Utrecht University

Maiborn Wolff

Mensch IT

Engaging Stakeholders in Scenario-Based Requirements Engineering with Gamification

Master Thesis

Philipp Lombriser, 4077911
Master in Business Informatics
Utrecht University
p.lombriser@students.uu.nl

1st Supervisor: dr. F. Dalpiaz (Utrecht University)
2nd Supervisor: prof. dr. S. Brinkkemper (Utrecht University)
Project Facilitator: F. Metzger (MaibornWolff)

Notice of Originality

I declare that this master thesis is my own work and that information derived from published or unpublished work of others has been acknowledged in the text and has been explicitly referred to in the list of references. All citations are in the text between quotation marks (“ ”). I am fully aware that violation of these rules can have severe consequences for my study at Utrecht University.

Signed: _____ **Name:** Philipp Lombriser

Date: _____ **Place:** _____

Abstract

The principles behind gamification are not entirely new. The term originated in 2002, but did not gain popularity until 2010. Since then, gamification has become a major trend in many business domains to increase customer loyalty and employee engagement. In requirements engineering, academics and practitioners are also exploring new opportunities to boost stakeholder participation with the aid of game mechanics and game elements.

The purpose of this thesis is to evaluate the effectiveness of gamification in requirements engineering in order to improve stakeholder engagement. We developed an online digital platform for scenario-based RE supported with gamification. Derived from an in-depth literature study, we selected user stories complemented with scenarios from behavior-driven development (BDD) as a method to express stakeholder requirements. Points, badges and leaderboards (PBL) are very common game elements in terms of gamification and are used as a starting point for the artifact design. In total, the platform consists of 17 different game mechanics and elements, which intension is to positively affect intrinsic and extrinsic stakeholder motivation.

Subsequently, the playful prototype is tested in a controlled experiment. A conceptual framework is constructed in order to measure the effect of gamification on user engagement and performance. The experimental findings expose that it is possible to change stakeholder's behavior effectively with gamification. Stakeholders who are exposed to the gamified platform produce more user stories, in better quality and with more creative ideas. The majority of their identified requirements are categorized as attractive, which lead to higher customer satisfaction. However, no differences concerning emotions and cognition between the experimental conditions were identified.

In contrast, the control group was likewise involved in the elicitation phase during the trial, but with collaboration as a dominating factor. Their requirements are qualitative rather poorly rated and contain less creative thinking. Furthermore, most user stories cover basic functionalities, which do not cause any positive customer satisfaction when implemented.

The reason for these different behaviors lies in the competitive environment that was simulated for the treatment group. Various game elements are designed to promote rivalries between players and therefore, diminish interpersonal communication. We hypothesize that creativity was inhibited by production blocking that emerged from the intensive discussion within the control group.

Overall, the study of gamification in RE with the utilization of user stories and scenarios as a mean to gather and document requirements yield positive results. Stakeholders interacting with a gamified version of the prototype not only outperform the control group, but are also more satisfied. Derived from the experimental results we agree that competitive game elements are advantageous for RE elicitation to mitigate collaboration and therefore, prevent production blocking. In sharp contrast, social game elements are favorable for RE analysis, specification and validation where cooperation is demanded.

Acknowledgement

It would not have been possible to write this master thesis without the support of the kind people around me. I am sincerely thankful to my first supervisor, dr. Fabiano Dalpiaz, for his guidance, support and great patience at all times. Our weekly Skype meetings were always very insightful and provided me with new scientific input. Thanks also to my second supervisor, prof. dr. Sjaak Brinkkemper for accepting to supervise me and take great interest in my thesis project.

Furthermore, I would like to thank Franziska Metzger for making the project possible at MaibornWolff and for all her time dedicated to answer my questions. I would also like to thank the people at MaibornWolff who volunteered in the experiment and evaluation, but also all other colleagues for being so friendly and helpful during my 8-month stay.

