

# Storified Quizzes in the Classroom: A 12-week Study in User Experience and Artificial Intelligence Courses

Luiz Rodrigues, SENAI University Center (UniSENAI PR) - Londrina - Brazil,  
*luiz.rodrigues@sistefiep.org.br*, *Orcid*: 0000-0003-0343-3701

Paula T. Palomino, FATEC Matão - Matão - Brazil,  
*paula.palomino@fatec.sp.gov.br*, *Orcid*: 0000-0002-9730-2253

Armando Toda, University of São Paulo (USP) - São Carlos - Brazil,  
*armando.toda@gmail.com*, *Orcid*: 0000-0003-2681-8698

Ana Klock, Tampere University - Tampere - Finland,  
*ana.tomeklock@tuni.fi*, *Orcid*: 0000-0003-3774-6511

**Abstract.** *Embedding game elements such as narrative and storytelling into learning activities is known as content gamification. Research argues content gamification might lead to improvements in psychological learning outcomes (e.g., motivation and preference). However, there is a lack of research experimenting with content gamification in real educational settings. This paper addresses this gap with a 12-week quasi-experimental study that compares content gamification to no gamification during real User Experience (UX) and Artificial Intelligence (AI) courses. Specifically, we used a narrative framework to storify classroom quizzes and analyze gamification's and contextual moderators' roles in students' learning experiences. Overall, the results indicate that content gamification did not positively affect students, and exploratory analyses suggest that AI students favored the standard quizzes. In contrast, UX students preferred the gamified quizzes, and their preferences changed over time. These findings contribute to practitioners and researchers by providing empirical evidence on how content gamification affects learning experiences in real contexts, compared to no gamification, revealing important considerations for designing, applying, and researching content gamification, especially as storified quizzes.*

**Keywords:** *Gamified Learning; Storification; Narrative; Education.*

## 1. Introduction

Gamification, the use of game elements outside games, has been widely applied to improve student engagement and motivation [Koivisto and Hamari 2019]. Studies demonstrate that gamification has a positive effect on motivational learning outcomes, such as preference and social relatedness [Sailer and Homner 2020]. However, if not well planned, gamification might lead to negative outcomes, such as reduced performance and undesired behaviors [Toda et al. 2018]. A common argument for such cases is the selection of which game elements a design features, which is important because distinct game elements are perceived and experienced differently depending on the user and the context [Koivisto and Hamari 2019].

The different effects from distinct game elements motivated calls for studying game elements in isolation [Landers et al. 2018]. Notably, despite previous research has explored several game elements, studies are mostly concerned with leaderboards, badges,

and points [Bai et al. 2020]. On the other hand, research suggests that fictional game elements (e.g., storytelling and narrative) hold great potential to improve gamification's effects [Sailer and Homner 2020]. Specifically, researchers argue that fictional game elements allow embedding game elements within the educational activity (i.e., *content gamification*) and, consequently, provide a more motivating learning experience [Kapp 2012].

However, there is a limited understanding of how content gamification impacts the learning experience. Prior research explored content gamification outside the educational domain [Wiley et al. 2020] and, in the context of learning activities, previous studies either failed to compare content gamification to non-gamified practices [Aura et al. 2023] or are limited to laboratory settings [Armstrong and Landers 2017]. Comparing gamification to non-gamified activities in real contexts is important to generate evidence on the intervention's impact, which might be different from that on laboratories [Wohlin et al. 2012]. Thus, there is a need to face the lack of experimental studies investigating the effect of content gamification on students' learning experiences, compared to non-gamified practices, within real settings.

Towards addressing this gap, this article presents a 12-week quasi-experimental study involving two higher education classes (i.e., User Experience - UX and Artificial Intelligence - AI). Throughout the 12 weeks, students completed quizzes that were either content-gamified or not and reported their preference right after finishing each quiz. Overall, this article contributes to practitioners and researchers towards fulfilling the need for empirical evidence on how content gamification affects learning experiences in real contexts compared to no gamification.

