













# GARFIELD: A Recommender System to Personalize Gamified Learning

Luiz Rodrigues<sup>1</sup> , Armando Toda<sup>1,2</sup> , Filipe Pereira<sup>2,6</sup> ,  
Paula T. Palomino<sup>1</sup> , Ana C. T. Klock<sup>3</sup> , Marcela Pessoa<sup>4</sup> ,  
David Oliveira<sup>4</sup>, Isabela Gasparini<sup>5</sup> , Elaine H. Teixeira<sup>4</sup> ,  
Alexandra I. Cristea<sup>2</sup> , and Seiji Isotani<sup>1</sup> 

<sup>1</sup> University of São Paulo, São Carlos, Brazil  
lalrodrigues@usp.br

<sup>2</sup> Durham University, Durham, UK

<sup>3</sup> Tampere University, Tampere, Finland

<sup>4</sup> Federal University of Amazonas, Manaus, Brazil

<sup>5</sup> Santa Catarina State University, Joinville, Brazil

<sup>6</sup> Federal University of Roraima, Boa Vista, Brazil

**Abstract.** Students often lack intrinsic motivation to engage with educational activities. While gamification has the potential to mitigate that issue, it does not always work, possibly due to poor gamification design. Researchers have developed strategies to improve gamification designs through personalization. However, most of those are based on theoretical understanding of game elements and their impact on students, instead of considering real interaction data. Thus, we developed an approach to personalize gamification designs upon data from real students' experiences with a learning environment. We followed the CRISP-DM methodology to develop personalization strategies by analyzing self-reports from 221 Brazilian students who used one out of our five gamification designs. Then, we regressed from such data to obtain recommendations of which design is the most suitable to achieve a desired motivation level, leading to our interactive recommender system: GARFIELD. Its recommendations showed a moderate performance compared to the ground truth, demonstrating our approach's potential. To the best of our knowledge, GARFIELD is the first model to guide practitioners and instructors on how to personalize gamification based on empirical data.

**Keywords:** Tailored gamification · Data-driven · Education · e-Learning

## 1 Introduction

Intrinsic motivation (IM) is a strong predictor of learning gains [6]. In this regard, gamification is one method with strong potential to improve motivational learning outcomes [8]. However, gamification's effect might vary from person to person, leading to adverse effects (e.g., demotivation) for some people [6]. Research

shows that if gamified designs are not tailored to users and contexts, they are likely to not achieve their full potential, which encourages studies on how to tailor gamification [2].

Most often, gamification is tailored through personalization: designers or the system itself change the gamified design according to predefined information [11], such as changing the game elements according to the learning task. However, personalization demands a user/task model, such as those developed in [11] and [7]. In common, those and similar models are based on *potential* experiences: they were built from data captured through surveys or after seeing mock-ups [2]. Thereby, they are limited because *potential* experiences might not reflect *real* experiences [4]. For instance, [9] developed a model based on both learners' profiles and motivation before using the gamification, but with no information of learners' real experiences (e.g., after actually using gamification). Hence, to the best of our knowledge, there is no data-driven model, based on users' real (instead of potential) experiences, for personalizing gamification designs.

To address that gap, this paper presents GARFIELD - Gamification Automatic Recommender for Interactive Education and Learning Domains, a recommender system for personalizing gamification built upon data from real experiences. Our goal was to indicate the most suitable gamification design according to students' intrinsic motivation due to its positive relationship with learning [6]. For this, we followed a two-step reverse engineering approach: we collected self-reports of users' intrinsic motivations from *actually* using a gamification design, then, regressed from such data ( $N = 221$ ) to obtain recommendations of which design is the most suitable to achieve a desired motivation level given the user's information. To the best of our knowledge, GARFIELD is the first model that guides practitioners and instructors on how to personalize gamification based on empirical data from *real* usage. Therefore, this paper contributes by creating and providing a motivation-based model for personalizing gamification, informing educators on how to personalize their gamified practices and researchers by performing a first step towards developing experience-driven models for designing gamification.

