Gamified Requirements Engineering: Model and Experimentation

Philipp Lombriser, Fabiano Dalpiaz, Garm Lucassen, and Sjaak Brinkkemper

Utrecht University - Department of Information and Computing Sciences,
Princetonplein 5, De Uithof, 3584 CC Utrecht, The Netherlands
{p.lombriser,f.dalpiaz,g.lucassen,s.brinkkemper}@uu.nl

Abstract. [**Context & Motivation**] Engaging stakeholders in requirements engineering (RE) influences the quality of the requirements and ultimately of the system to-be. Unfortunately, stakeholder engagement is often insufficient, leading to too few, low-quality requirements. [**Question/problem**] We aim to evaluate the effectiveness of gamification to improve stakeholder engagement and ultimately performance in RE. We focus on agile requirements that are expressed as user stories and acceptance tests. [**Principal ideas/results**] We develop the gamified requirements engineering model (GREM) that relates gamification, stakeholder engagement, and RE performance. To evaluate GREM, we build an online gamified platform for requirements elicitation, and we report on a rigorous controlled experiment where two independent teams elicited requirements for the same system with and without gamification. The findings show that the performance of the treatment group is significantly higher, and their requirements are more numerous, have higher quality, and are more creative. [**Contribution**] The GREM model paves the way for further work in gamified RE. Our evaluation provides promising initial empirical insights, and leads us to the hypothesis that competitive game elements are advantageous for RE elicitation, while social game elements are favorable for RE phases where cooperation is demanded.

Keywords: Gamification, Requirements Elicitation, Empirical Study, Agile Requirements, Gamified Requirements Engineering Model

1 Introduction

Despite the crucial role of requirements engineering (RE) in software development [30], many IT projects still fail to deliver on time, within cost, or expected scope [5]. Reasons for project failures include incorrect or unsatisfied requirements, often caused by poor collaboration and communication. Furthermore, the lack of stakeholder participation in RE workshops and review meetings are additional impediments to the completion of software projects [18, 3].

In this paper, we aim to improve the quality and increase the creativity of requirements by enhancing active participation of stakeholders in requirements elicitation workshops, especially when online digital platforms are used. We suggest gamification as a possible way to achieve this end.

The literature on gamification and RE is limited to two main studies [12, 31] that develop software tools to increase stakeholder engagement and evaluate it via a single case study. The former study [12] proposes the *iThink* tool that is designed to stimulate parallel thinking and increase group discussion. The latter study [31] introduces the *REfine* platform that aims at enlarging participation in RE by involving a crowd of both internal and external stakeholders [32]. In both works, the conducted case study showed that stakeholders felt more motivated and that participation rate increased in the requirements elicitation process.

Despite their novelty, these works have limitations. The researchers only evaluated their tool in the context of a case study, making it difficult to generalize the results and draw conclusions about causality. The impact of alternative causes, such as usability, design, and stakeholders' background were omitted.

We address these limitations by evaluating the gamification of RE in a controlled experimental setting that enables better determining patterns of cause and effect. Gamification is applied in the context of agile RE to the elicitation of user stories enriched with acceptance tests that are expressed as real-life examples. We make the following contributions:

- We propose a Gamified Requirements Engineering Model (GREM) to evaluate the impact of gamification on engagement and performance in requirements elicitation.
- We develop a gamified online platform for requirements elicitation that supports expressing requirements as user stories and acceptance tests.
- We evaluate the effectiveness of the platform through a controlled experiment with two equal balanced groups of stakeholders, and we conduct quantitative analyses on the results.
- Based on the outcomes of the evaluation, we propose a mapping between the different game elements and the RE phases they support best.

The rest of the paper is structured as follows. Sec. 2 reviews related work on agile RE and gamification. Sec. 3 presents the conceptual framework for our research. Sec. 4 describes our proposed gamified platform. Sec. 5 reports on the design and administration of the experiment, and Sec. 6 discusses the results. We analyze threats to validity in Sec. 7, and we conclude in Sec. 8.

2 Background

After reviewing scenario-based RE in the context of agile software development in Sec. 2.1, we introduce the principles behind gamification and its potential impact on motivation and engagement in Sec. 2.2.

2.1 Scenario-Based RE in Agile Development

In RE, a scenario is "an ordered set of interactions between partners, usually between a system and a set of actors external to the system" [14]. Scenarios can take many forms and provide various types of information on different levels of

abstraction. The specification spectrum can vary between informal descriptions to more formal representation. They can be expressed in natural language, diagrams, pictures, wireframes, mockups, storyboards, prototypes, customer journeys, and many other formats [34]. The selection of the appropriate scenario technique depends on many factors including acceptance, notation skills, specification level, type of system, complexity, consistency, and unambiguity [30].

User Stories. After evaluating different techniques, we decided to select user stories as a requirements documentation technique because of their simplicity, comprehensibility, and their popularity in agile development [23]. They are easy to learn and can be also applied by stakeholders without any notation or modeling skills. Furthermore, user stories stimulate collaboration and facilitate planning, estimation, and prioritization. Cohn [6] suggests to use the following tripartite structure when documenting user stories:

As a [**role**], I want to [**goal**], so that [**benefit**]

The *role* defines who will directly benefit from the feature, the *goal* specifies which feature the system should exhibit, and the *benefit* is the value that will be obtained by implementing the user story. An example of user story is the following: "As an Administrator, I want to be notified of incorrect login attempts, so that I can more easily detect attackers".

