Just beat it: Exploring the influences of competition and task-related factors in gamified learning environments

Luiz Rodrigues[1], Armando M. Toda[1], Wilk Oliveira[1], Paula T. Palomino[1], Seiji Isotani[1]

[1]Institute of Mathematics and Computer Science – University of São Paulo (ICMC-USP)
400 Trabalhador São-carlense Avenue São Carlos, SP 13566-590 – Brazil.

`{lalrodrigues, armando.toda, wilk.oliveira, paulatpalomino}@usp.br`

`sisotani@icmc.usp.br`

Abstract. *Understanding how each game element in isolation affects learners' motivation and contextual factors' moderator effects is needed to improve gamified interventions. Thus, this paper explored the impact of one of the most used game elements - Competition - on motivation and whether task-related contextual factors (e.g., familiarity with the task's subject) moderate that impact. In a within-subject quasi-experimental design, graduate students from an Artificial Intelligence course created a reflexive intelligent agent for a console-based fight simulator in a non-gamified condition. Then, they improved their agents to compete against their peers' agents (gamified condition). Based on motivation levels measured in both conditions, we found that Competition was positive for students and that task-related contextual factors influenced that effect. Therefore, suggesting i) Competition alone can be positive for motivation and ii) contextual moderators should be considered in defining gamified designs.*

1. Introduction

Gamification is the use of game elements outside their original context [Deterding et al. 2011]. Whereas its overall effect is positive, some studies show uncertain and negative findings (e.g., demotivation and disengagement), with gamification designs often blamed for these undesired outcomes [Koivisto and Hamari 2019]. Although scholars have presented approaches to aid that process (e.g., [van Roy and Zaman 2017, Toda et al. 2019a]), two main problems arise from this context. First, as approaches often lead to designs featuring multiple game elements, it is unclear which game element is actually causing the outcomes (positive or negative), and to what extent each one is contributing [Mekler et al. 2017]. Second, which factors moderate gamification's success are not completely defined, demanding studies to better understand pre-determinants of gamification's effectiveness [Sailer and Homner 2019].

Furthermore, in the context of gamification, the Competition element[1] is one of the most used in gamified environments and it is recommended for males [Klock et al. 2020]. Motivation is often considered the psychological outcome sought by gamification [van Roy and Zaman 2017] and context has been discussed as a moderator of gamification's success [Liu et al. 2017], with activities/tasks being a central part of the context

---

[1]As discussed by [Toda et al. 2019b], we consider Competition a game element that might be represented by, for instance, a leaderboard or player-to-player conflict.

[Savard and Mizoguchi 2019]. Hence, the purpose of this paper is to verify the effects of Competition and task-related factors on users' motivation. Given that gamification is often expected to affect user motivation, contextual data are likely to moderate that effect, and the aforementioned problems, we sought to answer (**RQ1**) *how does Competition affect learners' motivation?* and (**RQ2**) *how do task-related factors moderate Competition's impact on learners' motivation?*.

Compared to previous research, this study is different in two points. First, whereas most studies implement Competition as leaderboards (e.g., [Chan et al. 2018, Mekler et al. 2017]), we implemented it as a player-to-player (PvP) competition, which avoids having learners down in the leaderboard as the competition occurs amongst them. Second, studies evaluating Competition's impact often analyze moderators linked to users' characteristics (e.g., [Landers et al. 2019, Papadopoulos et al. 2015]), whilst we explore those related to the task, guided by literature research agenda [Hallifax et al. 2019, Rodrigues et al. 2019, Liu et al. 2017]. Thus, this study contributes by analyzing the impact of a single game element (i.e., Competition) on learners' motivation and investigating how task-related factors (i.e., those related to the gamified task) moderate that effect.

