Gamification of a BPMN Modeling Course: an Analysis of Effectiveness and Student Perception

Giacomo Garaccione, Riccardo Coppola, Luca Ardito, Marco Torchiano
first.last@polito.it
Politecnico di Torino
Torino, Italy

Abstract

Background. Gamification, i.e. the use of game-like elements in non-recreational contexts for increasing user participation and interest, has been adopted several times for Software Engineering activities for both practitioners and learners. There is, however, little evidence in the literature about applications of gamification to the discipline of Process Modeling, and even fewer experience reports are available about the use of gamified tooling in education.

Aims. Our research aims at filling this current gap by applying BIPMIN, a tool that employs gamification of Business Process Modeling Notation (BPMN), in the practical sessions of an Information Systems course, to assess whether gamification leads to improvements in the students' BPMN knowledge and their commitment in performing the exercises.

Method. All the exercises of the course were performed with a gamified tool, that embedded means to evaluate automatically the students' efforts in terms of completeness, syntax errors, and semantic errors. We analyzed the improvements over time during the course and compared the results obtained at the end of the year with those obtained by students of the previous edition, to assess whether gamification leads to higher grades.

Results. The grades of BPMN-related exercises were higher if gamified concepts were applied. Students managed to have a solid grasp of modeling fundamentals around halfway through the course, and no evident reduction in productivity or results is measured when the most difficult BPMN concepts are introduced.

Conclusions. The application of Gamification to an entire Information Systems course appears beneficial for process modeling education, with higher grades, favorable student reception, and positive learning effects as students keep interacting with the gamified tool.

CCS Concepts

• **Applied computing → Computer-assisted instruction**; **Interactive learning environments**.

Keywords

Gamification, Process Modeling, Software Engineering Education, BPMN, Software Modeling

ESEM '24, October 24–25, 2024, Barcelona, Spain

ACM ISBN 979-8-4007-1047-6/24/10
https://doi.org/10.1145/3674805.3686683

ACM Reference Format:
Giacomo Garaccione, Riccardo Coppola, Luca Ardito, Marco Torchiano. 2024. Gamification of a BPMN Modeling Course: an Analysis of Effectiveness and Student Perception. In *Proceedings of the 18th ACM / IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '24), October 24–25, 2024, Barcelona, Spain.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3674805.3686683

1 Introduction

Business Process Modeling Notation (BPMN) is a standard and widely used technique for the representation of a business process, based on flow-charting logic. The notation highlights the actors interacting with the system, and the nature and order of activities that they perform. BPMN is frequently adopted in the Software development process since it can complement functional requirements to describe in a more comprehensible way the actual flow of activities that can be performed with the system under development [24]. For this purpose, BPMN has gained attention in Software Engineering for Masters' students [23].

In recent years, Gamification (namely, the application of game design aspects to non-gaming contexts [10]) has been widely explored by Software Engineering literature as a way to increase the motivation, engagement, and participation in the activities of the software development process. Additionally, there is a vast literature about the benefits that can be provided by the use of Gamification aspects in an educational context, in terms of increased participation in class activities and students' results and grades [6]. While these aspects have been demonstrated by empirical research efforts for development and testing activities, in both professional and educational settings [20], limited evidence has been gathered regarding software modeling activities.

The objective of this work is therefore to provide evidence about the impacts of the adoption of gamified concepts to teaching Software Modeling with the use of the BPMN notation. We describe the implementation of a tool for gamified teaching of BPMN and its application for the entire duration of an Information Systems Master's course and evaluate the effects on productivity and correctness, as well as the usability of the tool.

The remainder of the paper is structured as follows: Section 2 contains a discussion regarding the state-of-the-art application of Gamification to software modeling; Section 3 describes the implementation of BIPMIN and its integration in a BPMN modeling course; Section 4 describes the experimental methodology used here as well as the threats to its validity; Section 5 presents the results of the experimentation, that are discussed in Section 6; finally, Section 7 reports the lessons learned and future research directions.

2 Background and Related Work

Several frameworks and methodologies have been proposed in recent years for the application of Gamification in educational contexts and in Software Engineering [12, 27].

A very popular framework for the definition of gamified professional and learning environments is Yu-Kai Chou's Octalysis [7], which is focused on the type of emotions that can be generated by game design mechanics on the users of a given system, rather on the results that can be obtained with them. The framework defines eight different dimensions (or core drives) for gamification mechanics: (i) *Epic Meaning and Calling*, i.e. the feeling that the users' activities belong to a bigger picture or to a greater goal; (ii) *Accomplishment*, i.e. driving the users' activities with specific goals and new elements to disclose and learn; (iii) *Empowerment*, i.e. giving the user the feeling of expressing themself and receiving feedback for their actions; (iv) *Ownership*, i.e. providing the users with goods and items that can be controlled; (v) *Social Influence*, i.e. providing interaction and connections with peers; (vi) *Scarcity*, i.e. designing elements of the gameplay that are rare by design and therefore more desirable; (vii) *Unpredictability*, i.e. ensuring that the users' experiences are always different; (viii) *Avoidance and Loss*, i.e. defining negative consequences for errors to foster the users' attention in their activities.

Gamification has a history of success in Software Engineering [9], especially in activities (e.g., software testing [14]) with limited creative aspects, where the developers' engagement might progressively degrade, potentially impacting their work's quality. Regarding Software Engineering Education, Alhammad and Moreno found through a systematic mapping that the main purpose of applying gamification in the SE field is mostly directly related to improving student engagement and, to a lesser extent, to improving student knowledge, although other targets are the application of SE best practices and socialization [1].

Several gamified instruments are described in the Software Modeling literature. Junior and Farias described [17] a reference framework for instructors to include gamified activities in their teaching activities, along with a parameterized way to evaluate UML models created by learners. The gamified activities conceptualized by the authors can refer to the Octalysis framework and include points, progress indicators, and visual feedback for the users.

Cosentino et al. describe a gamified environment for modeling in Papyrus in the form of a plugin providing game elements such as challenges, achievements and rewards, and scoring mechanisms [8]. Bucchiarone et al. describe Papygame [4], another gamified Papyrus modeling plugin intended to be integrated into the Eclipse IDE and provide a game experience with feedback, awards, progression, and points. The authors also described the gamification rules as a separate engine and API set to allow the creation of customized gamified activities and experiences.

Cagnazzo et al. described UMLegend, a tool for teaching the creation of class diagrams according to the UML language [5]. The tool implements gamified mechanics such as avatar creation, level progression, and live feedback.

Regarding BPMN modeling, no attempts are documented in the related literature about the application of gamified concepts. Studies are instead available that describe the development and utilization of serious games (i.e., games with an educational purpose [11]). While serious games share commonalities with gamified tools, the latter are not designed as games but only borrow individual mechanics from game design. Kutun et al. describe BPMN-Wheel, a team-based cooperative game where gamers can perform modeling activities while obtaining in-game currency and progressing towards a team objective [18]. Ivanova et al. describe BPMS-game, a tool to support the definition of games that promote sustainability in BPM environments [21]. Marín documents a positive experience in using a game to teach BPMN aspects to students, in the bigger context of a course where five different games were applied to different activities in Software Modeling [22].