Personas are often used to facilitate the discovery of user stories: a persona is a fictional character that represents roles and characteristics of end users [6]. Stakeholders can be assigned specific personas to obtain requirements from the perspective of specific user types.

Acceptance tests. Acceptance criteria complement user stories with conditions that determine when a story is fulfilled [6]. They specify how the system should behave to meet user expectations. We choose to use Dan North's template [25] for expressing acceptance tests:

Given [**context**], when [**event**], then [**outcome**]

In summary, our baseline for representing requirements consists of: (i) *personas* to distinguish between different types of users, (ii) *user stories* to explain what the users want to achieve through the system, and (iii) *acceptance tests* to determine the correctness criteria for the system to satisfy a user story.

Quality of User Stories. INVEST is an acronym that characterizes six core attributes to evaluate the quality of a user story [35]. According to INVEST, good user stories should be *Independent* from each other, *Negotiable* as opposed to a specific contract, *Valuable* for the stakeholder, *Estimable* to a good approximation, *Small* so as to fit within an iteration, and *Testable*.

The extrinsic value of a user story, however, can be better made explicitly visible using the Kano model [17], which can be utilized to determine how satisfied

or dissatisfied end users will be with the presence or absence of certain system features. Although initially developed for marketing, the Kano model can be effectively utilized in agile methodologies for prioritizing product backlog. The priority is determined by answering *functional* (what if a feature is included?) and *dysfunctional* (what if a feature is excluded?) questions [7]. The model characterizes features according to the customer value their implementation leads to:

- *Must-be*: implementation is taken for granted but exclusion from implementation leads to dissatisfaction;
- *One-dimensional*: satisfaction if implemented and dissatisfaction if excluded;
- *Attractive*: satisfaction if implemented but no dissatisfaction if excluded;
- *Indifferent*: neither satisfaction or dissatisfaction;
- *Reverse*: implementation leads to dissatisfaction.

2.2 Gamification

The principles behind gamification have existed for decades, but the term itself became mainstream only in 2010 with its initial definition of "the application of game design elements in non-gaming contexts" [11]. A principal reason why gamification has become so popular in recent years is that games have a strong "pull" factor [20]. Games affect positive emotions, relieve stress, create stronger social relationships, give a sense of accomplishment, and improve cognitive skills [15]. With gamification, the advantages of games are applied to existing business contexts in order to increase success metrics [37].

Game Elements. The classic triad of game elements in gamification consists of *points*, *badges*, and *leaderboards* (PBL) [37]. Many platforms use these elements because of their effectiveness and implementability. Points are tangible and measurable evidence of accomplishment; badges are a visual representation of achievements; and leaderboards allow players to compare themselves against each other. Next to PBL, a variety of game elements exist, including *levels*, *storytelling*, *chance*, *goals*, *feedback*, *rewards*, *progress*, *challenge*, *avatar*, and *status*. They allow for a compelling user experience and leverage motivation [37].

To understand the effects of gamification on player's behavior, a closer look at the theories of motivation and engagement is due.

Motivation. People have needs that motivate them to take action to satisfy their desires. The Maslow pyramid is one of the earliest theories describing people's needs [24]. Based on various research studies, Reiss identified 16 basic desires that guide human behavior [28]. The Self-Determination Theory (SDT) is concerned with people's inherent tendencies to be self-determined and self-motivated, without external interference [10]. SDT distinguishes between *intrinsic* and *extrinsic* motivation. People are intrinsically motivated when they

do something because they simply enjoy the activity, whereas extrinsically motivated people do something for external rewards or to avoid negative consequences. "Flow" is also considered to be a motivating force for excellence. Individuals experiencing flow are more motivated to carry out further activities [8]. Optimal flow can be obtained with progression stairs, whereas engagement loops are responsible to keep players motivated by providing constant feedback [37].

Engagement. User engagement in information science covers the study of people's experience with technology [26]. The term is an abstract concept and closely related to theories of *flow*, *aesthetic* and *play*. User engagement is defined as "the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource" [2]. Therefore, engaged people not only better accomplish their personal goals, but are also physically, cognitively, and emotionally closer connected to their endeavors [26].

3 The Gamified Requirements Engineering Model

We devise a conceptual model that aims to explain the effect of gamification on stakeholder engagement and RE performance. The gamified requirements engineering model (GREM) integrates the theories of *gamification* and *engagement* presented in Sec. 2.2 in the context of *performance* in RE. The relationships between these three concepts are shown in the conceptual model of Fig. 1. The model consists of three main abstract variables: the independent variable *gamification* and the dependent variables *stakeholder engagement* and *performance*. Furthermore, two control variables mitigate threats to internal validity: *motivation* and *stakeholder expertise*. For stakeholder engagement three sub-dimensions are defined: emotions, cognition and behavior [2]. Performance is sub-divided into productivity, quality and creativity, which are perceived as supportive concepts for measuring the output in requirements elicitation [19].