Moreover, we link gamification on practice to theory by exploiting [Pink 2011]'s theory when discussing the Competition's as well as the moderator's impacts and motivation. This author advocates intrinsic motivation is preferred over the extrinsic one. Intrinsic motivation relates to pleasure from the activity itself and extrinsic motivation relates to pleasure from what one receives in exchange for performing the activity [Pink 2011]. It might be split into external and identified regulations, the former relates to behaviors originated by interests in rewards and the latter relates to performing an activity due to its end (e.g., possible rewards) but assuming it as valuable and chosen by oneself. Amotivation relates to no intention in performing a task [Guay et al. 2000]. [Pink 2011] also proposes achieving intrinsic motivation is based on *autonomy* (i.e., freedom to do as desired), *mastery* (i.e., become better at something relevant), and *purpose* (i.e., a cause/reason).

2. Study

The goal of this study was to **analyze the impact of Competition on different motivation types and how task-related factors moderate this impact**. To achieve this goal, we performed a within-subject quasi-experimental study following the one factor with two treatments design. Treatments were a gamified condition, implemented through the Competition game element [Toda et al. 2019b], and a non-gamified condition. To compare conditions, we used a within-subject (paired) design that, according to [Wohlin et al. 2012], improves the experiment precision. Thus, there was no random assignment, characterizing a quasi-experiment. We also aimed to analyze it within a real learning setting (i.e., a real class), then, by convenience sampling, we recruited students from a graduate course where we were allowed to perform the study. Artificial Intelligence (AI) was the class' topic, in which students were introduced to Python programming language, AI itself, and intelligent agents types. Subjects were 15 Brazilian males with an average age of 31 (± 8.43) years. All of them agreed to participate in the study and were in accordance with the use of their (non)personal data for research ends.

In the study task, participants worked with a console-based fight simulator, which

was developed specifically to be used as a learning resource in the course this quasi-experiment was performed. As the quasi-experiment was performed in an introductory AI class, its tasks concerned developing the simplest intelligent agent type, a reflexive agent, which is a kind of agent that chooses actions based on its current perception of the environment [Russell and Norvig 2010]. Subjects were provided with a random reflexive agent that they could use to test their agents' performance and with instructions on which actions their agents could use. Then, the quasi-experiment's tasks were accomplished in the learning resources developed for the course. At first (Task A), participants were asked to implement their agents by analyzing the environment and selecting which action to do. All other functionalities were provided (e.g., simulating the fights and updating the environment based on agents' actions). No gamification was deployed in Task A as we do not consider fighting to the random agent a competition because learners were not considering competing or in conflict with another person [Toda et al. 2019b]. Thereafter (Task B), participants were required to improve their agents compared to the version developed in Task A. As such task is unlikely to be motivating to the students, and code improvement is necessary in many cases, so improving the motivation for this task is valuable.

The condition manipulation was performed by deploying a PVP unplugged competition [Toda et al. 2020] in Task B. Choosing this game is recommended to our male sample [Klock et al. 2020]. Right after participants finished Task A and before they started Task B, subjects were warned that the result of the task would be competing with their peers (i.e., fighting against other subjects' agents). This inserts participants into a competitive environment as they readily internalize they are performing a task with a competitive end [Toda et al. 2019b]. Henceforth, adding an unplugged gamification design into the task by inciting a PVP competition, exposing all participants to conditions with and without gamification to allow us to evaluate whether the Competition game element impacts their motivation.

To measure motivation, we used the Situational Motivation Scale (SIMS) [Guay et al. 2000] because it captures subjects' motivation while performing a learning activity and assesses intrinsic, identified, external, and amotivation. Also, it is psychometrically validated and has been widely used and cited in the literature. Furthermore, unlike most related works, we evaluated task-dependent factors as possible moderators. In that sense, analyzing users' previous knowledge contributes to understanding the effect of context-related moderators as the activity is fundamental to the context [Savard and Mizoguchi 2019]. Accordingly, given that participants worked with AI in the Python programming language, we assessed their self-reported familiarity with these aspects, which were all approached in the class: AI, programming, and Python. These data were measured in five-point Likert-scales. Hence, allowing us to collect contextual data related to the task they would perform. Table 1 demonstrates the sample's percentages in each familiarity degree for each measure.

Table 1. Participants' contextual-factors (five-point Likert scale), shown as N(%).