3 The BIPMIN tool

BIPMIN has been developed as a web BPMN modeler easily accessible to students. The tool is free to use and available online[1].

3.1 Gamification Mechanics

BIPMIN offers multiple gamified elements in its current implementation, spread between mechanics that directly affect the execution of a student's exercise and mechanics that are applied separately, as a consequence of their actions. The implemented mechanics are reported in table 1, along with the related dimension in the Octalysis framework and the motivation for its implementation.

3.2 Evaluation Engine

One of the main features of BIPMIN is an evaluation engine capable of automatically evaluating a student's diagram and providing them with immediate visual feedback.

Students can submit their diagram to an evaluation that checks their solution focusing on syntactic quality (handled by the external library *bpmn-js-bpmnlint*[2]) and semantic quality. The evaluation quantifies how well a diagram adheres to the context of the exercise.

The syntactic evaluation looks for elements that present errors such as not having a name, processes that lack either a start or end event, or present multiple start events; moreover, the default rules offered by the library have been extended with support for additional syntax checks such as ensuring that only specific elements (e.g. message events, service tasks) can send messages, or identifying errors related to boundary events.

The semantic evaluation, instead, allows teachers to define, for each exercise, a set of reference solutions: whenever a student performs a check, their diagram is compared with all reference solutions, and a correctness measure is computed for each solution; the solution with the highest correctness is taken as the reference solution whose feedback is provided to the student through the gamified mechanics (e.g. correctness in the progress indicator, errors, and process parts in the list, colored diagram feedback). Examples of how the feedback appears for each error type are presented in an online appendix[3].

A reference solution is composed of the following entities:

[1]https://softeng.polito.it/bipmin. Username: UserTest - Password: !!User_Test_s100030!!
[2]https://github.com/bpmn-io/bpmnlint (last accessed on 23/04/2024)
[3]https://doi.org/10.6084/m9.figshare.25738668

Table 1: Game mechanics implemented in BIPMIN

Game Mechanic	Octalysis Core Drive	Motivation
Levels and Experience	Development and Accomplishment, Loss and Avoidance	Representing the student's learning progression in modeling, providing a negative motivation with the risk of experience loss
Avatar Customization	Ownership	Providing a way to express one's individuality, offering unlockable props that provide rewards throughout the course
Avatar Feedback (Fig. 1.1)	Loss and Avoidance	Providing a negative motivator with a progressively sadder representation of the user the more mistakes are made
Leaderboards	Accomplishment	Offering a competition mechanism where students can compare their standing with their peers, with separate ranking types offering multiple ways for comparison. A leaderboard for the top ten students and relative one showing students with similar scores avoid having a huge and possibly discouraging leaderboard
Boss Icon (Fig. 1.2)	Epic Meaning	Representing the challenge with a direct metaphor, offering an enemy to defeat. Defeating a boss by completing an exercise is treated as a significant milestone, with the icon unlocked and visible in the homepage
Indicators (Fig. 1.3)	Accomplishment, Loss and Avoidance	Displaying both the completion rate of the current exercise and the experience reward that can be completed, with student actions directly affecting the two indicators
Error Lists (Fig. 1.4)	Empowerment of Creativity	Providing direct feedback on what the student needs to improve in their current diagram
Process Parts Lists (Fig. 1.5)	Empowerment of Creativity	Providing direct feedback on the correct elements of a student's diagram, as well as on the elements that can be improved
Diagram Feedback (Fig. 1.6)	Empowerment of Creativity	Reinforcing the feedback given by the error lists by showing a direct representation on the diagram of what is correct and what is an error.
Exercise Hints (Fig. 1.7)	Accomplishment	Offering unlockable content that can help students in completing an exercise

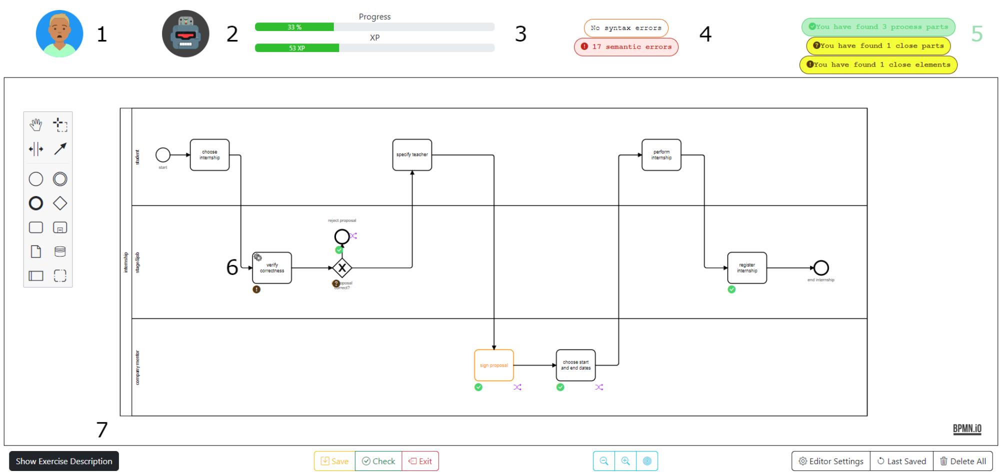


Figure 1: Exercise page of the tool showing: 1) Student Avatar, 2) Boss, 3) Indicator, 4) Error Lists, 5) Process Parts Lists, 6) Diagram Feedback, 7) Exercise Menu

- **Business Entities.** The pools that must be present in the diagram. Each business entity has its reference name and a set of allowed alternative names, and a pool in a diagram matches a business entity if its name is close enough to at least one name in the set. A business entity may also have a flag that specifies whether it must be represented with a collapsed pool or not, and a flag that specifies whether its absence in the diagram is considered an error or if it can be omitted.
- **Actors.** The set of lanes that must be present in a diagram. Actors, like business entities, have a name and a set of allowed synonyms that a lane in the diagram needs to have for the actor to be modelled correctly. Moreover, each actor specifies the business actor it belongs to, and a lane is considered correct if it has a matching name and belongs to a pool that matches the reference entity; in case a business entity does not have a matching pool in a diagram, all the actors that belong to that entity are automatically considered missing.
- **Forbidden Entities and Actors.** Sets of names that cannot be used for representing pools and lanes. Examples may be using a name intended for a pool for a single lane, using the name of an actor that is not involved with the process as a pool name, or using the name of a subsystem of the entire process as an actor.
- **Sub-processes.** The set of sub-processes that a diagram must contain. A reference sub-process specifies to which parent element (e.g. a pool or another sub-process) it belongs and whether it is an event sub-process or not with an additional constraint that specifies whether it must be interrupting or not in the first case.
- **Process Parts.** The set of different logical parts in which a process is divided. Each part can be represented by multiple