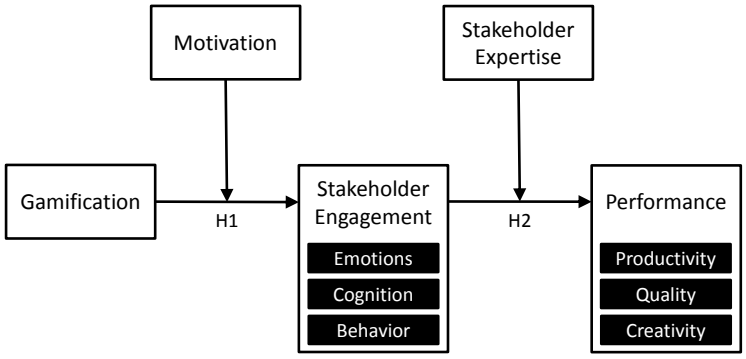

Fig. 1. The Gamified Requirements Engineering Model (GREM)

We operationalize each of the concepts of the model as follows:

- Gamification is measured with a dichotomous variable by dividing the sample into two equal balanced groups.
- Motivation is measured with the Reiss profile test [28], a rich and extensively tested tool to assess human strivings [16].
- Emotions are measured with the Positive and Negative Affect Schedule (PANAS) [36]. Since gamification is expected to provoke positive emotions, we only consider Positive Affect (PA), thereby excluding negative affective states such as distressed, upset, guilty, hostile and scared.
- Cognition is reported through the the Flow Short Scale (FSS), which consists of 10 items to measure the components of flow experiences using a 7-point scale [29].
- Behavior is observed through background analytics provided by the platform that is used to express requirements.
- Stakeholder expertise is measured with a pretest questionnaire on experience in IT, RE and user stories.
- Productivity is calculated with the number of user requirements produced.
- Requirements quality is assessed with INVEST and the Kano model.
- Creativity of user stories is determined with expert opinions on a 5-point Likert scale (1 = definitely not novel, 5 = definitely novel).

Based on this conceptual model, the following two hypotheses are defined:

H1 If a diversified gamification RE platform is deployed in alignment with motivation, then stakeholder engagement is significantly increased.

H2 If stakeholders are more engaged in requirements elicitation with respect to their expertise, then the overall performance of the process and outcomes is significantly increased.

4 A Gamified Requirements Elicitation Platform

To test the effect of gamification on engagement and on performance in RE, we designed and developed an online gamified platform for eliciting requirements through user stories and acceptance tests. Our platform is developed on top of Wordpress[1]. User stories are specified by adapting blog entries to the user story template, while acceptance tests are expressed as comments to blog entries. Furthermore, a chat is included to facilitate stakeholder collaboration.

We developed the platform in such a way that the gamification elements could be enabled or disabled easily, making it possible to design specific experiments between a control group (no gamification) and a treatment group (with gamification). We embedded support for a number of gamification elements using the *Captain Up* API[2], which enables turning a website into a game experience. Basic

[1] http://www.wordpress.com
[2] https://captainup.com/

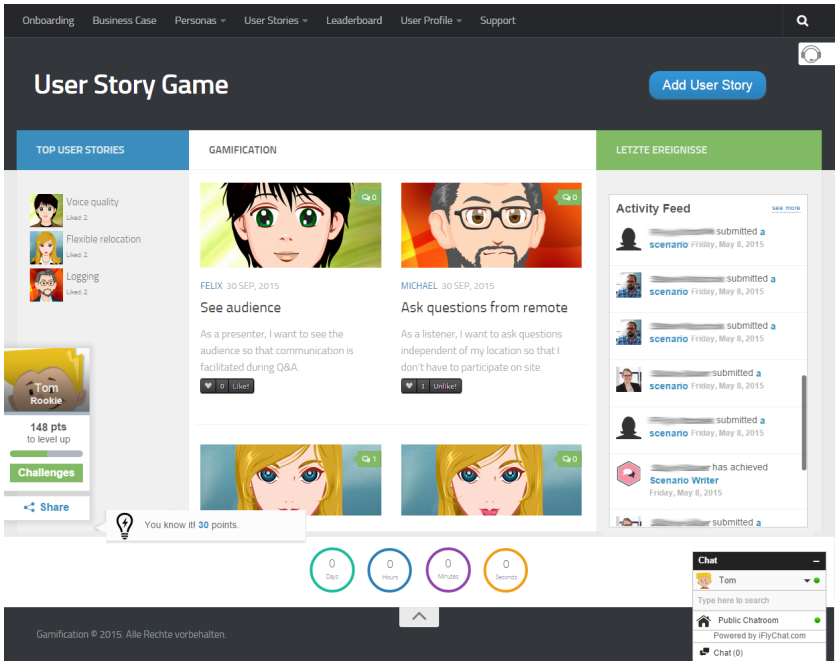

Fig. 2. A screenshot of our requirements elicitation platform showing two user stories

game elements that come out of the box include points, badges, leaderboards, levels, challenges and activity feeds.

