of not assessing the match between all game elements and all LAT from our methodology, which might be accomplished in steps (e.g., assessing one LAT per experiment) to cope with the complexity of testing all at once.

Appendix A

Appendixes are available at: https://osf.io/3a2fd/.

Acknowledgment

The authors would like to thank the funding provided by CNPq (141859/2019-9; 163932/2020-4; 308458/2020-6), CAPES (Finance code - 001), and FAPESP (Projects: 2018/07688-1; 2018/15917-0; 2013/07375-0).

References

[1] K. M. Kapp, *The gamification of learning and instruction*. Wiley San Francisco, 2012.

[2] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining "gamification"," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, ser. MindTrek '11, Tampere, Finland, Sep. 28-30, 2011, p. 9–15. [Online]. Available: 10.1145/2181037.2181040

[3] M. Sailer and L. Homner, "The gamification of learning: a meta-analysis," *Educational Psychology Review*, Aug 2019. [Online]. Available: https://doi.org/10.1007/s10648-019-09498-w

[4] A. M. Toda, P. H. D. Valle, and S. Isotani, "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*, A. I. Cristea, I. I. Bittencourt, and F. Lima, Eds. Cham: Springer International Publishing, 2018, pp. 143–156. [Online]. Available: 10.1007/978-3-319-97934-2_9

[5] B. Morschheuser, L. Hassan, K. Werder, and J. Hamari, "How to design gamification? a method for engineering gamified software," *Information and Software Technology*, vol. 95, pp. 219–237, Mar 2018. [Online]. Available: https://doi.org/10.1016/j.infsof.2017.10.015

[6] D. Liu, R. Santhanam, and J. Webster, "Toward meaningful engagement: A framework for design and research of gamified information systems," *MIS quarterly*, vol. 41, no. 4, pp. 1011–1034, Jun 2017. [Online]. Available: 10.25300/MISQ/2017/41.4.01

[7] L. Rodrigues, A. M. Toda, P. T. Palomino, W. Oliveira, and S. Isotani, "Personalized gamification: A literature review of outcomes, experiments, and approaches," in *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, ser. TEEM'20, Salamanca, Spain, Oct. 21-23, 2020, p. 699–706. [Online]. Available: https://doi.org/10.1145/3434780.3436665

[8] G. F. Tondello, "Dynamic personalization of gameful interactive systems," Ph.D. dissertation, University of Waterloo, 2019.

[9] N. Yee, "Gaming motivations align with personality traits," jan 2016, accessed in January 2020. [Online]. Available: quanticfoundry.com/2016/01/05/personality-correlates/

[10] L. Rodrigues and J. D. Brancher, "Improving players' profiles clustering from game data through feature extraction," in *Proceedings of SBGames 2018 - Computing Track*. Foz do Iguaçu, Brazil: IEEE, Oct. 29-01, 2018, pp. 177–186. [Online]. Available: 10.1109/SBGAMES.2018.00029

[11] R. Orji, J. Vassileva, and R. L. Mandryk, "Modeling the efficacy of persuasive strategies for different gamer types in serious games for health," *User Modeling and User-Adapted Interaction*, vol. 24, no. 5, pp. 453–498, Jul. 7-11, 2014. [Online]. Available: 10.1007/s11257-014-9149-8

[12] A. Knutas, R. Van Roy, T. Hynninen, M. Granato, J. Kasurinen, and J. Ikonen, "A process for designing algorithm-based personalized gamification," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13 593–13 612, Dec 2019. [Online]. Available: 10.1007/s11042-018-6913-5

[13] L. Rodrigues, A. Toda, W. Oliveira, P. Palomino, and S. Isotani, "Just beat it: Exploring the influences of competition and task-related factors in gamified learning environments," in *Proceedings of the Brazilian Symposium of Computers on Education*, Porto Alegre, Brasil, Nov. 24-28, 2020, pp. 461–470. [Online]. Available: 10.5753/cbie.sbie.2020.461

[14] J. Koivisto and J. Hamari, "The rise of motivational information systems: A review of gamification research," *International Journal of Information Management*, vol. 45, pp. 191–210, Apr 2019. [Online]. Available: 10.1016/j.ijinfomgt.2018.10.013

[15] G. F. Tondello, A. Mora, and L. E. Nacke, "Elements of gameful design emerging from user preferences," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, Amsterdam, The Netherlands, Oct. 15-18, 2017, pp. 129–142. [Online]. Available: 10.1145/3116595.3116627