## **2. Related Work**

Prior research highlights the importance of studying game elements in isolation [Landers et al. 2018]. However, the literature often explores multiple ones simultaneously, as found in the meta-analysis by [Bai et al. 2020]. Furthermore, the same meta-analysis found that the most studied elements are points, badges, and leaderboards, in detriment of empirical investigations focused on content gamification, despite arguments on their potential [Sailer and Homner 2020].

Nevertheless, a few studies have started to investigate content gamification. For instance, [Wiley et al. 2020] conducted a laboratory study to evaluate the effects of the narrative game element on users' experiences. For this, the authors added a theme to the application, which concerned a cognitive task (i.e., the dot-probe task). Hence, unlike our goal, the study by [Wiley et al. 2020] does not concern the educational domain.

In the educational domain, a few studies have explored content gamification. [Armstrong and Landers 2017] conducted a laboratory study to evaluate how adding a narrative affects information retention. For this, they relied on a method for structuring simple stories, which was applied to standard text. Different from this study, their results are based on a controlled laboratory setting, while this article presents an experimental study conducted in a real learning setting.

[Aura et al. 2023] also conducted a study based on adding content gamification to an educational context. They relied on an approach in which they used storification to add a Happy Potter theme to the school community, aiming to create a positive, social, and immersive learning experience. Based on that context, they conducted a 10-day ethnographic

study that involved educational staff and middle school students. However, in contrast to this article, they did not experimentally compare gamification to a non-gamified setting to generate evidence from its impact [Wohlin et al. 2012].

To summarize, researchers have advocated that content-based gamification offers motivational benefits for educational activities [Sailer and Homner 2020]. However, previous studies based on such design are either limited to laboratory settings [Wiley et al. 2020, Armstrong and Landers 2017] or do not experimentally compare it to non-gamified learning experiences [Aura et al. 2023]. Thus, this article advances the literature with an experimental study conducted in a real learning setting that compares standard learning activities to content gamification. For this, we rely on a framework focused on content gamification, as discussed in Section 3.

### 3. Designing Content Gamification based on Narratives

As the gamification design is prominent for its success, researchers recommend following frameworks to define it [Koivisto and Hamari 2019]. Accordingly, numerous gamification frameworks have been proposed, but few concentrate on content-based approaches [Mora et al. 2017]. One of those few is The Gamification Journey framework, which was recently proposed and emphasizes the role of narrative and storytelling in content gamification for education [Palomino et al. 2022]. Thereby, we chose to use it because it aims to harness the full potential of content-based educational approaches by exploring immersive experiences often neglected in previous frameworks.

Specifically, a key component of The Gamification Journey framework is its Learning Journey, which is inspired by the Hero's Journey narrative structure and Bloom's Revised Taxonomy of learning objectives [Palomino et al. 2022]. Fundamentally, it is based on a learning journey divided into four quadrants, as summarized below:

- **Call to Action:** This initial quadrant focuses on the user experience (UX). It represents the student's first contact with the learning system. It aims to spark their motivation and curiosity to embark on the learning journey.
- **Trials:** In this second phase, the student enters their "special world" and begins their learning experience. The pedagogical content of this stage is designed to align with the Remembering, Understanding, and Applying levels of Bloom's Taxonomy, ensuring that students grasp foundational knowledge and concepts.
- **Transformation:** The third phase of the Learning Journey involves students transferring and applying their newly acquired knowledge in various contexts. Pedagogical content in this stage corresponds with the Analyzing, Evaluating, and Creating levels of Bloom's Taxonomy, promoting higher-order thinking and problem-solving skills.
- **Result:** The final quadrant evaluates the student's overall learning journey, considering what they have learned, their emotional experiences, and how they have changed throughout the process, emphasizing reflection on the user experience.