Using this plugin, we assign points and badges based on the actions that the user is performing on the website, such as writing user stories or acceptance tests, visiting specific pages, and sending chat messages. The number of points that are awarded is calculated based on the estimated achievement time for the different tasks that lead to points and badges.

For instance, submitting a user story is rewarded with *30 points* and adding an acceptance test with *10 points*, based on our estimation that writing a good-quality user story would take about three minutes, while creating a single acceptance test for a specified story would take approximately one minute. After writing 3 user stories a '*User Story Writer*' badge plus *90 bonus points* are credited to the user's account. Based on collected points, players can level up and compare their rank on a highscore list. The primary goal of the gamification API is to allow players to pursue mastery with a progression stair and keep them actively engaged with a positive reinforcement cycle [37]. To give points a specific, tangible meaning, a prize is awarded to the winner of the game. The player with the most points and likes receives a gift card with a value of €25. A screenshot of the platform's front-end is shown in Fig. 2.

Table 1. Summary of game elements and mechanics that we implemented

Game Element	**Affected Motivation [28, 27]**
Points: the basis means to reward users for their activities	Order, Status, Saving
Badges: visualizations of achievements to give a surprise effect	Power, Order, Saving
Leaderboard: a ranking of the players	Power, Order, Status
Levels: phases of difficulty in a game to enable progression	Order, Independence Status
Challenges: steps towards a goal, which are rewarded with badges and points	Curiosity, Independence, Power
Activity feed: a stream of recent actions of the community	Power, Order, Status
Avatar: graphical representation of the current player	Power, Independence, Status
Onboarding: the process of getting familiar with the platform	Curiosity, Independence, Tranquility
Game master: the moderator of the game	Curiosity, Social Contact, Status
Storytelling: a background narrative to arouse positive emotions	Curiosity, Independence, Tranquility
Video: media to explain user stories and the business case	Curiosity, Order, Tranquility
Facial animation: animated characters to introduce personas	Curiosity, Order, Tranquility
Progress bar: a bar showing the player's current state in a process	Order, Tranquility
Quiz: a test to let players check their new acquired knowledge	Curiosity, Independence, Order
Timer: a clock that shows remaining time and that puts pressure	Order, Tranquility
Liking: a feature for users to support certain content	Power, Status, Vengeance
Prize: physical award given to the winner of the game	Power, Independence, Status

In addition, we include further game elements that we implemented to enhance user experience and stimulate intrinsic motivation. We created a video introduction of the business case that makes use of *video animation*; the case is explained by a fictional character called Tom. Moreover, we devised a *storyline* that guides the player into the platform's basic actions, such as learning about the business case, creating a user story, specifying an acceptance test, etc.

Facial animation is used to make our personas more vivid. Talking characters are responsible to make their background stories more memorable. The primary goal of facial animation is to develop stakeholder trust in the personas by increasing empathy and provoking a fun and novel experience.

A complete list of game elements and mechanics is captured in Table 1. The purpose of this broad selection is to affect a variety of human needs. For example,

while leaderboards satisfy people with desire for status and power, storytelling is more suitable for people with a demand for curiosity [27].

5 Experiment

We investigate the effect of gamification on stakeholder engagement and performance in a controlled experimental setting based on the GREM model introduced in Sec. 3. The intervention on the treatment group consists of the 17 game elements that were included in our online platform presented in Sec. 4. These game elements were disabled for the platform that was used by the control group. Our aim is to measure the response of the gamification intervention by means of an ex-post test. All details on the experiment can be found online [21].

The experiment was conducted at MaibornWolff[3], an IT consultancy company in Munich (Germany) that employs over 160 people and was founded in 1989. Our experiment involved 12 potential stakeholders. Participants were divided into two equal balanced groups with consideration to gender, motivation and expertise. The grouping used their Reiss profile test results and an experience pre-test on IT, RE, and user stories.

Before the experiment, all participants were simultaneously briefed and provided with a real business case. The company is currently lacking an efficient *video conferencing system* (VCS) for corporate team meetings. Stakeholders were asked to gather user requirements that could serve as a checklist to compare different existing VCS solutions. Both groups were given a time range of two hours to fill an initial VCS backlog with user stories together as a team.

To avoid interferences between the experimental groups, participants were told that they are working on two different cases. Furthermore, the impression was given that the aim of the experiment was to test remote requirements engineering and that communication is only allowed within the team via the integrated chat feature. The investigation of gamification was never mentioned to the participating subjects (neither in the control nor in the treatment group).

5.1 Results

The operation of the experiment went smoothly with an issue facing the treatment group. One participant from the control group dropped out after 10 minutes, leaving the group with 5 stakeholders. The data from this participant is omitted from the analysis.

The following sections present the aggregated findings from the experiment, which were statistically analyzed in SPSS. Quality was rated by 5 Scrum experts, while creativity was assessed by 13 potential end users. While reading the results, bear in mind the limited size of our experiment, which threatens the generality of the results (see also Sec. 7).