Familiarity with	1	2	3	4	5
Programming	0 (0.00)	1 (0.07)	3 (0.20)	7 (0.47)	4 (0.27)
Python	2 (0.13)	9 (0.60)	2 (0.13)	1 (0.07)	1 (0.07)
AI	6 (0.40)	6 (0.40)	2 (0.13)	1 (0.07)	0 (0.00)

The study was conducted in six steps. **First**, learners self-grouped into groups of three to discuss the activity. We adopted this approach because the study concerns the last activity students performed at the course's first class (after roughly 6 hours of learning, without considering the intervals) and we wanted to foster collaborative learning (e.g., exchanging knowledge, opinions and/or strategies regarding the activity itself, the programming language, and the programming approaches they could adopt during the tasks), as well as get to know each other, which is facilitated by this kind of grouping intervention [Swaray 2012]. The remaining steps were accomplished by each student individually. **Second**, each participant completed the contextual info questionnaire. This was necessary to provide data to analyze the possible moderation effect of contextual, task-related factors on Competition's impact. **Third**, they had to perform Task A, which was the accomplishment of the first part (control condition) of the within-subject design[2]. **Fourth**, subjects completed the SIMS (i.e., pre-test, the first part of the paired comparison). **Fifth**, they had to perform Task B, the second and final part (experimental condition) of the within-subject design as we were analyzing two conditions only. This was the gamified task as participants performed in within the unplugged PVP competition previously introduced. **Sixth**, students completed the SIMS again (i.e., post-test, the second part of the paired comparison) right after finishing Task B. The reliability of participants' answers (Cronbach's alpha $\alpha \geq 0.7$) was adequate for all motivation types.

Lastly, for data analysis, we analyzed paired measures' difference to evaluate Competition's impact following the suggestion by [Wohlin et al. 2012]. This difference was calculated subtracting the pre- from the post-measure, which yields higher values for larger impacts, and vice-versa. Due to our reduced sample size, we chose not to perform inferential statistical tests as those would have no power to yield reliable results. Considering our quasi-experiment context (N = 15; paired), a large effect size (ES; ≥ 0.8) would be needed for a t-test to achieve the usual 0.8 power (Calculated using the *pwr* package from R) under the standard 0.05 significance level; similar for an ANOVA to assess moderations. However, gamification's effects mostly range from small to moderate [Sailer and Homner 2019], and the evaluation of moderator effects was exploratory. Therefore, we used descriptive measures (mean (M); standard deviation (SD); and ES) for data analysis, especially considering ES is sample size-independent, unlike inferential tests [Wohlin et al. 2012]. To calculate ES, we used Hedge's g, which is recommended over Cohen's d for small samples, such as in this study's case [Ellis 2010].

3. Results

Overall, subjects reported high levels of intrinsic and identified motivation (> 6), *indifferent* levels of external motivation (± 4), and low levels of amotivation (± 2), as shown in Table 2. Additionally, the impact of adding Competition to the tasks the participants performed was small (between $|0.14|$ and $|0.27|$) [Ellis 2010] for all motivation types. However, the *difference's* standard deviations in Table 2 suggest that either the motivation change was positive or negative depended on the subject. Hence, reinforcing that one gamification design for all (one-size-fits-all) is unlikely to be completely successful, demonstrating the need for understanding what factors moderate/pre-determine gamification effectiveness [Koivisto and Hamari 2019]. We approach this need by analyzing the

---

[2]Note that the subjects experienced the control condition first because once a subject experience a gamified condition, they cannot be readily *ungamified* [Thom et al. 2012].

impact of subjects' familiarity with aspects related to the activities' context, in accordance with literature suggestions [Hallifax et al. 2019, Liu et al. 2017].

Table 2. Descriptive statistics (SD: Standard Deviation) for each motivation type.