Overall, The Gamification Journey framework is designed to be both iterative and incremental. Accordingly, it allows educators to implement gamification strategies in blocks or modules across multiple cycles. Alternatively, it can be used as an overarching framework for an entire course, progressively introducing content as students advance through their learning journey, or for specific activities. This flexibility makes the

---

framework adaptable to various educational contexts, as well as learning objectives and activities, according to teachers' and students' needs.

#### 4. Method

To understand how content gamification affects students' learning experiences, we conducted a quasi-experimental study. Overall, the study followed a within-subject, repeated-measures design based on one factor (quiz design) with two levels: standard and storified. The within-subject approach avoids group contamination (i.e., participants from one condition leaking information to those from the other), and the repeated-measures mitigate order-effects as well as confounding effects due to differences in the topics of the quizzes [Wohlin et al. 2012].

The experimental task was completing quizzes on in-person lessons. Specifically, those were deployed in two higher education courses (User Experience - UX and Artificial Intelligence - AI) of a Software Engineering undergrad course within UniSENAI PR - Londrina. The first author taught both classes and developed all quizzes according to each lesson's plan. Note that, although the study was conducted during real lessons, students could choose not to participate in the study (i.e., they completed the activities, because they were part of the lesson, but their answers were not considered in this study). Based on that, 36 students (AI = 19; UX = 17) participated in this study. Within this group, 34 identified as men and two as women, an unbalance common to several technology-related, higher education courses. The sample's average age was 21 years.

The second author developed the storified version of all quizzes, which were used to manipulate the *quiz design* factor. In this task, the second author received the standard version of the quiz, which was independently developed by the first author; the class's lecturer. Then, the second author used the framework by [Palomino et al. 2022] to storify the quiz. To exemplify, a standard item would read *Define two data collection methods (use up to four sentences)*, while its storified version reads *You started to create your UX design portfolio to start working on some freelance job. In your resume, you would like to specify which data collection methods you know and use. Use up to four sentences to describe two of them.*<sup>1</sup> Importantly, while this framework is thought for long learning blocks/courses, we adapted it to the context of the lessons' quizzes. Accordingly, each quiz features four items, wherein each one represented one of the four quadrants of the framework by [Palomino et al. 2022], as discussed in Section 2.

Furthermore, as this study involved two classes (AI and UX), we used between-classes counterbalancing, in contrast to within classes (see Table 1). Because the study was conducted during real lectures, the aim of counterbalancing was to ensure equality among students and prevent undesirable effects on students from experimental manipulation within their lectures. Therefore, for a given quiz in one of the two classes, the same design was provided for all students. Table 1 also summarizes the schedule of the classes wherein we conducted the study, demonstrates in which weeks participants completed a quiz, and shows how the study design was employed throughout the semester.

To understand how content gamification affects students' experiences, our primary outcome was *user preference*, a psychological learning outcome that has been explored

---

<sup>1</sup>The items, originally written in Portuguese, were translated to English for presentation in the article.

**Table 1. Schedule of the classes involved in this study, along with details concerning their relationship to quiz design and data collection.**

Week	Activity	Data Collection?	AI Design	Quiz Design	UX Design
1	Class Introduction + Content 1	No	NA	NA	NA
2	Quiz 1 + Content 2	No	Storified	Standard	Standard
3	Quiz 2 + Content 3	Yes	Standard	Storified	Storified
4	Quiz 3 + Content 4	Yes	Storified	Standard	Standard
5	Quiz 4 + Content 5	Yes	Standard	Storified	Storified
6	Quiz 5 + Activity	Yes	Storified	Standard	Standard
7	Activity	No	NA	NA	NA
8	Content 6	No	NA	NA	NA
9	Quiz 6 + Content 7	No	Standard	Storified	Storified
10	Quiz 7 + Content 8	Yes	Storified	Standard	Standard
11	Quiz 8 + Content 9	Yes	Standard	Storified	Storified
12	Quiz 9 + Content 10	Yes	Storified	Standard	Standard
13	Quiz 10 + Activity	Yes	Standard	Storified	Storified