[3] http://www.maibornwolff.de/en

Performance We report on the results about the performance dependent variable in Fig. 1, which are measured in terms of productivity, quality and creativity.

Productivity. The average number of provided user stories within the treatment group was much higher than those of the control group. A significant difference was also identified in the total number of submitted acceptance tests between the treatment group and the control group. The total number of produced user stories and acceptance tests per group can be found in Fig. 3, whereas Table 2 reports the statistical results.

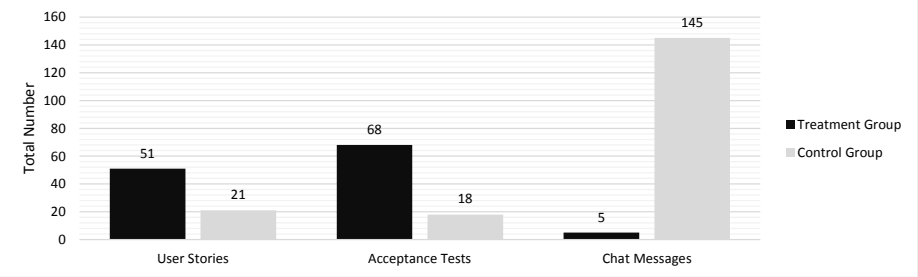

Fig. 3. Total number of produced user stories, acceptance tests and chat messages

Table 2. Independent t-test results for performance: : M = mean, SD = standard deviation, t = t-value, p = p-value

	Treatment Group		Control Group			
	M	SD	M	SD	t	p
User Stories (Productivity)	10.000	2.345	3.500	2.258	4.673	.001
Acceptance Tests (Productivity)	13.400	5.727	3.000	3.847	3.597	.006
Independent	4.022	.950	3.436	1.302	3.025	.003
Negotiable	3.985	1.099	3.891	1.048	.543	.558
Valuable	3.933	1.052	4.055	1.061	-.718	.473
Estimable	3.504	1.177	2.418	1.213	5.714	<.001
Small	3.244	1.187	2.364	1.007	4.837	<.001
Testable	4.193	1.040	3.418	1.370	3,772	<.001
Creativity	3.044	1.0850	2.236	.922	4.853	<.001

Quality. For the quality aspect, the requirements were stratified sampled and evaluated by 5 certified Scrum experts (between 1 and 9 years of experience) with the INVEST model [35].

User stories gathered by the treatment group were more independent (I), allowed for better estimations (E), were smaller (S), and better testable (T)

than those of the control group. Negotiable (N) and valuable (V) did not report any significant differences between the two groups. The mean score for each characteristic is presented in Table 2 and visualized in Fig. 4.

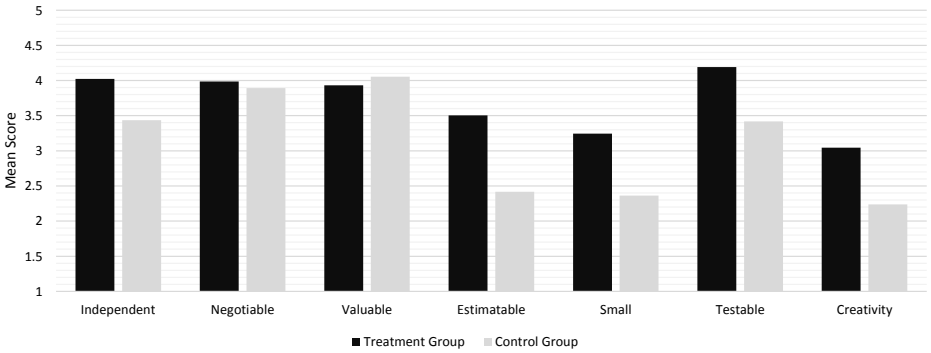

Fig. 4. INVEST and creativity scores that were rated by Scrum experts

To determine the extrinsic value of user stories, the Kano questionnaire [17] was answered by 13 employees adopting the role of future end users and disjoint from the participants in the experimental groups. The results from Fig. 5 indicate that nearly half of the requirements within the treatment group were categorized as attractive requirements. Must-be requirements account for one third, and indifferent requirements for approximately a quarter of all user stories.

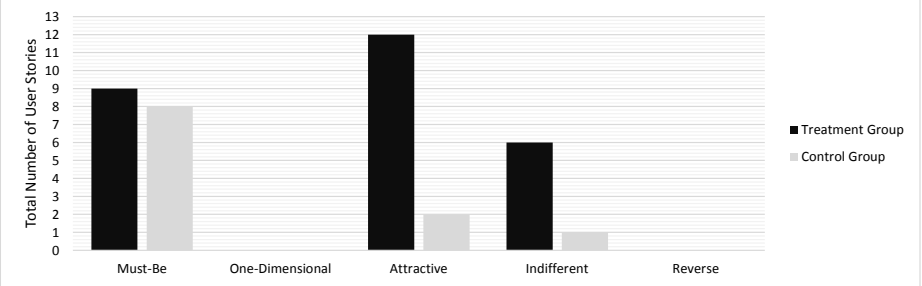

Fig. 5. Total number of user stories per Kano category classified by 13 future end users

Most of the requirements in the control group were prioritized as must-be requirements, followed by a few attractive and indifferent requirements. No requirements were classified as one-dimensional or reverse quality.