Table 2: Semantic Evaluation Algorithm Steps

Step	Description
1	**For each** Business entity in the reference solution Add an error if there is no pool with a matching name and the business entity is not optional, or if there is a collapsed pool with a matching name but the business entity cannot be collapsed (or vice versa)
2	**For each** Actor in the reference solution Add an error if there is no pool in the diagram that matches the expected business entity for the actor, if there is a matching pool but there is no lane with a matching name for the actor, or if there is a lane with a matching name that does not belong to the pool
3	**For each** Pool in the diagram that does not match a business entity Add an error if the pool has a name close to at least one forbidden business entity name
4	**For each** Lane in the diagram that does not match an actor Add an error if the lane has a name close to at least one forbidden actor name
5	**For each** Sub-process in the reference solution Add an error if there is no element in the diagram that matches the expected parent of the sub-process, or if there is such an element but no sub-process that matches the expected one exists as a child of the parent element
6	**For each** Process part in the reference solution **For each** Group of elements that can model the part Search, for each element in the group, for a diagram element that has a close name, a type equal to one of the allowed types, and belongs to the same pool as the reference element. If all group elements have a matching diagram element the group is found and the part is modeled correctly. Otherwise, check if the group of matching elements is close enough (at least half of the expected elements have a matching diagram element and the group has more than the previous close group, with the first close group having 0 elements), and take the group as the new close group in that case. The process part is found if there is a matching group of elements, close if there is a close group, and missing otherwise. Add an error for the latter two cases.
7	**For each** Process part that has been found Add for each violation found for each element of the part (e.g. being out of order in the part, wrong message exchange, being in the wrong lane) Add an error if the process part has a preceding part and none of its elements is preceded by an element of the preceding part,

alternative, equivalent groups of elements (e.g. a User Task followed by a Message Throw Event, or a User Task followed by a Service Task).

Each element inside a possible group is characterized by a set of allowed names, a set of possible element types (the same element may be modelled using a blank Task or with a UserTask, for instance), and the pool and lane it must belong to. Additional information may also be specified to define the interaction between elements such as the preceding elements in a group, the elements with which messages are exchanged, or the interaction with boundary events.

An element inside a group is considered modelled if there is an element that has a name close enough to one name of the set of allowed ones, has a type equal to one of the allowed types, and belongs to the expected pool; elements that do not satisfy the name constraint and only one of the type and pool constraints are considered *close elements* and reported directly. The other constraints are not necessary for an element to be correct, but instead represent separate errors with a lower penalty score.

The evaluation engine works by performing the operations defined in Table 2. It repeats the process above for each reference solution defined by the teachers, supporting different alternative representations of the same process in terms of both the overall high-level structure and the low-level elements of a single representation.

An important limitation of the currently implemented engine, however, is the fact that all the alternative solutions (different process representations or ways to express process parts, and synonyms for each word used in the solution) must be imagined and defined by the teacher.

3.3 Course setup and BIPMIN integration

We adapted BIPMIN in an Information Systems course held at Politecnico di Torino, where the course is part of the first-year curriculum of the Master's Degree in Engineering and Management.

The course is organized in theory lectures and practical laboratory hours where students are encouraged to solve exercises on the different course topics such as conceptual modeling, use case analysis, cost estimation, and process modeling using BPMN.

In the 2024 edition of the course, BIPMIN was used as the tool for solving the BPMN exercises of the practical labs, for a total of five exercises implemented on the platform; this marked a difference in comparison to the previous edition of the course, where students used a non-gamified tool (the Signavio Academic environment[4]) for all different exercises.

After a brief theory lecture that acted as the introduction to BIPMIN, students were given complete freedom in how and when they could tackle the exercises: the only specific deadline for the exercises was the end of the course, to allow students enrolled late to use the tool as well.

The usage of the tool was also rewarded with two additional points added to the grade of the written exam. To qualify for these two points, students were asked to perform at least one attempt on each of the six exercises before the aforementioned deadline. Moreover, we did not set a minimum quality threshold to qualify for the points, meaning that students would be rewarded even in case their attempts were not marked as *completed* by the tool, as long as their diagram showed at least some effort (i.e., submission of empty diagrams were disregarded).

The course also involved a group-based project work consisting of the analysis of a full case study: students worked on the Signavio platform for the BPMN-related analysis of the project work instead of on BIPMIN, as the former was used for theory lectures and offered the ability to share diagrams among groups of users.

4 Experimental Design

We conducted an experiment with the purpose of investigating the impact of gamification on students' learning habits, as well as to assess whether gamification leads to improvements in the students' final performances in BPMN-related tasks both at the end of the course and during the course. The research objectives are defined in Table 3 according to the Goal-Question-Metric (GQM) template.

We provide a replication package containing the data analyzed and the participants' diagrams fetched directly from the tool in the form of an online appendix[5]. Additionally, the appendix also

[4]https://signavio.com (last accessed on 21/02/2024)

[5]https://doi.org/10.6084/m9.figshare.25738668

Table 3: GQM template for the experiment

Object of Study	Learning of BPMN Modeling
Purpose	Assess the impact of gamification
Focus	Grades, Learning Progress, Commitment
Context	Modeling tools, Education
Perspective	Teachers, Learners

contains the textual description of the five exercises used during the experiment, as well as the reference solutions for each exercise.

4.1 Research Questions

We define the following Research Questions to frame the analysis of our experiment:

RQ1 Effectiveness: does gamification affect grades?
We refine the question further into:
RQ1.1 Does the application of gamification impact the students' grades in BPMN exercises for the final exam of the course?
RQ1.2: Does the application of gamification impact the students' grades in BPMN exercises for the end-of-course project work?
RQ2 Progress: does gamification support progressive learning during the course?
We expect gamification to ease a incremental acquisition of modeling skills leading to progressively increasing quality of produced models.
We refine the question further into:
RQ2.1: Does the application of gamification impact the correctness of the students' exercises during the course?
RQ2.2: Does the application of gamification impact the errors made by the students throughout the course?
RQ2.3: Does the application of gamification impact the completion rate of the exercises?
RQ3 Perception: is gamification well received by the students?
We assume gamification is positively perceived by the students with more involvement and willingness to adopt it.
We refine the question further into:
RQ3.1: How is the gamified tool perceived by the students?
RQ3.2: What are the issues and improvement areas found in the students' opinions?

4.2 Participants

The participants in the experiment were students of an Information Systems course, offered as a mandatory course in the first year of the Master's Degree in Engineering and Management at Politecnico di Torino. The course offers fundamentals of information systems such as conceptual modeling and process modeling as parts of its topics and does not require any specific prerequisite. We did not impose any kind of background knowledge on process modeling as a prerequisite for taking part in the experiment.