## 2 Method: CRISP-DM

Because we had an apriori goal, we followed the CRISP-DM reference model, which is suggested for goal-oriented projects [14]<sup>1</sup>.

CRISP-DM's first phase is **business understanding**. In this phase, we first defined the project's goal: creating a model based on students' intrinsic motivation captured after real system usage to allow the personalization of gamified educational systems. Additionally, we defined two requirements: i) the model must consider user characteristics and ii) the model must be interactive. The former is based on research showing users characteristics affect their experiences with gamified systems [6, 11]. The latter aims to facilitate practical usage.

---

<sup>1</sup> For transparency, this link details our dataset and all analysis: [osf.io/nt97s](https://osf.io/nt97s).

The second phase is **data understanding**. Openly sharing data extends a paper's contribution because it enables cheaper, optimized exploratory analyses [13] and is especially valuable for educational contexts wherein data collection is expensive. Accordingly, we opted to work with a dataset collected and made available by [5]. This dataset has data from students enrolled in STEM undergraduate courses of three Brazilian northwestern universities (ethical committee approval: 42598620. 0.0000.5464). Students self-reported their motivations to complete in-lecture assessments after using one of the following gamified designs: i) points, acknowledgments, and competition (PBL)<sup>2</sup>, ii) acknowledgments, objectives, and progression (AOP), iii) acknowledgments, objectives and social pressure (AOS), iv) acknowledgments, competition, and time pressure (ACT), and v) competition, chance, and time pressure (CCT). We analyzed those designs by convenience because we used data shared by a previous study, which aimed to tailor gamification to user characteristics and learning activity type [5].

When available, each game element functions as follows. Students received *points* after completing a mission. After finishing each mission, they were *acknowledged* with a badge depending on their performances (e.g., getting all items right). Students could *compete* with each other based on a leaderboard that ranked them based on the points they made during the week. Within the leaderboard, a clock provoked *time pressure* by highlighting the time available to climb the leaderboard before the week's end. Additionally, a progress bar indicated student's *progression* within missions, a notification aimed to provoke *social pressure* by warning that peers just completed a mission, and a skill tree represented short-term *objectives* (i.e., completing 10 missions).

The third phase is **data preparation**. First, we ran *attribute selection*, choosing columns related to students' characteristics, intrinsic motivation, status, and the game elements they interacted with. Next, we proceeded to *data cleansing*, removing answers from students with less than 18 years ( $N = 1$ ) due to ethical aspects and participants that provided their motivations without using the system ( $N = 4$ ). Then, we conducted *data transformations* by: i) transforming the intrinsic motivation variable (captured through a seven-point Likert-scale using the respective subscale of the Situational Motivation Scale (SIMS) [1]) to range between zero and six to facilitate regression coefficients' interpretation; and ii) removing observations ( $N = 8$ ) from levels representing less than 5% of the dataset, unless grouping them with another level was feasible, to avoid overfitting. Additionally, we *constructed new attributes* for highly skewed continuous variables by categorizing: i) weekly playing time into whether the student plays an average of at least one hour per day or more than that; and ii) age, into those below Brazilian undergraduate STEM students average (i.e.,  $<21$ ) and those at or above it. Lastly, we analyzed the *game elements* column, our dependent variable, and found a single observation of the ACT design; we removed it, lead-

---

<sup>2</sup> We consider Badges and Leaderboards implementations of Acknowledgments and Competition, respectively [10], but use PBL to maintain the standard nomenclature.

ing to the prepared dataset featuring 221 observations (see our supplementary materials for details: [osf.io/nt97s](https://osf.io/nt97s)).