Creativity. Creativity was rated by the 5 Scrum experts and was significantly higher in the treatment group compared as well. The average creativity score per group is shown in Fig. 4. The statistical results from SPSS are listed in Table 2.

Creativity strongly correlated with the Kano categories. Higher creative requirements were classified as attractive or indifferent, whereas requirements with low creativity score were classified as must-be [$r(36) = .632$, $p < .001$].

Stakeholder Engagement As per our GREM conceptual model in Fig. 1, we measure stakeholder engagement in terms of emotions, cognition, and behavior. All the statistical results are reported in Table 3.

Emotions. Users interacting with the gamified platform did not report higher positive emotions (PA) than did the control group.

Cognition. The treatment group experienced slightly more flow compared to the control group, according to the Flow Short Scale. However, this difference was not statistically significant.

Behavior. Participants interacting with the gamified platform caused more page visits than did the control group as shown in Table 3. In sharp contrast, the control group wrote more text messages compared to the treatment group. The total number of written messages is shown in Fig. 3 and reported in Table 3.

Table 3. Independent t-test results for stakeholder engagement

	Treatment Group		Control Group			
	M	SD	M	SD	t	p
Emotions	36.800	4.025	37.000	4.000	-.082	.936
Cognition	50.400	7.635	43.333	5.645	1.767	.111
Page Visits (Behavior)	161.000	40.367	88.833	38.338	3.036	.014
Chat Messages (Behavior)	1.000	1.732	24.167	19.995	-2.560	.031

6 Discussion

Our experiment shows that a gameful experience in requirements elicitation can be used to effectively influence user behavior and to increase performance. The obtained results enable an evaluation of the hypotheses H1 and H2:

– **We retain the null hypothesis for H1.** *Emotions* and *cognition* did not exhibit statistical differences between the two experimental conditions, whereas *behavior* did. Stakeholders exposed to gamification were active with requirements production, whereas the control group was intensively collaborating during the operational phase. Therefore, it is not possible to reject the null hypothesis for H1, for stakeholder engagement was high in both groups.

– **We reject the null hypothesis for H2.** Findings from both experimental groups reported significant variations in all sub-dimensions of the performance concept. The treatment group did not only *produce* more user requirements, but their *quality* and *creativity* was higher as well. Performance was indirectly impacted by gamification, which caused a change in the behavioral dimension. Consequently, our second hypothesis provides evidence to be true and therefore, we reject the null hypothesis for H2.

Furthermore, we draw some conclusions on the role of gamification in RE.

Productivity, quality, and creativity *may be* increased by gamification. The treatment group not only *produced more* unique requirements, but their *quality and creativity* were significantly *higher*. Their stories were more independent and written in smaller chunks, and the resulting product backlog allowed for better estimations and testing. Most stories were attractive requirements, which have great impact on *customer satisfaction* [17]. Moreover, they outperformed the control group in creativity: most of their requirements were *more novel*.

Competitive gamification *may* reduce stakeholder communication. On the other side, the control group was very *communicative* during the execution phase of the experiment. The recorded data indicate a continuous discussion from the very beginning until the end of the experiment. This group apparently approached the task more as a team, while interpersonal communication between the other group subjects was barely present. Nonetheless, from the intensive discussion observed in the chat, we are able to deduce that this group was causing *mutual obstruction*. As a result, not only was *creativity* of their user stories *lower*, but also intrinsic and extrinsic *quality suffered* as well.

No differences concerning emotions & cognition were identified. We presume that an *optimal flow* was not present in our game design, because players were not challenged enough throughout the game [8]. A second possible explanation is that the achievement system was too *extrinsically rewarding*, which might have caused an emotional and cognitive decrease [9]. A further conjecture might be that the control group was engaged by a social dimension. While the treatment group was primarily progressing in a virtual game and enhancing their competences, the control group was *socially engaged* in the requirements elicitation process [10].

Collaboration in elicitation *may* have negative consequences. Although positive collaboration is deemed as a key success factor for RE [18], our case has shown that it may also have negative consequences during elicitation. The chat discussion in the control group has probably absorbed people's attention and blocked productivity, in line with the cognitive theory of idea generation [33].

7 Validity Evaluation

We discuss the main threats to internal and external validity, and explain how we dealt with them in our research.

7.1 Internal validity

It refers to the causal conclusion between two variables [4]. Despite our efforts to precisely characterize gamification and its effect on motivation and performance, we cannot claim the GREM model to be comprehensive. However, the use of a control group helps eliminate many potential causal relationships [38].

By choosing a wide set of game mechanics and game elements, in order to support participants with different personalities, we collect limited evidence on the impact size of individual elements. We measured this impact by posing a set of questions regarding the enjoyment of individual elements in the posttest.

To mitigate this risk of poor wording and bad instrumentation, we decided to use standardized questionnaires with high validity and reliability, such as the Short Scale Flow [29] and the Positive and Negative Affect Schedule [36].