NA = Not applicable.

in prior research [Sailer and Homner 2020]. We captured it through a 9-point Likert scale (1 meaning "I totally prefer the design of last class' quiz"; 5 meaning "I'm indifferent concerning quizzes' design"; and 9 meaning "I totally prefer this class' quiz") because the study was conducted during real lessons, and we predicted completing questionnaires with multiple items would decrease students' participation. Therefore, we decided to use a single-item measure to mitigate maturation effects. As this choice might increase measurement error, we remedied it with a 9-point Likert scale [Wohlin et al. 2012]. Accordingly, we collected student feedback 8 times: one per quiz, except for quizzes 1 and 6, because students had no comparison baseline from the previous week for those.

To analyze such data, we first reverse-coded some Likert scores so that 1 always refers to preferring the standard design and 9 always refers to preferring the storified design. This was necessary because the measure asks about "last class'" and "this class'" quizzes. Hence, this transformation ensures that 9 always means preferring the storified design and 1 means preferring the standard one. Next, we normalized the 9-point Likert scale to range between -4 and 4, instead of 1 and 9, so that -4 always refers to preferring the standard quiz, 0 concerns indifference, and 4 refers to preferring the storified version. This transformation ensures that, in the context of regression analyses, an intercept significantly different from zero will indicate a preference towards some design: the standard one in case the intercept is negative, the storified otherwise [Gelman and Hill 2006].

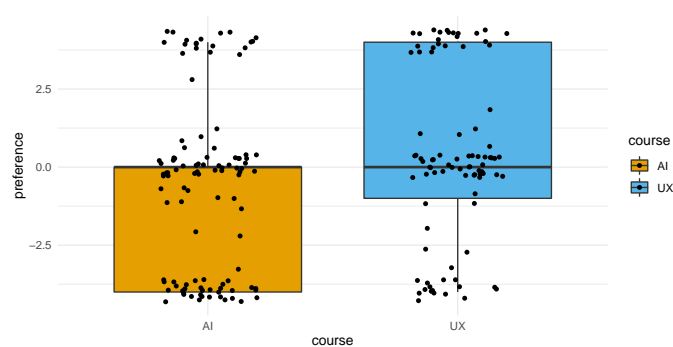
Then, we proceeded to the confirmatory data analysis. Note that our data is dependent (i.e., we have two or more data points from the same person). Standard regression analyses are not suitable for such situations, therefore, we used a multilevel regression [Gelman and Hill 2006]. According to our goal, *quiz design* was the predictor of interest, and we grouped our data by *participant id* to cope with the within-subject, repeated-measures design. To do so, we used the *lme4* and *lmerTools* R packages. Based on that, we tested the quiz design's overall effect considering the standard 95% confidence level.

Furthermore, we conducted exploratory analyses following the confirmatory ones. In those, we sought evidence to explain the confirmatory results based on prior research suggesting that gamification's effect changes depending on how long it has been used as well as its usage context [Koivisto and Hamari 2019, Rodrigues et al. 2022]. Accordingly, addressed those issues by analyzing whether student preference changes depending on usage time (i.e., number of weeks since the start of the experiment) and context (i.e., class: AI or UX). Importantly, recent literature argues that inferential statistics and p-values are likely to mislead readers when presented for exploratory analysis because readers might interpret the results as definitive conclusions (e.g., due to a p-value being smaller than 0.05). Therefore, we follow literature recommendations and present our exploratory analyses based on descriptive statistics and data visualization [Cairns 2019].