Finally, I would like to thank my family and friends who supported me at all times; especially my girlfriend who probably is a gamification expert in the meantime as well. ;-)

Table of Contents

1 Introduction 1
1.1 Problem Statement 1
1.2 Research Objective 2
1.2.1 Gamification 2
1.2.2 Scenario-Based RE 2
1.3 Research Questions 3
1.4 Research Method 4
1.4.1 Awareness of Problem 4
1.4.2 Suggestion 4
1.4.3 Development 5
1.4.4 Evaluation 5
1.4.5 Conclusion 5
1.5 Evaluation Method 5
1.6 Relevance 5
1.6.1 Scientific Relevance 5
1.6.2 Social Relevance 6
2 Theoretical Background 7
2.1 Requirements Engineering 7
2.1.1 Requirements 9
2.1.2 RE Process 9
2.2 Scenario-Based RE 14
2.2.1 Scenarios 14
2.2.2 Representing Scenarios 14
2.2.3 Documenting Scenarios 15
2.2.4 Scenario Techniques Compared 24
2.2.5 Selected Technique 25
2.3 RE Quality 26
2.3.1 Quality Assurance for Requirements 26
2.3.2 Verification & Validation 29
2.3.3 Customer Satisfaction 32
2.3.4 Quality of User Stories 33
2.3.5 RE Process Quality 36
2.4 Gamification 37
2.4.1 Concept of Game 37
2.4.2 Concept of Gamification 38
2.4.3 Game Elements 40
2.4.4 Motivation 42
2.4.5 Engagement 46
2.4.6 Design Process 49
2.4.7 Gamification in RE 50
3 Conceptual Framework 53
3.1 Relationships 53
3.1.1 Gamification and User Engagement 53
3.1.2 User Engagement and Performance 54

3.2 Conceptual Model 55
3.2.1 Concepts 55
3.3 Hypotheses 60
4 Artifact Design 62
4.1 System Requirements 62
4.2 Architectural Design 64
4.2.1 Technical Architecture 64
4.2.2 Gamification API 65
4.2.3 UML Class Diagram 67
4.3 Graphical User Interface 69
4.3.1 Front Page 69
4.3.2 Onboarding Program 71
4.3.3 Video-Based Character Animation 72
4.3.4 Animated Personas 73
4.3.5 Add User Story Form 73
4.3.6 Add Scenario Form 74
4.3.7 Mobile User Interface 75
4.4 Game Elements Portfolio 75
5 Experiment 77
5.1 Scope 77
5.2 Design 77
5.2.1 Selection of Subjects 78
5.2.2 Design Type 78
5.3 Process 79
5.3.1 Pilot Study 79
5.3.2 Pretest 80
5.3.3 Operation 81
5.3.4 Posttest 81
5.3.5 Focus Group 81
5.3.6 Analysis 81
5.4 Instrumentation 82
5.4.1 Business Case Description 82
5.4.2 Gamification API Configuration 83
5.4.3 Motivation 87
5.4.4 Game Elements 87
5.4.5 User Engagement 88
5.4.6 Performance 90
5.4.7 Stakeholder Expertise 92
5.5 Validity Evaluation 92
5.5.1 Internal Validity 92
5.5.2 External Validity 93
6 Results 94
6.1 Gamification 94
6.1.1 Emotions 95
6.1.2 Cognition 96
6.1.3 Behavior 96

6.2 Performance 99
6.2.1 Productivity 99
6.2.2 Quality 101
6.2.3 Creativity 108
7 Discussion 111
7.1 Gamification 111
7.1.1 Competition vs. Cooperation 111
7.1.2 Extrinsic Rewards 112
7.2 User Engagement 113
7.2.1 Theory of Flow 113
7.2.2 Extrinsic Motivation 113
7.2.3 Self-Determination Theory 113
7.3 Performance 114
7.3.1 User Stories & Scenarios 114
7.3.2 INVEST & Kano 114
7.3.3 Cognitive Theory of Idea Generation 115
7.4 Hypotheses Evaluation 115
7.4.1 Hypothesis 1 116
7.4.2 Hypothesis 2 116
7.4.3 Revised Conceptual Model 117
7.5 Conclusion 118
7.6 Research Questions Answered 119
7.7 Limitations 120
7.8 Future Research 121
7.8.1 Gamification Is More Than Just PBL 121
7.8.2 End-To-End Gamified Approach 122
7.8.3 Motivational Design 122
8 References 123
9 Appendix 144
Appendix A: Game Elements 144
Appendix A1: Jon Radoff's 42 Fun Factors 144
Appendix A2: Mapping Game Elements to Fun Factors 147
Appendix A3: Mapping Game Elements to Motivation 148
Appendix B: Pilot Study 150
Appendix B1: Questionnaire 150
Appendix B2: Answers 150
Appendix C: Heatmap Example 153
Appendix D: Survey Results 154
Appendix D1: Pretest 154
Appendix D2: Experimental Groups 154
Appendix D3: Reiss Profile Scores 155
Appendix D4: Controlling Motivation 155
Appendix E: Experiment Results 158
Appendix E1: Gamification 158
Appendix E2: Motivation 159
Appendix E3: Emotions 160