4.3 Operationalization of Variables

We present in this section the variables we have selected for consideration in our experiment for each of the Research Questions we defined. Independent and dependent variables are summarized in Table 4.

Table 4: Summary of variables and corresponding RQs

	Type	Variable	Value Type	RQ
RQ1	Independent	Academic year	{2023, 2024}	
	Dependent	Exam BPMN grade	numeric [0,8]	RQ1.1
		Project BPMN grade	numeric [0,5]	RQ1.2
RQ2	Independent	Exercise number	{1, 2, 3, 4, 5}	
	Dependent	Completeness	percentage	RQ2.1
		Syntax errors	count	RQ2.2
		Semantic errors	count	RQ2.3
RQ3	Dependent	TAM - usage	likert 1-5	RQ3.1
		TAM - ease of use	likert 1-5	
		TAM - intention to use	likert 1-5	
		TAM - usefulness	likert 1-5	
		GAMEX - enjoyment	likert 1-5	RQ3.1
		GAMEX - creative	likert 1-5	
		GAMEX - dominance	likert 1-5	
		GAMEX - absorption	likert 1-5	
		GAMEX - activation	likert 1-5	
		GAMEX - negative	likert 1-5	

To answer RQ1 we considered the Academic year as the independent variable, by comparing the current edition of the course, where gamification was applied, and the previous one, where a standard approach was adopted.

The dependent variables we considered for RQ1 were two separate grades: the points obtained in the written exam for BPMN-related exercises, and the points obtained by students in the BPMN exercise of the end-of-course project work. The grades obtained by the participants have been compared with the grades obtained by the students of the 2023 edition of the course.

To answer RQ2 we considered the Exercise number as the independent variable: the tool offered five exercises of increasing difficulty that were made available for completion in a predefined order at different dates during the course (i.e., students performed Exercise 1 first, then Exercise 2, and so on until reaching Exercise 5).

We measured the students' long-term learning progress by considering three metrics taken directly from their submitted exercises:

- The exercise completeness computed directly by the evaluation engine and considered as the correctness percentage of a student's diagram (number of correct elements over the total expected amount of elements) in comparison with the reference solution.
- The number of syntax errors found in the diagram.
- The number of semantic errors found in the diagram.

To answer RQ3, we considered as dependent variables the constructs offered by the TAM [19] and GAMEX [13] questionnaires; the analysis of the questionnaires consisted of a qualitative assessment of the distribution of the students' answers to identify the general opinions of the participants. Both questionnaires were used in their original form by considering all questions, with *BIPMIN* replacing the original placeholder values (the game/the application). The full questionnaire is presented in an online appendix[6].

[6]https://doi.org/10.6084/m9.figshare.25738668

4.4 Analysis Procedure

The null hypotheses we formulated to address RQ1 are:

- $H_{1.1_0}$: The application of gamification has no statistically significant impact on the grade obtained by students for the BPMN-related exercises of the final written exam;
- $H_{1.2_0}$: The application of gamification has no statistically significant impact on the grade obtained by students for the BPMN-related exercises of the end-of-course project work;

To evaluate the above hypotheses, we performed a Wilcoxon test on the distribution of the grades, considering the academic year when the exam took place as the independent variable: this means comparing the grades for an edition of the course with gamification applied with one where BPMN modeling was taught with a traditional approach. The goal of the comparison is to assess whether exam and project grades are higher in the former, that is, in the 2024 edition of the course.

The null hypotheses we formulated to address RQ2 are:

- $H_{2.1_0}$: The application of gamification does not lead to a statistically significant improvement in the correctness score obtained in the exercises throughout the course;
- $H_{2.2_0}$: The application of gamification does not lead to a statistically significant improvement in the number of syntax errors made in the exercises throughout the course;
- $H_{2.3_0}$: The application of gamification does not lead to a statistically significant improvement in the number of semantic errors made in the exercises throughout the course.

To evaluate them, we performed statistical tests using two distinct encoding methods of the variable representing the exercise:

- Fixed reference: the first exercise represents the reference level (intercept) while each following exercise corresponds to a dummy variable (1 for that exercise, 0 otherwise), therefore the coefficients (and the relative significance) represent the difference w.r.t. the first exercise.
- Progressive reference: the first exercise represent the reference level (intercept) while each following exercise corresponds to a variable that is set to 1 for that exercise and all the following ones and set to 0 for the preceding exercises; therefore the coefficients (and the relative significance) represent the difference of any exercise w.r.t. the previous one exercise.

First, we performed, for each metric, an ANOVA of a fixed effects linear model analysis using the fixed reference encoding, to assess any possible difference compared to when students first started using BIPMIN (i.e., the first exercise).

Second, we performed a similar analysis using the progressive reference encoding, to precisely identify changes in student performance over time when moving from one exercise to the next.

Since conducting multiple ($n = 8$) univariate ANOVAs may increase the chance of Type I error (false positives), we adjusted the significance level using Bonferroni correction to account for multiple comparisons; the p-value threshold we consider as statistically significant is $\alpha_C = \alpha/n = 0.00625$.

To answer RQ3.1, we administered to our students a questionnaire at the end of the course: the questionnaire was composed of the TAM and GAMEX questions: the former was used to gauge the perceived usability of BIPMIN, with four metrics focusing on Attitude towards usage (ATU), Perceived ease of use (PE), Behavioral intention to use (BI), and Perceived usefulness (PU).

The GAMEX items allowed evaluation of the gamified experience based on six different constructs: Enjoyment, Absorption, Creative Thinking, Activation, Possible Negative Effects, and Dominance. For both questionnaires, answers were given in the form of Likert-scale values ranging from 1 (Strongly Disagree) to 5 (Strongly Agree).

To answer RQ3.2, we included an open question at the end of the questionnaire which asked students to present their opinion on any issue they had encountered throughout the course when using BIPMIN, as well as how they felt we could better integrate the tool into the course activities.

The answers to this open question have been analyzed following the *open coding* approach defined by the Straussian Grounded Theory [26]: we aimed at identifying topics present in the answers, and we would then group the answers depending on the topic. Following the guidelines by Stol et al. [25] for the application of grounded theory to software engineering, we describe the process we adopted and the role distribution below.

One author analyzed each answer singularly, searching for a suitable topic among the already identified ones, assigning the topic to the answer if it was deemed suitable, or creating a new one otherwise. This process started with an empty set of topics and led to an incrementally growing set.

After having assigned a topic to all questions, the same author reviewed the answers once again to assess whether a more suitable topic among the present ones existed; in this second step, no new topic was defined.

The final step consisted of a review, performed by all four authors, of the topics and of their assignment, reaching a final consensus.

4.5 Threats to Validity

We analyzed the potential threats to the study's validity according to the four categories defined by Wohlin et al. [28].