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

[17] J. Baldeón, I. Rodríguez, and A. Puig, "Lega: A learner-centered gamification design framework," in *Proceedings of the XVII International Conference on Human Computer Interaction*, Salamanca, Spain, Sep. 13-16, 2016, pp. 45:1–45:8. [Online]. Available: http://doi.acm.org/10.1145/2998626.2998673

[18] S. Hallifax, A. Serna, J.-C. Marty, and É. Lavoué, "Adaptive gamification in education: A literature review of current trends and developments," in *Transforming Learning with Meaningful Technologies*, M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, and J. Schneider, Eds. Cham: Springer International Publishing, 2019, pp. 294–307. [Online]. Available: 10.1007/978-3-030-29736-7_22

[19] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The gamification user types hexad scale," in *Proceedings of the 2016 annual symposium on computer-human interaction in play*, Austin, Texas, USA, Oct. 16-19, 2016, pp. 229–243. [Online]. Available: 10.1145/2967934.2968082

[20] S. Hallifax, A. Serna, J.-C. Marty, G. Lavoué, and E. Lavoué, "Factors to consider for tailored gamification," in *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, ser. CHI PLAY '19, Barcelona, Spain, Oct. 22-25, 2019, p. 559–572. [Online]. Available: https://doi.org/10.1145/3311350.3347167

[21] A. Mora, G. F. Tondello, L. Calvet, C. González, J. Arnedo-Moreno, and L. E. Nacke, "The quest for a better tailoring of gameful design: An analysis of player type preferences," in *Proceedings of the XX International Conference on Human Computer Interaction*, Donostia Gipuzkoa, Spain, Jun. 25-28, 2019, p. 1. [Online]. Available: 10.1145/3335595.3335625

[22] P. Buckley and E. Doyle, "Individualising gamification: An investigation of the impact of learning styles and personality traits on the efficacy of gamification using a prediction market," *Computers & Education*, vol. 106, pp. 43–55, Mar 2017. [Online]. Available: 10.1016/j.compedu.2016.11.009

[23] S. Borges, V. Durelli, H. Reis, I. I. Bittencourt, R. Mizoguchi, and S. Isotani, "Selecting effective influence principles for tailoring gamification-based strategies to player roles," in *Brazilian Symposium on Computers in Education*, vol. 28, Recife, Brazil, Oct. 30-02 2017, pp. 857–866. [Online]. Available: 10.5753/cbie.sbie.2017.857

[24] A. C. T. Klock, I. Gasparini, M. S. Pimenta, and J. Hamari, "Tailored gamification: A review of literature," *International Journal of Human-Computer Studies*, vol. 144, p. 102495, Dec 2020. [Online]. Available: https://doi.org/10.1016/j.ijhcs.2020.102495

[25] K. Bovermann and T. J. Bastiaens, "Towards a motivational design? connecting gamification user types and online learning activities," *Research and Practice in Technology Enhanced Learning*, vol. 15, no. 1, pp. 1–18, Jan 2020. [Online]. Available: 10.1186/s41039-019-0121-4

[26] G. F. Tondello, R. Orji, and L. E. Nacke, "Recommender systems for personalized gamification," in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, Bratislava, Slovakia, Jul. 09-12 2017, pp. 425–430. [Online]. Available: 10.1145/3099023.3099114

[27] F. F.-H. Nah, Q. Zeng, V. R. Telaprolu, A. P. Ayyappa, and B. Eschenbrenner, "Gamification of education: a review of literature," in *International conference on hci in business*, Heraklion, Greece, Jun 22-27 2014, pp. 401–409. [Online]. Available: 10.1007/978-3-319-07293-7_39

[28] A. M. Toda, W. Oliveira, A. C. Klock, P. T. Palomino, M. Pimenta, I. Gasparini, L. Shi, I. Bittencourt, S. Isotani, and A. I. Cristea, "A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation," in *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, vol. 2161-377X, Maceió, Brazil, Jul. 15-18 2019, pp. 84–88. [Online]. Available: 10.1109/ICALT.2019.00028

[29] A. M. Toda, A. C. Klock, W. Oliveira, P. T. Palomino, L. Rodrigues, L. Shi, I. Bittencourt, I. Gasparini, S. Isotani, and A. I. Cristea, "Analysing gamification elements in educational environments using an existing gamification taxonomy," *Smart Learning Environments*, vol. 6, no. 1, p. 16, Dec 2019. [Online]. Available: 10.1186/s40561-019-0106-1

[30] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of personality*, vol. 60, no. 2, pp. 175–215, Jun 1992. [Online]. Available: 10.1111/j.1467-6494.1992.tb00970.x.