Phase four is **modeling**. Here, we used Multinomial Logistic Regression [3] through the *nnet* R package with the maximum number of iterations set to 1000 to ensure the algorithm's convergence. This form of machine learning enables working with nominal dependent variables, such as gamification designs, based on the null hypothesis significance testing framework. Hence, allowing us to evaluate coefficients' contributions to the model based on their significance. This technique works similarly to standard Logistic Regression, but comparing the dependent variable's reference value to all others. In our analysis, we defined the PBL design as the reference value because PBL is the most used gamification design in educational contexts [8]. As independent variables, we started with all of those of the prepared dataset. Additionally, because recommendations should consider how students' intrinsic motivation from using a gamification design change depending on their characteristics, our model assumes intrinsic motivation interacts with all other variables.

Phase five **evaluates** modeling alternatives to determine the best option. Here, we used recursive feature elimination with p-values as the elimination criteria because we followed the standard of working within the null hypothesis significance testing framework. As this project has an exploratory nature, we considered a 90% confidence level, following similar research (e.g., [6]). After selecting the final model, we evaluated it based on its predictions according to Cohen's Kappa and F-measure, calculated using R packages *vcd* and *caret*, respectively, because those metrics are reliable for multi-class problems wherein data is unbalanced.

### 3 Evaluation Results and Deployment

After running the Multinomial Logistic Regression, we found significant interactions between all user's characteristics and intrinsic motivation. Hence, we removed no features and defined the initial model as the final one. In evaluating the model, we found the Cohen's Kappa for the agreement between its predictions and the ground truth is 0.43. This value is significantly different from zero ( $p < 0.001$ ), with its 95% confidence interval ranging from 0.34 to 0.52, revealing a moderate agreement [12]. To further understand the model's predictions, Table 1 shows the confusion matrix along with the F-measure of each category, demonstrating the model performed the best for designs AOP and CCT. Differently, its performance for designs PBL and AOS were slightly worse. Additionally, the confusion matrix reveals the model's misclassifications (e.g., wrongly predicting AOS design should be PBL and AOP 13 and 18 times, respectively). Therefore, phase five shows the model recommends gamified designs with moderate performance, despite variations from one design to another. Thus, demonstrating its potential as well as room for improvement.

In terms of deployment, we developed GARFIELD, our interactive recommender system (access it here: [osf.io/nt97s](https://osf.io/nt97s)). Its interface receives user input and

**Table 1.** Confusion Matrix of the models predictions against the ground truth.

	PBL	AOP	AOS	CCT	Balanced Accuracy	F-measure
PBL	40	11	13	04	0.68	0.57
AOP	17	54	18	01	0.75	0.67
AOS	14	06	24	02	0.65	0.48
CCT	02	01	00	14	0.83	0.74

passes it to our model. Then, our model predicts the probability of recommending each possible design and presents it as a barplot. Accordingly, practitioners can use it to get recommendations for personalizing their gamified designs in a simple, interactive way. Thus, attending to our project's second requirement.

## 4 Discussion

Overall, our goal was to facilitate the personalization of gamification with a model that recommends a gamified design given an expected intrinsic motivation level. Additionally, we aimed that such recommendations considered user characteristics and could be used interactively. Ultimately, our recommender system - GARFIELD - achieves these goals, allowing educators to use it in an interactive, web-based way to receive design recommendations based on the aforementioned input. Thus, this research expands the literature by i) creating personalization guidelines from feedback collected after *real* experiences, in contrast to prior research that developed personalization guidelines based on potential experiences (e.g., [7,9] and ii) providing concrete, interactive recommender system unlike the conceptual tools related work has contributed (e.g., [2]).

As implications for future research, our contribution is twofold. First, the lack of data-driven strategies likely poses a challenge for researchers interested in developing similar approaches. In developing our approach, we demonstrate how one can create personalization strategies step-by-step through the CRISP-DM reference model, contributing with a concrete example that can be followed to implement data-driven personalization guidelines. Second, we understand that modeling users efficiently is challenging, especially for tasks that depend on people's subjective experiences (e.g., intrinsic motivation). In this paper, we created a model using 221 observations with inputs of self-reported intrinsic motivation and demographic characteristics (e.g., age, gender, and gaming preferences). Yet, our model yielded a moderate predictive power (Cohen's Kappa = 0.43). Thus, our results inform future research that while such information contributes to understanding which gamification design to use, we likely need additional information to personalize gamification more accurately.