Threats to internal validity. In the context of our experiment, there may be effects on the grades and quality of the exercise results based on the students' background and abilities, and on the relative difficulty of the different exercises that were executed. In the context of the edition 2023-2024 of the course (i.e., the gamified one) this threat was mitigated by trying to ensure a constant increase in difficulty between consecutive exercises over the year. To compare the scores between different years of the course, efforts has been made to make the exams in the different editions comparable in terms of size and difficulty of the exercise. An internal validity threat is related to the format of the project work in the two considered years: while in the year preceding gamification the project was individual, in the year with gamification the project was performed as a group work. Although the provided grades were normalized to cover this difference, this factor can have an heavy influence on the distribution of project grades among the students. As an additional confounding factor, a possible preliminary experience of students in BPMN modeling could have an influence on the grades and thereby on the results of this experiment.

The tool has been validated internally and updated after the results of a previous usability study that highlighted potential bugs

Table 5: General characteristics of the Effectiveness metrics

Type	Year	Min	Max	Mean (SD)	Median
Exam Grade	2023	0.0	8	5.83 (1.65)	6.30
Exam Grade	2024	1.6	8	6.56 (1.78)	7.00
Project Grade	2023	0.0	5	2.05 (1.20)	2.33
Project Grade	2024	2.5	5	4.45 (0.56)	4.50

and unclear elements in the GUI [3]. The prototypal nature of the tool may however lead to issues in the user experience that could have an influence on the outcome of the experiment.

Threats to external validity. Our results and the data we collected are representative only of the educational environment in the context of a Master's course. Replications of the study would be needed to validate the findings in other learning environments or even industrial contexts.

Threats to construct validity. As far as learning outcomes, we used the grades of two distinct activities. Concerning the progress, we limited these threats by referring to related literature for the selection and computation of metrics regarding the quality of the produced diagrams.

Threats to conclusion validity. All the non-parametric tests that we performed have no statistical prerequisite so they are valid for the type of analyses that we performed. The comparisons of progress metrics have assumed all exercises to have been performed in their intended order; it may be possible for students to have performed their exercises in a different order, and this variation is something that has not been considered.

5 Results

We report in this section the results of our experiment organized by Research Question.

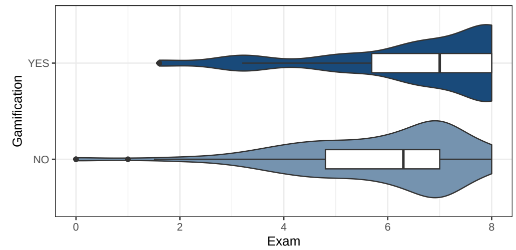
At the end of the course, a total of 264 students had performed at least one exercise on the platform: of these students, 254 had performed all five exercises offered during the course. Since the experiment design takes into account the grade obtained in the final written exam, we did not consider in our analyses the results of the students who did not pass the exam by the end of February 2024, leading to a total of 200 students. No filtering was performed on the students of the 2023 edition of the course, leading to a total of 302 grades considered against the 200 of the 2024 edition.

5.1 RQ1: Effectiveness

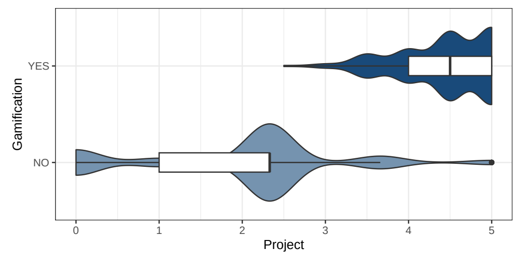
In Table 5 we report the summarized characteristics of the scores obtained by students of the 2023 and 2024 editions of the course for the BPMN-related exercises in the final exam and for the BPMN portion of the final project work.

Figure 2(a) shows combined box and violin plots for the distribution of grades depending on the academic year (2023 without gamification, 2024 with gamification). BPMN scores are on average higher for the year where gamification was adopted.

By performing a Wilcoxon test on the mean values of the exam grades, we observe that the p-value for hypothesis $H_{1.1_0}$ is statistically significant (p = 1.09e-08), meaning that we can reject the null hypothesis and state that applying gamification leads to a significant improvement in students' grades in written exams.



(a) Box and violin plot for the distribution of exam grades



(b) Box and violin plot for the distribution of project grades

Figure 2: Distribution of Effectiveness metrics

Table 6: Progress metrics average by exercise

Exercise	Completeness	Syntax Errors	Semantic Errors
1	76.32	1.08	8.04
2	72.00	0.71	7.11
3	77.40	0.57	4.42
4	77.46	1.44	5.78
5	78.73	1.24	5.64

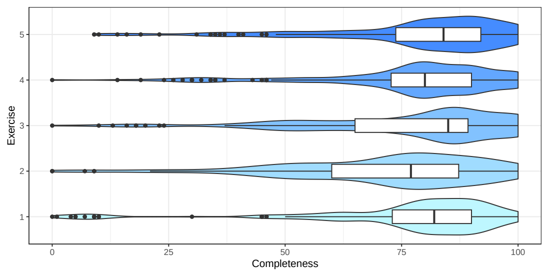
Figure 2(b) displays the combined box and violin plots related to the distribution of grades obtained by students in the BPMN part of the final project work depending on the application of gamification. Grades are, on average, much higher in the year when gamification was used.

The Wilcoxon test performed on the project grades yielded a statistically significant p-value (p = < 2.2e-16): we can thus reject the null hypothesis $H_{1.2_0}$ and affirm that gamification can lead to higher grades in the final project work.

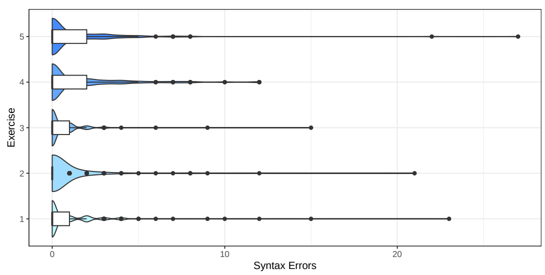
5.2 RQ2: Learning Progress

We present in Table 6 the average values of the three metrics measured to answer RQ2, for each of the five exercises performed by the students during the course.

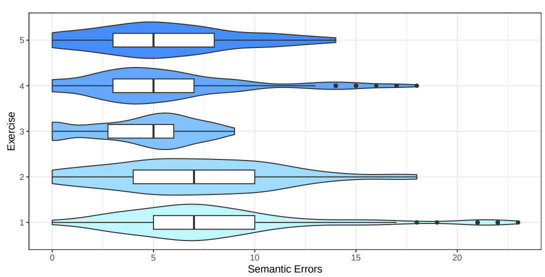
Figures 3(a), 3(b), and 3(c) show, respectively, the distribution of exercise correctness, syntax, and semantic errors depending on the exercise. We can see that exercise completeness is around 75%, with Exercise 2 being the one with the lowest average score, indicating a lower percentage for said exercise; syntax errors, on average, increase a bit around Exercise 4, while semantic errors see an average decrease towards the middle of the course, to then slightly increase again.