	Intrinsic		Identified		External		Amotivation	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Pre	6.10	1.08	6.03	0.97	3.63	1.52	2.25	1.42
Post	6.30	0.70	6.28	0.80	3.85	1.37	2.03	1.55
Difference	0.20	1.39	0.25	1.37	0.22	2.24	-0.22	2.41
Effect size	0.21	-	0.27	-	0.15	-	-0.14	-

Figures 1 to 4 present participants' motivation changes for intrinsic, identified, external, and amotivation, respectively, grouped by their familiarity degree for the three moderators (programming, Python, and AI). Colors are used to represent specific degrees (e.g., yellow for two) along the X-axis, independent of the moderator, and boxes represent the distribution of motivation changes for participants within that group. The Y-axis represents the extent of the change. Next, as these analyses are exploratory, we discuss findings and rise some hypotheses on possible reasons for them based on the motivation type's definition and [Pink 2011]'s theory.

For intrinsic motivation (Figure 1), our findings suggest previous general programming familiarity had the highest impact. Users with less familiarity ($<$ four) were demotivated by competition, whereas roughly all of those with higher familiarity reported increased intrinsic motivation. Similarly, those with the highest AI familiarity (three and four) presented higher intrinsic motivation gains than other subjects, although the small difference and the overlap with those that reported two. We speculate that participants' background supported them in feelings of mastery, therefore, increasing their desire to do so for no reason besides themselves. On the other hand, familiarity with Python on intrinsic motivation changes appears to have a U function. Possibly, those with some to moderate experience (two and three) felt the activity was not challenging enough (no purpose) whereas those with none (i.e., one) or more than moderate (four and five) were motivated to do so by themselves either to try to achieve or because already felt mastery.
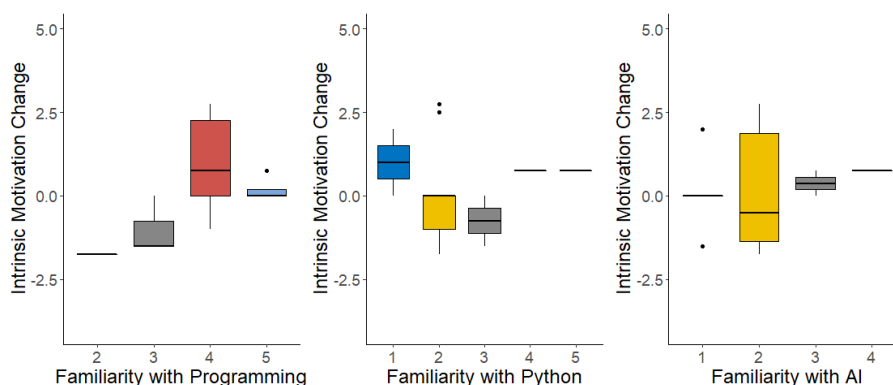

Figure 1. Intrinsic motivation changes per contextual factor.

For identified motivation (Figure 2), it is possible to note a roughly U function on Python experience as well. A possible reason is those around the middle not being

interested in the outcomes of the competition due to their moderate familiarity (e.g., no purpose), while those with the smallest and the two highest familiarity scores were concerned with either improving or showing up their skills seeking for acknowledgment (i.e., external reward; mastery). In terms of AI familiarity, a weak positive association appears to exist from this factor to identified motivation gains, in which those that reported higher familiarity also reported higher gains. Regarding programming familiarity, the results show those with high familiarity (four and five) reported positive changes whereas the others' reports demonstrated a decrease. In the last two cases, it might be that learners with more familiarity assumed they had to perform better in order to show their skills (external reward), or that they felt more confidence and interest in achieving the competition's outcomes compared to the remaining subjects (external reward; mastery).
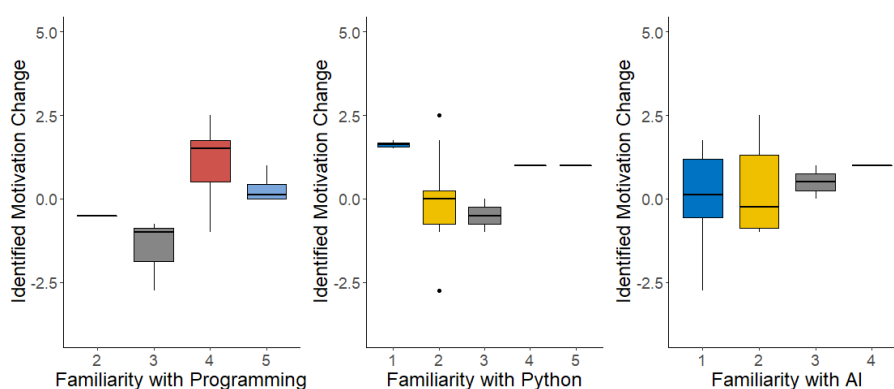