[31] M. Denden, A. Tlili, F. Essalmi, and M. Jemni, "An investigation of the factors affecting the perception of gamification and game elements," in *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*, Muscat, Sultanate of Oman, Dec. 19-21 2017, pp. 1–6. [Online]. Available: 10.1109/ICTA.2017.8336019

[32] A. Toda, W. Oliveira, L. Shi, I. Bittencourt, S. Isotani, and A. Cristea, "Planning gamification strategies based on user characteristics and dm: A gender-based case study," in *12th International Conference on Educational Data Mining*, Montreal, Canada, Jul. 02-05, 2019, pp. 438–443.

[33] A. M. Toda, R. M. do Carmo, A. P. da Silva, I. I. Bittencourt, and S. Isotani, "An approach for planning and deploying gamification concepts with social networks within educational contexts," *International Journal of Information Management*, vol. 46, pp. 294 – 303, Jun 2019. [Online]. Available: https://doi.org/10.1016/j.ijinfomgt.2018.10.001

[34] S. Deterding, "The lens of intrinsic skill atoms: A method for gameful design," *Human–Computer Interaction*, vol. 30, no. 3-4, pp. 294–335, May 2015. [Online]. Available: 10.1080/07370024.2014.993471

[35] L. Rodrigues, W. Oliveira, A. Toda, P. Palomino, and S. Isotani, "Thinking inside the box: How to tailor gamified educational systems based on learning activities types," in *Proceedings of the Brazilian Symposium of Computers on Education*, Brasilia, Brazil, Nov. 11-14 2019, pp. 823–832. [Online]. Available: 10.5753/cbie.sbie.2019.823

[36] R. E. Wood, "Task complexity: Definition of the construct," *Organizational Behavior and Human Decision Processes*, vol. 37, no. 1, pp. 60–82, Feb 1986. [Online]. Available: https://doi.org/10.1016/0749-5978(86)90044-0

[37] D. Diaper and N. Stanton, *The handbook of task analysis for human-computer interaction*. CRC Press, 2003.

[38] D. R. Krathwohl, "A revision of bloom's taxonomy: An overview," *Theory into practice*, vol. 41, no. 4, pp. 212–218, Jun 2002. [Online]. Available: 10.1207/s15430421tip4104_2

[39] B. S. Bloom, *Taxonomy of educational objectives: The classification of educational goals*. New York: Longman, 1956.

[40] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 1–35. [Online]. Available: 10.1007/978-0-387-85820-3_1

[41] D. Dicheva, K. Irwin, and C. Dichev, "Motivational factors in educational gamification," in *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, Mumbai, India, Jul. 09-13, 2018, pp. 408–410. [Online]. Available: 10.1109/ICALT.2018.00102

[42] Y. Xu and Y. Tang, "Based on action-personality data mining, research of gamification emission reduction mechanism and intelligent personalized action recommendation model," in *Cross-Cultural Design Methods, Practice and Impact*, P. Rau, Ed. Cham: Springer International Publishing, 2015, pp. 241–252. [Online]. Available: 10.1007/978-3-319-20907-4_22

[43] B. Monterrat, É. Lavoué, and S. George, "Toward an adaptive gamification system for learning environments," in *Computer Supported Education*, S. Zvacek, M. T. Restivo, J. Uhomoibhi, and M. Helfert, Eds. Cham: Springer International Publishing, 2015, pp. 115–129. [Online]. Available: 10.1007/978-3-319-25768-6_8

[44] J. Lazar, J. H. Feng, and H. Hochheiser, *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[45] T. Hothorn, K. Hornik, and A. Zeileis, "Unbiased recursive partitioning: A conditional inference framework," *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 651–674, Jan 2006. [Online]. Available: https://doi.org/10.1198/106186006X133933

[46] S. Bai, K. F. Hew, and B. Huang, "Does gamification improve student learning outcome? evidence from a meta-analysis and synthesis of qualitative data in educational contexts," *Educational Research Review*, vol. 30, p. 100322, Jun 2020. [Online]. Available: https://doi.org/10.1016/j.edurev.2020.100322

[47] R. N. Landers, K. N. Bauer, R. C. Callan, and M. B. Armstrong, *Psychological Theory and the Gamification of Learning*. Cham: Springer International Publishing, 2015, pp. 165–186. [Online]. Available: https://doi.org/10.1007/978-3-319-10208-5_9