(a) Box and violin plot for the distribution of the Completeness metric



(b) Box and violin plot for the distribution of the Syntax Errors metric



(c) Box and violin plot for the distribution of the Semantic Errors metric

Figure 3: Distribution of the Learning Progress metrics

Table 7: Coefficients and p-values for linear model using fixed reference encoding for each progress metric. Statistically significant p-values are in bold.

Exercise	Completeness	Syntax Errors	Semantic Errors
2	-4.37 (**4.86e-3**)	-0.37 (1.01e-1)	-0.93 (**4.83e-3**)
3	1.08 (4.87e-1)	-0.51 (2.38e-2)	-3.6 (< **2e-16**)
4	1.14 (4.63e-1)	0.36 (1.15e-1)	-2.26 (**1.03e-11**)
5	2.41 (1.20e-1)	0.16 (4.92e-1)	-2.40 (**6.19e-13**)

Tables 7 and 8 present the difference between the average values and the p-values resulting, respectively, from the linear model that compares each exercise with the first and with the preceding one.

Regarding exercise correctness, we observe that, for the comparison with the first exercise, the only statistically significant p-value is the one obtained when comparing it with the second exercise: moreover, the difference is actually negative, meaning that there is a drop in performance when performing the second exercise.

The comparison with subsequent exercises, however, shows a statistically significant increase in completeness when going from Exercise 2 to Exercise 3, meaning that there is an improvement as students keep using the gamified tool.

Table 8: Coefficients and p-values for linear model using progressive encoding for each progress metric. Statistically significant p-values are in bold.

Exercise Pair	Completeness	Syntax Errors	Semantic Errors
Ex. 1 -> Ex. 2	-4.37 (**4.86e-3**)	-0.37 (1.01e-1)	-0.93 (**4.83e-3**)
Ex. 2 -> Ex. 3	5.55 (**4.58e-4**)	-0.14 (5.34e-1)	-2.69 (**8.33e-16**)
Ex. 3 -> Ex. 4	0.06 (9.69e-1)	0.87 (**1.33e-4**)	1.36 (**3.85e-05**)
Ex. 4 -> Ex. 5	1.27 (4.12e-1)	-0.20 (3.75e-1)	-0.14 (6.80e-1)

The absence of other statistically significant results means that we cannot reject the null hypothesis $H_{2.1_0}$: gamification does not lead to a significant improvement of exercise correctness over time.

Regarding syntax errors, the comparison with the first exercise shows only a slightly significant p-value when comparing it with Exercise 3: the negative difference between the mean values implies that students tend to have a good understanding of BPMN theory concepts and modeling rules at around halfway through the course.

When comparing each exercise with the preceding one, however, we see that there is a statistically significant increase in syntax errors when going from Exercise 3 to Exercise 4: the latter introduces concepts such as messages, events, and multiple pools as necessary to complete the exercise, so it makes sense that students may struggle with new concepts.

The resulting p-values are not enough to safely reject the hypothesis $H_{2.2_0}$: gamification does not lead to a statistically significant reduction of the syntax errors made by students over time.

Regarding semantic errors, the comparison with the first exercise yields statistically significant p-values for all four exercises: moreover, the differences in mean values are all negative, meaning that students perform fewer syntax errors compared to when they begin using BIPMIN as they keep performing exercises.

The comparison of an exercise with the preceding one yields statistically significant p-values for all pairs except between Exercises 4 and 5: we observe negative differences for the first two pairs, meaning fewer errors, and a positive difference for the latter pair, implying a higher number of semantic errors. As mentioned above, Exercise 4 requires detailed knowledge of many new topics that were optional in the previous exercises, meaning that students may risk making more errors.

The resulting p-values show that students perform fewer semantic errors over time and compared to the beginning of the course, meaning that we can reject $H_{2.3_0}$ and assert that gamification can impact the number of semantic errors made by students by reducing them over time.

5.3 RQ3: Perception of the gamified experience

The perception of student was assessed by means of two questionnaires. Out of the 200 participants, five of them did not answer the questionnaire, leading to a total of 195 answers analyzed.

Regarding the TAM and GAMEX questions, we computed the mean of the answers given by each student to each construct as the general opinion of the student on the construct and then computed the distribution of these mean answers to get the general picture

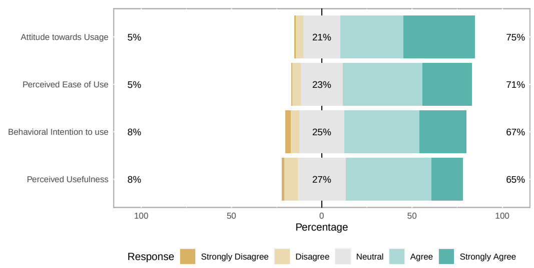


Figure 4: Distribution of answers to the TAM questionnaire

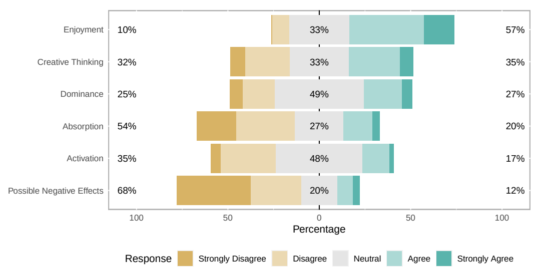


Figure 5: Distribution of answers to the GAMEX questionnaire

of the participants' opinions. Figure 4 shows the distribution of answers given to the TAM questionnaire.

We observe an encouragingly positive distribution of answers given to the TAM questionnaire: for all four fields, there are at least 65% of the students who either *Agree* or *Strongly Agree* as their general opinion regarding each field. The high percentage of positive answers for the Attitude towards Usage construct and the low percentage of negative answers for Perceived Ease of Use and Perceived Usefulness are particularly encouraging results, as they mean BIPMIN is perceived by the majority of our students as a fairly easy to use tool that can lead to benefits for its users.

We present the distribution of answers to the GAMEX questions in Figure 5: the most appreciated construct is Enjoyment, with at least half of the participants having a positive opinion and a total absence of *Strongly Disagree* answers, implying that using BIPMIN is generally perceived as a fun and enjoyable activity; the high percentage of negative answers for Possible Negative Effects further reinforces this result, with around two-thirds of the participants agreeing that the tool is neither frustrating nor upsetting.

The answers to the four remaining constructs do not show similarly polarized distributions: regarding Creative Thinking, the three general opinions (negative, neutral, positive) all have around one-third of each of the participants, showing that BIPMIN does not offer a completely free environment where its users are free to be creative; around half of the participants have a neutral opinion for the Dominance construct, showing that they did not feel completely in charge of their actions but somewhat limited by the tool. Lastly, the Absorption and Activation constructs present distributions that are more on the negative side, implying that the gamified experience may not have been perceived as immersive or stimulating.