## 5. Results

Our confirmatory analysis revealed a non-significant difference in student preference towards one quiz design or another. Students' average preference was close to zero (mean,  $M = -0.241$ ; standard deviation,  $SD = 2.82$ ). Accordingly, the multilevel regression's intercept was not significantly different from zero ( $p = 0.286$ ; estimate,  $E = -0.256$ ; standard error,  $SE = 0.24$ ; Lower and Upper Confidence Interval,  $CI = [-0.728; 0.215]$ ). Hence, there is no evidence to support the hypothesis that, overall, storified quizzes had a positive effect on students' learning experiences based on their preferences. Therefore, we proceeded to the exploratory analyses to further understand this result.

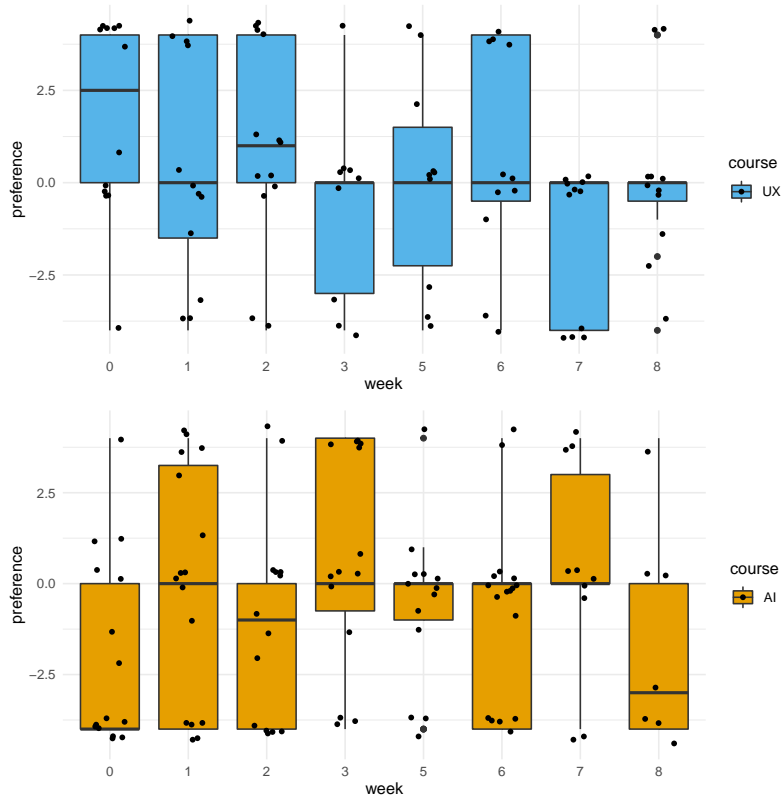
First, we analyzed whether student preference changed depending on the usage context. That is, whether the quizzes were completed in one course or another. Figure 1 summarizes this analysis results. For the AI course, the average preference was  $-0.63$  ( $SD = 2.84$ ), whereas that was  $0.22$  ( $SD = 2.74$ ) for students of the UX course. These findings suggest a difference in student preference: while AI students seemed to prefer the standard quizzes, the UX ones seemed to prefer the storified activities.



**Figure 1. Students' preferences towards storified (values greater than zero) or standard (values smaller than zero) quizzes depending on their course.**

Second, we analyzed whether student preference changed depending on how long they had been completing the quizzes. Because our previous finding suggested that AI and UX students' preferences differ, we conducted this analysis for each course separately. Figure 2 summarizes these results. For UX students, it seems the preference started favoring the storified quizzes, but decreased over time. Differently, for AI students, their

preference seems to favor the standard quizzes most of the time, with few variations towards the storified activities. These findings suggest that student preference changed over time, however, they do not reveal a clear pattern for such changes.



**Figure 2. Students' preferences towards storified (values greater than zero) or standard (values smaller than zero) quizzes depending on usage time.**

In summary, our results suggest students had no clear preference towards storified or standard quizzes. Nevertheless, our exploratory analyses revealed that context and usage time seem to play a role in this finding. It seems that AI and UX students seemed to prefer the standard and the storified quizzes, respectively; and that UX students' preferences toward the storified quizzes seemed to decrease over time, whereas that of participants enrolled in the AI course exhibited no such patterns. Importantly, readers should interpret the exploratory analyses as to-be-tested hypotheses, not as conclusive evidence.