Figure 2. Identified motivation changes per contextual factor.

For external motivation (Figure 3), those with high levels (four or five) of familiarity in both Python and AI were highly demotivated. A hypothesis for that is those subjects were confident in their skills or were not afraid of possible punishments (no purpose). In addition, for AI familiarity, overall motivation practically did not change for those with moderate familiarity or less ($<$ four), which might be that the lack of this skill had no effect on learners' external regulations. For Python familiarity, from one to three it is possible to identify a positive association to motivation gain, a cue that as learners' familiarity approached a moderate point, their motivation also increased. Contrary, from programming familiarity two to four, a negative association is shown, which might be because as the more the participants were familiar with programming, the less they cared about external regulations (loss of purpose).

For amotivation (Figure 4), it appears that subjects with moderate or higher familiarity with AI were not affected by competition whereas the less their familiarity in this factor the more their amotivation decreased. This is similar to the results for Python familiarity, with exception to those in the middle point (i.e., three). Probably, Competition helped to foster interest in performing the activity for those that were not interested due to their lack of background on the previously mentioned factors (purpose; mastery). For the aforementioned exception, it might be that those subjects are not interested in competition itself (no purpose) or that their moderate familiarity led to discouragement (no mastery), therefore, demotivating them to perform the task. On the other hand, in terms of programming familiarity, those with high levels (four and five) presented decreased amotivation whereas the remaining became more amotivated with the competition, which
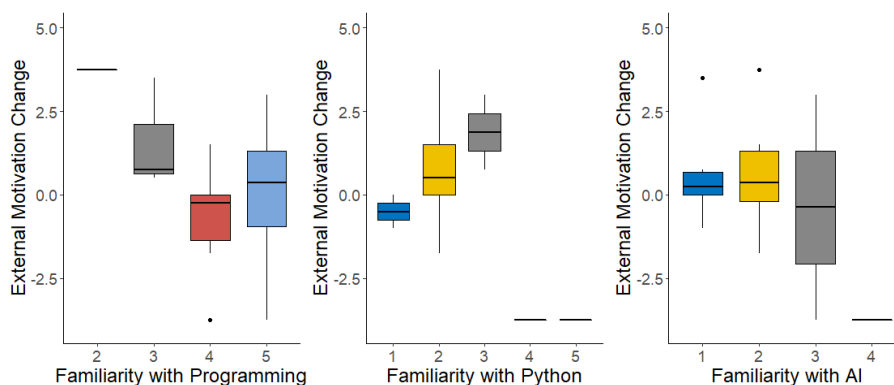
Figure 3. External motivation changes per contextual factor.

possibly emerged as those with more background were more confident whilst the others felt insecure (mastery).
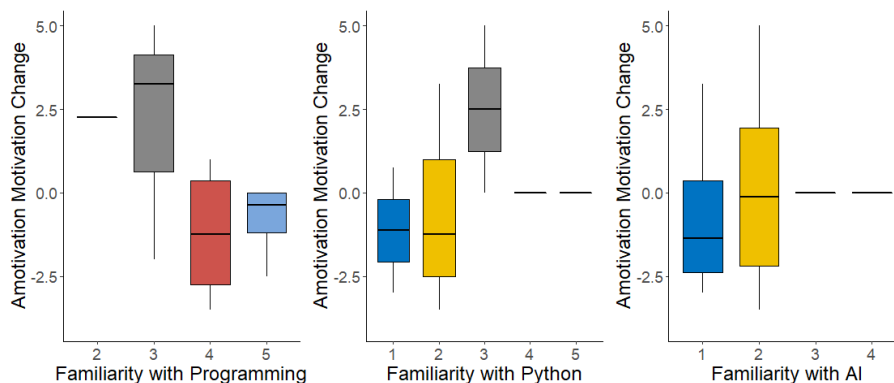

Figure 4. Amotivation motivation changes per contextual factor.