In summary, with our results practitioners have technological support to help them personalize their gamified practices. This can be achieved using

GARFIELD, an interactive, ready-to-use recommender system to get design suggestions. Additionally, with this paper, researchers have a concrete guide on how to use CRISP-DM for creating data-driven personalization strategies based on real (instead of potential) experiences. Note, however, that our recommender's predictions are limited to moderate predictive power. We understand that limits its practical usage as it is. Nevertheless, to our best knowledge, GARFIELD is the first tool to provide gamification design recommendations based on real experiences. Thus, we believe it provides practitioners with a reliable starting point and paves the way for researchers to expand and improve it in future research.

**Acknowledgments.** This research received financial support from the following Brazilian institutions: CNPq (141859/2019-9, 163932/2020-4, 308458/2020-6, 308513/2020-7, and 308395 /2020-4); CAPES (Finance code - 001; PROAP/AUXPE); FAPESP (2018/ 15917-0, 2013/07375-0); Samsung-UFAM (agreements 001/2020 and 003/2019).

## References

1. Guay, F., Vallerand, R.J., Blanchard, C.: On the assessment of situational intrinsic and extrinsic motivation: the situational motivation scale (sims). *Motiv. Emot.* **24**(3), 175–213 (2000)
2. Klock, A.C.T., Gasparini, I., Pimenta, M.S., Hamari, J.: Tailored gamification: a review of literature. *Int. J. Hum.-Comput. Stud.* (2020)
3. Kwak, C., Clayton, A.: Multinomial logistic regression. *Nurs. Res.* **51** (2002)
4. Palomino, P., Toda, A., Rodrigues, L., Oliveira, W., Isotani, S.: 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, pp. 1127–1130 (2020)
5. Rodrigues, L., et al.: How personalization affects motivation in gamified review assessments, May 2022. <https://doi.org/10.17605/OSF.IO/EHM43>
6. Rodrigues, L., Toda, A.M., Oliveira, W., Palomino, P.T., Avila-Santos, A.P., Isotani, S.: 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, pp. 184–190 (2021)
7. Rodrigues, L., Toda, A.M., dos Santos, W.O., Palomino, P.T., Vassileva, J., Isotani, S.: Automating gamification personalization to the user and beyond. *IEEE Trans. Learn. Technol.* (2022)
8. Sailer, M., Homner, L.: The gamification of learning: a meta-analysis. *Educ. Psychol. Rev.* **32**(1), 77–112 (2019). <https://doi.org/10.1007/s10648-019-09498-w>
9. Stuart, H., Lavoué, E., Serna, A.: 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 (2020)
10. Toda, A.M., et al.: Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learn. Environ.* **6**(1), 16 (2019). <https://doi.org/10.1186/s40561-019-0106-1>
11. Tondello, G.F.: Dynamic personalization of gameful interactive systems. Ph.D. thesis, University of Waterloo (2019)
12. Viera, A.J., Garrett, J.M., et al.: Understanding interobserver agreement: the kappa statistic. *Fam. Med.* **37**(5), 360–363 (2005)

13. Vornhagen, J.B., Tyack, A., Mekler, E.D.: Statistical significance testing at chi play: challenges and opportunities for more transparency. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play, pp. 4–18 (2020)
14. Wirth, R., Hipp, J.: CRISP-DM: towards a standard process model for data mining. In: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, vol. 1. Springer, London, UK (2000). <https://www.tib.eu/en/search/id/BLCP:CN039162600/CRISP-DM-Towards-a-Standard-Process-Model-for-Data?cHash=16bb58abe6455a858f809fd55fb0ca8f>