Concerning the selection of subjects, we could not perform a random selection, but rather had to use a convenience sampling technique. The experiment was announced on the corporate intranet where people could voluntarily enroll. However, these people already might were intrinsically motivated, which could significantly influence the statistical results. We did ensure, however, that both groups had similar characteristics and professional work experience.

In previous studies on gamification in RE [12], the researchers concluded that the graphical user interface had an impact on user satisfaction. To avoid the same problem, we employed the same aesthetic theme for both prototypes.

7.2 External validity

This type of threat measures the extent to which the obtained results are valid outside the actual context in which the experiment was run. Concerning the experimental condition, the sample size is relatively small to make significant conclusions [13]. Due to the fact that this research project was conducted within a single software engineering company, we were bound to the available resources. On the other hand, it could get confusing to manage user stories on an online platform when too many stakeholders are interacting at the same time. To mitigate this threat, we strove to make the experimental environment as realistic as possible by providing them with a real company internal business case.

The experiment lasted two hours, due to practical constraints. Thus, we cannot draw conclusions on the long-term effect of gamification. Extrinsic rewards were effective in the short-term, but their long-term effect is unknown.

To mitigate the threat of interference between the two groups, we told the groups they would be working on two separate and independent cases, and we did not mention gamification as the treatment we were measuring (see Sec. 5).

8 Conclusion

We have shown how gamification can positively influence the elicitation process in agile RE. We did so by conducting a thorough *controlled experiment* where

the treatment group was given the gamification intervention in the form of game elements added to the elicitation platform. To the best of our knowledge, this is the first controlled experiment that studies gamification in RE.

The success of gamification heavily depends on the choice of game mechanics and game elements, as they can affect different psychological needs. Our experiment shows that an individual leaderboard and the opportunity to win a prize incentivizes competition in a positive manner. Stakeholder rivalries increased requirements production, resulting in higher quality and more creative ideas.

We found that simulating competition with gamification can help gather basic and novel requirements, and contributes greatly to creativity. However, individual leaderboards or activity feeds might not always be the right choice. In later development stages, that focus on the creation of a shared conceptualization [1], more cooperative game elements could be more adequate for the analysis, specification and validation of requirements. Social game elements, such as team leaderboards or team challenges, can stimulate cooperation and collaboration [37]. Thus, we build a new hypothesis to validate in future studies:

H3 While requirements elicitation is positively supported by competitive game elements, cooperative game elements are more suitable for requirements analysis, specification and validation.

Future Research. More experiments are required to generalize the results and the applicability of GREM. First of all, the experiment should be executed again, but with the removal of the chat function. This would prevent the control group from being socially engaged and presumably decrease production blocking [33].

It would be valuable to conduct trials with different sample sizes and game elements. Game mechanics and elements should be tested in isolation and in partial combinations to measure their influence on motivation and behavior.

The experiment can also be repeated using different quality frameworks for user stories. For example, it would be interesting to use the Quality User Story (QUS) framework [22] that defines quality in terms of syntactic, semantic, and pragmatic attributes that go well beyond the simple INVEST mnemonic.

To generalize our claims beyond agile RE, experiments are needed with alternative notations to represent requirements. A particularly interesting facet is to explore gamification for the elicitation of non-functional requirements, either in general or looking at specific aspects such as security. Furthermore, we have not tested the long-term trends with respect to stakeholder engagement.

The GREM model contains no elements that are apply uniquely to the RE field, as it stems from theories from management science, psychology, etc. An interesting direction is to explore GREM beyond software engineering as a general model that relates gamification to performance through engagement.

Acknowledgments

We thank everyone at MaibornWolff for hosting our research; in particular, we are grateful to Franziska Metzger for her support throughout the project and to all the participants in the experiment.