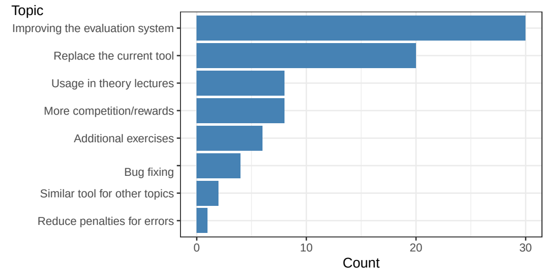


Figure 6: Topics found in the open-ended question after performing open coding on the answers

The answers to the GAMEX questionnaire show that BIPMIN has been perceived, on average, as an enjoyable tool that does not make the modeling activity frustrating; however, the tool is still lacking in some aspects (e.g. creativity, freedom of action, immersivity) that make the gamified experience incomplete.

We present in Figure 6 the results of the open coding we performed on the answers given to the open-ended question asking for the students' opinions. Out of the 200 total answers, we identified 74 answers that did not contain any opinion, issue, or suggestion, and 42 answers where students declared themselves satisfied with the way BIPMIN was integrated in the course; we have excluded these answers from our analysis, leading to a total of 79 answers.

The most common topic found (30 answers) highlights the current limits of the evaluation system: the comparison of a student's diagram with reference solutions means that teachers need to consider all possible alternative modeling solutions for a given exercise, in terms of both diagram elements and possible element names; many students found themselves in a situation where a diagram that was assigned a low score was judged much higher after minor changes such as using a synonym for an element's name or changing some elements in the diagram.

Other relevant topics emerged as a consequence of using BIPMIN exclusively in the practical laboratories: theory lectures and the final project work were both conducted using Signavio Academic. This disparity in usage, as well as the absence of gamified tools for the other course topics, led many students to ask for a more thorough integration of gamification, spanning over the entire course instead of just the laboratories.

6 Discussion

For this study, we performed a longitudinal experiment to assess whether the application of a gamified modeling tool throughout the duration of an Information Systems course can be beneficial in terms of effectiveness on student grades and learning progress over time. Moreover we evaluated the perception of the tool by students.

We evaluated the **effectiveness** of the gamification by comparing the grades obtained by the students of the 2024 edition of the course with those obtained during the 2023 edition, where traditional learning methods were used. Two separate grade types have been compared: the points obtained in the written exam for BPMN-related exercises and the points obtained for the BPMN part of the end-of-course project work.

By applying statistical tests, we found that gamification had a significant positive impact on both exam (increased by 0.73, +12.52%) and project (increased by 2.4, +117.03%) grades.

We evaluated the students' **learning progress** by analyzing three metrics: the exercise completeness, computed as the number of correct elements present in the student diagrams, the syntax errors, and semantic ones made by the students. The longitudinal experiment consisted of five distinct exercises attempted by the students throughout the course, and the aforementioned metrics have been considered for each exercise. The statistical analysis for student progress considered two different scenarios: one where the metrics of each exercise were compared with those of the first exercise, to assess the students' status in comparison with the beginning of the course, and another where each exercise was compared with the preceding one, to evaluate the presence of improvements.

Regarding exercise *completeness*, we found that the only significant difference in comparison with the beginning of the experiment was a drop in performance when performing the second exercise; moreover, the comparison among pairs of exercises showed a significant increase when going from the second to the third exercise. These results are too inconclusive to state that gamification can lead to an increase of exercise correctness over time.

Concerning *syntax errors*, we found a statistically significant decrease, on average, at around halfway through the experiment, and a statistically significant increase with the introduction of mandatory advanced topics; these results show that gamification does not lead to a significant change over time when it comes to BPMN syntax errors.

As far as *semantic errors*, the comparison with the first exercise yielded statistically significant p-values for all four exercises, with a decrease in the average number of errors for all comparisons. The comparison between subsequent exercises also showed a significant reduction in errors around the midpoint of the experiment, with a significant increase tied to the introduction of advanced topics.

Based on the above findings, we conjecture that the continuous use of a gamified tool during the course made the students more accustomed to the type of solutions that are expected for the provided exercises, thanks to the semantic errors promptly shown in the tool UI. We deem this beneficial since it can render the students used to the way the final project of the course – or even the exam if it contained practical modeling exercises – is eventually evaluated.

Overall, we can see that, if an exercise has a difficulty level higher than expected, in relation to its placement in time, can translate to worse student performance; the difficulty does not discourage students, however, who improve their understanding of modeling at around halfway through the course. The introduction of advanced topics as necessary to complete an exercise also causes drops in performance, but students can still make good use of the knowledge they gained over time and improve for the later activities.

The **students' perception** of the gamified experience was measured through the TAM and GAMEX questionnaire, focused on the usability and the game aspects, respectively. The usability of BIPMIN was appreciated by more than half of the participants, being perceived as easy to use and useful for its modeling purpose. The answers to the GAMEX questionnaire, however, yielded less positive results: while the tool was judged as enjoyable and not frustrating there were still issues such as it not being immersive or not allowing enough creative thinking and freedom of action.

Lastly, the analysis of the students' opinions regarding issues and improvement areas found a glaring issue in the current implementation of the evaluation engine, as many students expressed their dissatisfaction with semantically equivalent solutions not being accepted by the tool; many students also felt that adopting a gamified approach for the other course topics, and for the theory aspects of the course as well, would have been more effective.

7 Conclusion and Future Work

This paper described a longitudinal experiment where gamification was applied to the BPMN modeling assignments in a Master Degree Information Systems course.

The results showed that, with the application of gamification, students managed to improve their knowledge of BPMN modeling, with this improvement translating to higher grades in comparison to a previous edition of the course, where no gamification was employed. The application of gamification led the students to comply better with the requirements of the provided assignments and eventually with those of the final course project, which in an educational environment translates to better student satisfaction. Our results are in line with existing evidence in literature about the increase of grade and students' motivation [16], and – to the best of our knowledge – represent the first effort in providing a large-scale, empirical experimentation of such impacts. The results can be as well transferred to other modeling languages and techniques, as we verified with experiences with the UML modeling language that employed a similar framework of gamification mechanics [15].

The tool was perceived as easy to use, effective, and enjoyable, with some issues related to its evaluation engine, considered by students not flexible enough. In no case the use of the tool was considered as detrimental to learning the core concepts of the course.

As our future research plans, we plan to extend the evaluation engine of BIPMIN, which showed some limitations in how it handled alternative correct representations of the same process fragment, which were not foreseen in advance by the lecturers. A possible strategy would be the usage of Large Language Model agents, which have shown promising capabilities in understanding the semantic variations of a scenario [2]. Finally, it is worth underlying that teaching adjustments were applied between the first and second years considered in this study. Even though these threats have been mitigated by normalizing the scores given in the considered years, future research efforts may replicate the evaluation by comparing the effects on equivalent course setups.