## 6. Discussion

Surprisingly, our confirmatory analysis suggested that narrative-based gamification did not affect students' preferences. Instead, the results indicated students had no clear preference neither towards the gamified, nor the standard quizzes. On the one hand, those findings are surprising because prior research has highlighted the potential of such an approach to improve learning outcomes [Sailer and Homner 2020]. On the other hand, previous studies have also pointed out that several issues might lead gamification not to achieve the expected outcomes [Koivisto and Hamari 2019]. Based on that, we discuss three assumptions that might explain our findings.

A possible reason is how we designed the narrative-based quizzes. The literature argues that the reason for gamification not achieving the expected outcomes is its design [Toda et al. 2018]. In this study, we relied on the framework by [Palomino et al. 2022] to design the gamified quizzes, and this framework's guidance is grounded in gamification and education literature. While the framework is flexible and applicable to a wide range of learning activities, we applied its whole learning journey within each quiz. Therefore, this shortened application might have contributed to not achieving the narrative-based gamification's full potential. Thus, it might explain our unexpected findings.

Another possible explanation is the storified quizzes' formatting. The narrative-based quizzes immersed students within a story, whereas the standard ones presented straightforward statements (see Section 4). Due to those characteristics, students might have somewhat disliked the storified quizzes because they were longer and likely demanded more cognitive processing [Goldstein 2014]. Hence, this higher demand might have mitigated the narrative gamification's effect and, thus, led to our surprising findings.

A third reason for our unexpected findings relates to the several factors that moderate gamification's effect (e.g., usage time and context). Research demonstrates that understanding which factors affect gamification's outcomes, either maximizing or minimizing it, remains an open problem in the educational field [Sailer and Homner 2020]. Nevertheless, empirical studies have contributed a number of directions that shed light on this understanding. For instance, studies have found evidence that gamification's effect might both decrease and increase over time, which has been referred to as novelty and familiarization effects, respectively [Rodrigues et al. 2022].

Furthermore, research has argued that contextual factors moderate the effectiveness of gamification [Koivisto and Hamari 2019, Rodrigues et al. 2021]. Similarly, our exploratory analyses provided indications that students' preferences towards the narrative-based quizzes changed depending on their course (i.e., usage context) and how much they had been interacting with the quizzes (i.e., usage time). Thereby, possible explanations for finding unexpected results might be the gamification design not being aligned to the students' courses and its effect on their preference increasing and decreasing as time passed.

Given that context, this study contributes empirical evidence that sheds light on how narrative-based gamification, implemented in real-class quizzes, affects students' preferences. Previous research has often advocated towards the potential of such content-based gamification [Sailer and Homner 2020, Kapp 2012]. However, prior studies have not applied it to real learning contexts and experimentally compared it to standard gamification. This article contributes to that gap with a 12-week empirical study within the context of AI and UX students completing quizzes during real lessons.

Based on our contribution, this article has four main implications. First, it informs practitioners that while narrative-based quizzes might help with student immersion, one must be cautious not to increase student overload due to longer writing. Our findings also inform that the usage context and usage time are likely to influence whether storified quizzes positively affect student preference or not. Third, our results highlight the importance of carefully using frameworks to prevent adapting them to specific contexts in which the resulting design might deviate from the expectations. Lastly, those findings inform researchers on the need for further investigating narrative-based gamification's role in real



learning contexts and the need for specific resources to storify such learning activities.