[48] K. Casler, L. Bickel, and E. Hackett, "Separate but equal? a comparison of participants and data gathered via amazon's mturk, social media, and face-to-face behavioral testing," *Computers in human behavior*, vol. 29, no. 6, pp. 2156–2160, Nov 2013. [Online]. Available: 10.1016/j.chb.2013.05.009

[49] R. N. Landers and T. S. Behrend, "An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples," *Industrial and Organizational Psychology*, vol. 8, no. 2, pp. 142–164, Mar 2015. [Online]. Available: 10.1017/iop.2015.13

[50] C. Wohlin, P. Runeson, M. Hst, M. C. Ohlsson, B. Regnell, and A. Wessln, *Experimentation in Software Engineering*. Springer Publishing Company, Incorporated, 2012.

[51] S. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660–674, May 1991. [Online]. Available: 10.1109/21.97458

[52] J. Mingers, "Expert systems—rule induction with statistical data," *Journal of the Operational Research Society*, vol. 38, no. 1, pp. 39–47, Dec 1987. [Online]. Available: 10.1057/jors.1987.5

[53] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp. 1–13, Jul 2007. [Online]. Available: 10.4018/jdwm.2007070101

[54] X. Amatriain and J. Basilico, "Past, present, and future of recommender systems: An industry perspective," in *Proceedings of the 10th ACM Conference on Recommender Systems*, Boston, Massachusetts, USA, Sep. 15-19, 2016, p. 211–214. [Online]. Available: 10.1145/2959100.2959144

[55] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC, 2003.

[56] M. Khemaja and F. Buendia, "Building context-aware gamified apps by using ontologies as unified representation and reasoning-based models," in *Serious Games and Edutainment Applications*. Springer, 2017, pp. 675–702. [Online]. Available: 10.1007/978-3-319-51645-5_29

[57] A. M. Toda, P. T. Palomino, W. Oliveira, L. Rodrigues, A. C. Klock, I. Gasparini, A. I. Cristea, and S. Isotani, "How to gamify learning systems? an experience report using the design sprint method and a taxonomy for gamification elements in education," *Journal of Educational Technology & Society*, vol. 22, no. 3, pp. 47–60, Dec 2019.

[58] R. N. Landers, G. F. Tondello, D. L. Kappen, A. B. Collmus, E. D. Mekler, and L. E. Nacke, "Defining gameful experience as a psychological state caused by gameplay: Replacing the term 'gamefulness' with three distinct constructs," *International Journal of Human-Computer Studies*, vol. 127, pp. 81–94, Jul 2019. [Online]. Available: 10.1016/j.ijhcs.2018.08.003

Luiz Rodrigues Luiz Rodrigues is Lecturer at Faculty of Industries - SENAI Londrina and a Ph.D. candidate at the University of São Paulo and a Lecturer. His research focuses on understanding and improving gamification effectiveness within the context of educational systems. His main research interests are gamification, personalization, and user modeling.

Armando M. Toda Armando Toda is, currently, a PDRA at Durham University and University of Sao Paulo, working with gamification design in educational contexts. Other main areas of research are concerned with: data mining, serious games, and human-computer interaction.

Wilk Oliveira Wilk Oliveira is a Ph.D. candidate at the University of São Paulo and MSc. in Computer Science for the Federal University of Alagoas with exchange programs at the University of Saskatchewan. He was a guest lecturer at the Tiradentes University Center. His main topics include Artificial Intelligence in Education, Human-computer Interaction, and Gamification.

Paula T. Palomino Paula Palomino is a PhD. candidate at University of São Paulo, researching and developing a gamification content-based framework for educational purposes. Other areas of study include Games, HCI and Cyberculture.

Julita Vassileva Julita Vassileva is Professor at the Department of Computer Science, University of Saskatchewan. She researches Social Computing, Persuasive Technologies, Personalization, User Modeling, AI in Education, Peer-to-Peer systems, Trust and Reputation Mechanisms, and Human-computer Interaction. Her interests are Personalized Persuasive Systems, Learning Data-Driven Persuasion, and Distributed Ledgers for Storing and Sharing Personal Data.

Seiji Isotani Seiji Isotani is a Full Professor of Computer Science and Learning Technology at the University of São Paulo (USP). His research career has been devoted to the conception, design, development, testing and deployment of intelligent and interactive educational systems. He has published over a 100 scientific articles, books, and book chapters on gamification, educational technology, ontological engineering, artificial intelligence in education (AIED), among other topics.