In summary, these findings suggest that Competition positively affected users' motivations, especially the intrinsic and identified ones (RQ1), and that the task-related factors influenced such effect, increasing and decreasing participants' motivation change depending on their affinity level (RQ2). Next, we further discuss these findings.

4. Discussion

According to our results, Competition has an overall small positive effect on learners' motivation, decreasing their amotivation whilst increasing intrinsic, identified, and external motivations (RQ1). This finding corroborates those of a recent meta-analysis of gamification on education, indicating gamification has overall small impacts on motivational outcomes [Sailer and Homner 2019]. Those are valuable findings for educators as the motivation types influenced the most (i.e., identified and intrinsic) are related to autonomous motivation, which is discussed as the ideal motivation type to education [van Roy and Zaman 2017].

Furthermore, although the average impact was positive, the effect on some learners was more than small (positive), for others were none, and for others was negative. These findings also corroborate indications that gamification designs should consider people's characteristics [Liu et al. 2017] as research has shown, in different contexts, the impact

of individual characteristics [Rodrigues and Brancher 2019, Hallifax et al. 2019]. Hence, reinforcing the need for providing personalized gamified interventions to improve gamification effectiveness whilst preventing it from delivering undesired outcomes.

Also, our findings tackle a perspective of gamification studies that demand research and that has increasingly attracted researchers attention: the analysis of predeterminants related to the context (operationalized as the activity in this study; RQ2) [Koivisto and Hamari 2019, Savard and Mizoguchi 2019]. We found the task-dependent factors we assessed moderated Competition's effect on learners' motivation in varied ways: AI familiarity appears to have no overall impact; more familiarity with general programming appears to slightly improve motivation, except that it decreased external motivation for those with high familiarity; and the familiarity with Python programming language showed a U function impact on autonomous motivation and, for external and amotivation, a positive effect up to moderate familiarity with a decrease thereafter.

These findings are valuable to inform both researchers and practitioners. They contribute to the knowledge on how a specific game element impacts different types of motivation, which is the main psychological aspect scholars argue gamification seeks to affect. Hence, these results can be used to inform the design of gamified interventions by suggesting that even the isolated use of competition can lead to motivation gains, as suggested by previous research [Landers et al. 2019]. Furthermore, the insights can be used to inform research on personalized gamification in terms of how Competition worked for our sample and the hypotheses we raised for context-related moderations, exploiting [Pink 2011]'s theory, can be analyzed in future studies.

4.1. Limitations and Threats to Validity

This study has some limitations and threats to validity that must be considered. Concerning the sample, our findings are based on a small, homogeneous one, which reduces findings generalization; however, it was necessary to improve the study by performing it in a real class due to the costs involved. Additionally, participants' gender might have affected gamification's effects as all participants were males and given that the Competition game element was selected accordingly. Nevertheless, Competition might not work even for males - as our findings suggest - and while participants of a single gender limit findings generalization, it reduces the possible confound of multiple genders. Concerning the intervention, participants were just introduced to the gamified (i.e., Competition) context, weakening our understanding about to what extent the improvements would last, despite it is not clear whether the novelty effect plays a role in gamified environments [Sailer and Homner 2019]. However, competition is something people are often exposed to in their daily lives, mitigating this limitation. Concerning the instruments, the gamification design was not defined based on any gamification framework because of the study goal. Rather than following a framework to choose the gamification design, we deliberately implemented an ad-hoc PVP competition [Toda et al. 2020] to identify its effects, which can inform choices of those following some framework based on the effects found on our study sample and context. Furthermore, although we selected a psychometric validated scale to measure motivation, its language (English) is different from that of the participants, which is mitigated because participants belong to a field study that often interacts with English and studied it in high school.