Importantly, this study has limitations that must be considered. Although it was conducted during a 12-week period with two courses, the sample is limited to 36 students from the same institution. One should also consider that participants interacted with one quiz design each week, due to the real context, to prevent students of the same class from completing different quizzes. While we counterbalanced the design order based on the two courses, this choice still might affect our findings. Furthermore, we opted for a single item to measure student preference towards one design or another to better understand how the quiz designs compared to one another; and mitigate maturation effects because students completed them throughout eight lessons. We also made it based on a 9-point Likert scale to cope with measurement errors. Despite that, these choices should be considered when interpreting our findings. Finally, we reiterate that our exploratory analyses are not definitive evidence, but are meant to reveal insights for future testing. Accordingly, we backgrounded them with the relevant literature to support readers in grasping their message and exploring them in the future.

## 7. Final Remarks

Research often advocates towards embedding game elements into learning materials (i.e., content gamification) as an alternative to maximize gamification's effectiveness. However, little research has explored such an approach, and previous studies fall outside the educational domain, lack a comparison to standard learning activities, or are limited to controlled settings. Hence, this paper presents a step towards addressing that lack with a 12-week quasi-experimental study comparing content gamification to no gamification, in a real learning setting, by using a narrative framework to storify classroom quizzes.

Unexpectedly, our results indicated content gamification did not positively affect the students' learning experiences. In further analyzing these findings, we discussed important considerations, such as our adaptation of the narrative framework, the writing of the storified quizzes, and contextual moderators (e.g., usage time and class' subject). Based on that, we contribute initial evidence on how content gamification, implemented through storified quizzes, affects learning experiences in real contexts and reveal important issues for designing, applying, and researching similar interventions.

## Acknowledgments

The authors are thankful to all students who collaborated on this study. This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie [grant number 101029543, GamInclusive]

## References

- Armstrong, M. B. and Landers, R. N. (2017). An evaluation of gamified training: Using narrative to improve reactions and learning. *Simulation & Gaming*, 48(4):513–538.
- Aura, I., Hassan, L., and Hamari, J. (2023). Transforming a school into hogwarts: storification of classrooms and students' social behaviour. *Educational Review*, pages 1–20.
- Bai, S., Hew, K. F., and Huang, B. (2020). Is gamification “bullshit”? evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, page 100322.

- 
- Cairns, P. (2019). *Doing better statistics in human-computer interaction*. Cambridge University Press.
- Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- Goldstein, E. B. (2014). *Cognitive psychology: Connecting mind, research and everyday experience*. Nelson Education.
- Kapp, K. M. (2012). *The gamification of learning and instruction*. Wiley San Francisco.
- 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., Auer, E. M., Collmus, A. B., and Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, 49(3):315–337.
- Mora, A., Riera, D., González, C., and Arnedo-Moreno, J. (2017). Gamification: a systematic review of design frameworks. *Journal of Computing in Higher Education*, 29:516–548.
- Palomino, P. T., Isotani, S., and Nacke, L. P. E. (2022). Gamification of virtual learning environments: A narrative and user experience approach. Master’s thesis, Universidade de São Paulo.
- Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., Fernandes, D., Oliveira, E. H., Cristea, A. I., and Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education*, 19(1):1–25.
- Rodrigues, L., Toda, A. M., Oliveira, W., Palomino, P. T., Avila-Santos, A. P., and Isotani, S. (2021). Gamification works, but how and to whom? an experimental study in the context of programming lessons. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pages 184–190.
- Sailer, M. and Homner, L. (2020). The gamification of learning: a meta-analysis. *Educ Psychol Rev*, 32:77–112.
- Toda, A. M., Valle, P. H. D., and Isotani, S. (2018). The dark side of gamification: An overview of negative effects of gamification in education. In Cristea, A. I., Bittencourt, I. I., and Lima, F., editors, *Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions*, pages 143–156, Cham. Springer International Publishing.
- Wiley, K., Vedress, S., and Mandryk, R. L. (2020). How points and theme affect performance and experience in a gamified cognitive task. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–15.
- Wohlin, C., Runeson, P., Hst, M., Ohlsson, M. C., Regnell, B., and Wessln, A. (2012). *Experimentation in Software Engineering*. Springer Publishing Company, Incorporated.