References

1. Abdullah, N.N.B., Honiden, S., Sharp, H., Nuseibeh, B., Notkin, D.: Communication Patterns of Agile Requirements Engineerings. In: Proceedings of the 1st Workshop on Agile Requirements Engineering (AREW). pp. 1:1–1:4 (2011)
2. Attfield, S., Kazai, G., Lalmas, M., Piwowarski, B.: Towards a Science of User Engagement (Position Paper). In: Proceedings of the Workshop on User Modelling for Web Applications (IWUM) (2011)
3. Bano, M., Zowghi, D.: A Systematic Review on the Relationship between User Involvement and System Success. Information and Software Technology 58, 148–169 (2015)
4. Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices (2012)
5. Charette, R.N.: Why Software Fails [Software Failure]. IEEE Spectrum 42(9), 42–49 (2005)
6. Cohn, M.: User Stories Applied: For Agile Software Development. Addison-Wesley (2004)
7. Cohn, M.: Agile Estimating and Planning. Pearson Education (2005)
8. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience, vol. 41. HarperPerennial New York (1991)
9. Deci, E.L., Koestner, R., Ryan, R.M.: A Meta-analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin 125(6), 627–668 (1999)
10. Deci, E.L., Ryan, R.M.: Self-Determination. Wiley Online Library (2010)
11. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From Game Design Elements to Gamefulness: Defining Gamification. In: Proceedings of the International Academic MindTrek Conference. pp. 9–15. ACM (2011)
12. Fernandes, J., Duarte, D., Ribeiro, C., Farinha, C., Pereira, J.M., da Silva, M.M.: iThink: A Game-Based Approach Towards Improving Collaboration and Participation in Requirement Elicitation. Procedia Computer Science 15, 66–77 (2012)
13. Fisher, R.A.: Statistical Methods for Research Workers. No. 5, Genesis Publishing Pvt Ltd (1936)
14. Glinz, M.: Improving the Quality of Requirements with Scenarios. In: Proceedings of the World Congress on Software Quality (WCSQ). pp. 55–60 (2000)
15. Granic, I., Lobel, A., Engels, R.C.M.E.: The Benefits of Playing Video Games. American Psychological Association (2013)
16. Havercamp, S.M., Steven Reiss, S.: A Comprehensive Assessment of Human Strivings: Test-retest Reliability and Validity of the Reiss Profile. Journal of Personality Assessment 81(2), 123–132 (2003)
17. Kano, N., Seraku, N., Takahashi, F., Tsuji, S.: Attractive Quality and Must-be Quality. Journal of the Japanese Society for Quality Control 14(2), 147–156 (1984)
18. Kappelman, L.A., McKeeman, R., Zhang, L.: Early Warning Signs of IT Project Failure: The Dominant Dozen. Information Systems Management 23(4), 31–36 (2006)
19. Koopmans, L., Bernaards, C.M., Hildebrandt, V.H., Schaufeli, W.B., de Vet Henrica, C.W., van der Beek, A.J.: Conceptual Frameworks of Individual Work Performance: A Systematic Review. Journal of Occupational and Environmental Medicine 53(8), 856–866 (2011)
20. Lazzaro, N.: Why We Play Games: Four Keys to More Emotion Without Story. In: Proceedings of the Game Developers Conference (GDC) (2004)

21. Lombriser, P.: Engaging Stakeholders in Scenario-Based Requirements Engineering with Gamification. M.Sc. thesis, Utrecht University (2015), http://dspace.library.uu.nl/handle/1874/317766
22. Lucassen, G., Dalpiaz, F., van der Werf, J.M., Brinkkemper, S.: Forging High-Quality User Stories: Towards a Discipline for Agile Requirements. In: Proceedings of the International Requirements Engineering Conference (RE). pp. 126–135. IEEE (2015)
23. Lucassen, G., Dalpiaz, F., van der Werf, J.M., Brinkkemper, S.: The Use and Effectiveness of User Stories in Practice. In: Proceedings of the International Working Conference on Requirements Engineering: Foundation for Software Quality (REFSQ) (2016)
24. Maslow, A.H.: A Theory of Human Motivation. Psychological review 50(4), 370–396 (1943)
25. North, D.: Introducing Behaviour Driven Development. Better Software Magazine (2006)
26. O'Brien, H.L., Toms, E.G.: What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology. Journal of the American Society for Information Science and Technology 59(6), 938–955 (2008)
27. Radoff, J., Kidhardt, E.: Game on: Energize Your Business with Social Media Games. Wiley (2011)
28. Reiss, S.: Who Am I?: 16 Basic Desires that Motivate Our Actions Define Our Persona. Penguin (2002)
29. Rheinberg, F., Vollmeyer, R., Engeser, S.: Die Erfassung des Flow-Erlebens (2003)
30. Rupp, C., Simon, M., Hocker, F.: Requirements-Engineering und -Management. Hanser Fachbuchverlag, Nürnberg, 5 edn. (2009)
31. Snijders, R., Dalpiaz, F., Brinkkemper, S., Hosseini, M., Ali, R., Özüm, A.: REfine: A Gamified Platform for Participatory Requirements Engineering. In: Proceedings of the International Workshop on Crowd-Based Requirements Engineering (CrowdRE) (2015)
32. Snijders, R., Dalpiaz, F., Hosseini, M., Shahri, A., Ali, R.: Crowd-Centric Requirements Engineering. In: Proceedings of the International Workshop on Crowdsourcing and Gamification in the Cloud (CGCloud 2014) (2014)
33. Stroebe, W., Nijstad, B.A.: Warum Brainstorming in Gruppen Kreativität vermindert. Psychologische Rundschau 55(1), 2–10 (2004)
34. Sutcliffe, A.: Scenario-based Requirements Engineering. Requirements Engineering 3(1), 48–65 (1998)
35. Wake, B.: Invest in Good Stories, and SMART Tasks. http://xp123.com/articles/invest-in-good-stories-and-smart-tasks (2003), accessed: 2015-01-20
36. Watson, D., Clark, L.A., Tellegen, A.: Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. Journal of Personality and Social Psychology 54(6), 1063 (1988)
37. Werbach, K., Hunter, D.: For the Win: How Game Thinking Can Revolutionize Your Business. Wharton Digital Press (2012)
38. Wohlin, C., Runeson, P., Höst, M., Ohlsson, M.C., Regnell, B., Wesslén, A.: Experimentation in software engineering. Springer (2012)