Acknowledgments

This study was carried out within the "EndGame - Improving End-to-End Testing of Web and Mobile Apps through Gamification" project (2022PCCMLF) – funded by European Union – Next Generation EUithin the PRIN 2022 program (D.D.104 - 02/02/2022 Ministero dell'Università e della Ricerca). This manuscript reflects only the authors' views and opinions and the Ministry cannot be considered responsible for them.

References

[1] Manal M Alhammad and Ana M Moreno. 2018. Gamification in software engineering education: A systematic mapping. *Journal of Systems and Software* 141 (2018), 131–150.

[2] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. 2021. Program Synthesis with Large Language Models. _eprint: 2108.07732.

[3] Kylie Bedwell, Giacomo Garaccione, Riccardo Coppola, Luca Ardito, and Maurizio Morisio. 2023. BIPMIN: A Gamified Framework for Process Modeling Education. *Information* 14, 1 (2023). https://doi.org/10.3390/info14010003

[4] Antonio Bucchiarone, Maxime Savary-Leblanc, Xavier Le Pallec, Antonio Cicchetti, Sébastien Gérard, Simone Bassanelli, Federica Gini, and Annapaola Marconi. 2023. Gamifying model-based engineering: the PapyGame experience. *Software and Systems Modeling* (2023), 1–21. Publisher: Springer.

[5] Christian Cagnazzo, Giacomo Garaccione, Riccardo Coppola, Luca Ardito, and Marco Torchiano. 2023. UMLegend: A Gamified Learning Tool for Conceptual Modeling with UML Class Diagrams. In *Proceedings of the 2nd International Workshop on Gamification in Software Development, Verification, and Validation (Gamify 2023)*. Association for Computing Machinery, New York, NY, USA, 2–5. https://doi.org/10.1145/3617553.3617883

[6] Ilaria Caponetto, Jeffrey Earp, and Michela Ott. 2014. Gamification and education: A literature review. In *European conference on games based learning*, Vol. 1. Academic Conferences International Limited, 50.

[7] Y. k Chou. 2015. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Createspace Independent Publishing Platform. https://books.google.it/books?id=jFWQrgEACAAJ

[8] Valerio Cosentino, Sébastien Gérard, and Jordi Cabot. 2017. A Model-based Approach to Gamify the Learning of Modeling.. In *SCME-iStarT@ ER*. 15–24.

[9] Tommaso Dal Sasso, Andrea Mocci, Michele Lanza, and Ebrisa Mastrodicasa. 2017. How to gamify software engineering. In *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*. IEEE, 261–271.

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

[11] Damien Djaouti, Julian Alvarez, and Jean-Pierre Jessel. 2011. Classifying Serious Games: the G/P/S model. *Handbook of Research on Improving Learning and Motivation through Educational Games: Multidisciplinary Approaches* (Jan. 2011). https://doi.org/10.4018/978-1-60960-495-0.ch006

[12] Daniel J. Dubois and Giordano Tamburrelli. 2013. Understanding Gamification Mechanisms for Software Development. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering* (Saint Petersburg, Russia) *(ESEC/FSE 2013)*. Association for Computing Machinery, New York, NY, USA, 659–662. https://doi.org/10.1145/2491411.2494589

[13] René Eppmann, Magdalena Bekk, and Kristina Klein. 2018. Gameful experience in gamification: Construction and validation of a gameful experience scale [GAMEX]. *Journal of interactive marketing* 43, 1 (2018), 98–115. Publisher: SAGE Publications Sage CA: Los Angeles, CA.

[14] Tommaso Fulcini, Riccardo Coppola, Luca Ardito, and Marco Torchiano. 2023. A review on tools, mechanics, benefits, and challenges of gamified software testing. *Comput. Surveys* 55, 14s (2023), 1–37.

[15] Giacomo Garaccione, Riccardo Coppola, Luca Ardito, and Marco Torchiano. 2024. Gamification of business process modeling education: an experimental analysis. *Software and Systems Modeling* (2024), 1–26.

[16] Mantas Jurgelaitis, Lina Čeponienė, Jonas Čeponis, and Vaidotas Drungilas. 2019. Implementing gamification in a university-level UML modeling course: A case study. *Computer Applications in Engineering Education* 27, 2 (2019), 332–343.

[17] Ed Júnior and Kleinner Farias. 2021. ModelGame: A Quality Model for Gamified Software Modeling Learning. In *Proceedings of the 15th Brazilian Symposium on Software Components, Architectures, and Reuse (SBCARS '21)*. Association for Computing Machinery, New York, NY, USA, 100–109. https://doi.org/10.1145/3483899.3483910 event-place: Joinville, Brazil.

[18] Bahar Kutun and Werner Schmidt. 2019. BPMN wheel: Board game for business process modelling. In *European conference on games based learning*. Academic Conferences International Limited, 1008–1012.

[19] Younghwa Lee, Kenneth A Kozar, and Kai RT Larsen. 2003. The technology acceptance model: Past, present, and future. *Communications of the Association for information systems* 12, 1 (2003), 50.

[20] Jenni Majuri, Jonna Koivisto, and Juho Hamari. 2018. Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd international GamiFIN conference, GamiFIN 2018*. CEUR-WS.

[21] Javier Mancebo, Felix Garcia, Oscar Pedreira, and Maria Angeles Moraga. 2017. BPMS-Game: tool for business process gamification. In *Business Process Management Forum: BPM Forum 2017, Barcelona, Spain, September 10-15, 2017, Proceedings 15*. Springer, 127–140.

[22] Beatriz Marín. 2022. Lessons Learned About Gamification in Software Engineering Education. 1473–1496. https://doi.org/10.4018/978-1-6684-3710-0.ch071

[23] Ceres Germanna Braga Morais, Daniela Pedrosa, Mario Madureira Fontes, José Cravino, and Leonel Morgado. 2020. Detailing an e-Learning course on software engineering and architecture using BPMN. In *First International Computer Programming Education Conference (ICPEC 2020)*, Vol. 81. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 17–1.

[24] OMG. 2011. *Business Process Model and Notation (BPMN), Version 2.0*. Needham, MA, USA.

[25] Klaas-Jan Stol, Paul Ralph, and Brian Fitzgerald. 2016. Grounded Theory in Software Engineering Research: A Critical Review and Guidelines. In *2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE)*. 120–131. https://doi.org/10.1145/2884781.2884833

[26] Mai Thai, Li Chong, and Narendra Agrawal. 2011. Straussian Grounded Theory Method: An Illustration. *Qualitative Report* 17 (Sept. 2011). https://doi.org/10.46743/2160-3715/2012.1758

[27] Vladimir Uskov and Bhuvana Sekar. 2014. Gamification of software engineering curriculum. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. 1–8. https://doi.org/10.1109/FIE.2014.7044098

[28] Claes Wohlin, Per Runeson, Martin Höst, Magnus C Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in software engineering*. Springer Science & Business Media.