5. Concluding Remarks

This paper analyzed the impact of the Competition game element on the motivation of learners from a graduate AI class. Mainly, we found using Competition positively affected a specific population (which may provide some guidelines to educators, designers, and developers on how to use Competition to achieve better results) and indication that learners' familiarity to task's topic, which concerns contextual characteristics, is likely to moderate gamification's effectiveness. Hence, contributing to the scarce literature on analyzing the students' motivation based on isolated game elements, as well as how context-related factors moderate gamification's success, besides using a design that can be replicated in other studies to further analyze this and other game elements.

Acknowledgments

References

Chan, E., Nah, F. F.-H., Liu, Q., and Lu, Z. (2018). Effect of gamification on intrinsic motivation. In *International Conference on HCI in Business, Government, and Organizations*, pages 445–454. Springer.

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

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge University Press.

Guay, F., Vallerand, R. J., and Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (sims). *Motivation and emotion*, 24(3):175–213.

Hallifax, S., Serna, A., Marty, J.-C., Lavoué, G., and Lavoué, E. (2019). Factors to consider for tailored gamification. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, CHI PLAY '19, page 559–572.

Klock, A. C. T., Gasparini, I., Pimenta, M. S., and Hamari, J. (2020). Tailored gamification: A review of literature. *International Journal of Human-Computer Studies*.

Koivisto, J. and Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, 45:191–210.

Landers, R. N., Collmus, A. B., and Williams, H. (2019). The greatest battle is within ourselves: An experiment on the effects of competition alone on task performance. *International Journal of Human-Computer Studies*, 127:51–61.

Liu, D., Santhanam, R., and Webster, J. (2017). Toward meaningful engagement: A framework for design and research of gamified information systems. *MIS quarterly*, 41(4).

Mekler, E. D., Brühlmann, F., Tuch, A. N., and Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71:525 – 534.

Papadopoulos, P. M., Lagkas, T., and Demetriadis, S. N. (2015). How revealing rankings affects student attitude and performance in a peer review learning environment. In *International Conference on Computer Supported Education*, pages 225–240. Springer.

Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Penguin.

Rodrigues, L. and Brancher, J. D. (2019). Playing an educational game featuring procedural content generation: which attributes impact players' curiosity? *RENOTE*, 17(1).

Rodrigues, L., Oliveira, W., Toda, A., Palomino, P., and Isotani, S. (2019). Thinking inside the box: How to tailor gamified educational systems based on learning activities types. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 30, page 823.

Russell, S. J. and Norvig, P. (2010). Intelligent agents. In *Artificial Intelligence: A Modern Approach*. Prentice Hall Press.

Sailer, M. and Homner, L. (2019). The gamification of learning: a meta-analysis. *Educational Psychology Review*.

Savard, I. and Mizoguchi, R. (2019). Context or culture: what is the difference? *Research and Practice in Technology Enhanced Learning*, 14(1):1–12.

Swaray, R. (2012). An evaluation of a group project designed to reduce free-riding and promote active learning. *Assessment & Evaluation in Higher Education*, 37(3).

Thom, J., Millen, D., and DiMicco, J. (2012). Removing gamification from an enterprise sns. In *Proceedings of the acm 2012 conference on computer supported cooperative work*, pages 1067–1070. ACM.

Toda, A. M., do Carmo, R. M., da Silva, A. P., Bittencourt, I. I., and Isotani, S. (2019a). An approach for planning and deploying gamification concepts with social networks within educational contexts. *International Journal of Information Management*, 46:294–303.

Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., Bittencourt, I., Gasparini, I., Isotani, S., and Cristea, A. I. (2019b). Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, 6(1):16.

Toda, A. M., Palomino, P. T., Oliveira, W., Rodrigues, L., Klock, A. C., Gasparini, I., Cristea, A. I., and Isotani, S. (2020). 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 and society.*, 22(3):1–14.

van Roy, R. and Zaman, B. (2017). Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In *Serious Games and Edutainment Applications: Volume II*, pages 485–509.

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