Appendix E4: Cognition 161
Appendix E5: Correlation Between Emotions & Cognition 162
Appendix E6: Experienced Difficulty 163
Appendix E7: Overall Satisfaction 164
Appendix E8: Behavior 165
Appendix E9: Chat Messages 166
Appendix E10: Productivity 167
Appendix E11: Social Interest 169
Appendix E12: Quality 169
Appendix E13: Creativity 175
Appendix E14: Relationship Between Creativity & Kano Categories 177

Table of Figures

Figure 1-1. Design Research Process 4
Figure 2-1. Bridging Function (Adapted from Gunter, Gunter, Jackson, & Zave, 2000) 8
Figure 2-2. Requirements Engineering Disciplines (Adapted From Wiegers & Beatty, 2013) 8
Figure 2-3. RE Process Cycle (Dorfman, 1997) 9
Figure 2-4. Use Case Diagram 16
Figure 2-5. Use Case Slicing (Jacobson et al., 2011) 17
Figure 2-6. Activity Diagram 18
Figure 2-7. Message Sequence Diagram Example 19
Figure 2-8. Collaboration Diagram 19
Figure 2-9. State Machine Diagram 20
Figure 2-10. Goal ModeL 21
Figure 2-11. Wireframe 22
Figure 2-12. Storyboard 23
Figure 2-13 User Journey 23
Figure 2-14. Quality Attribute Scenario (Bass, 2007) 28
Figure 2-15. Validation Formality Spectrum (Wiegers, 2001b) 30
Figure 2-16. Taxonomy of Reading Techniques (do Prado Leite et al., 2005) 31
Figure 2-17. Kano Model (Kano et al., 1984) 33
Figure 2-18. Types of Game Thinking 39
Figure 2-19. Amazon Progress Bar (“Amazon,” 2014) 40
Figure 2-20. Pyramid of Gamification Elements (Werbach & Hunter, 2012) 40
Figure 2-21. Bartle's Player Types Model 44
Figure 2-22. Engagement Loop 46
Figure 2-23. Progression Stair (Werbach & Hunter, 2012) 46
Figure 2-24. Gamification Process 49
Figure 2-25. iThink Screenshot (Fernandes et al., 2012) 50
Figure 2-26. REfine Overview (Snijders et al., 2015) 51
Figure 2-27. Gamestorming Core Principles (Gray et al., 2010) 52
Figure 3-1. Conceptual Model 55
Figure 3-2. Impact Size of Gamification on Motivation 58
Figure 3-3. Construct of User Engagement 58
Figure 3-4. Construct of Performance 60
Figure 4-1. Technical Architecture 65
Figure 4-2. Captain Up HUD and Game Center 66
Figure 4-3. UML Class Diagram 68
Figure 4-4. Front Page Screen 70
Figure 4-5. Game Center Screen 71
Figure 4-6. Onboarding Example Page 72
Figure 4-7. Animation Video 72
Figure 4-8. Facial Animation 73
Figure 4-9. Add User Story Screen 74
Figure 4-10. Add Scenario Screen 74
Figure 4-11. Mobile User Interface 75
Figure 5-1. Laboratory Experiment 78

Figure 5-2. Covariance Design 79
Figure 5-3. Experiment Process 79
Figure 5-4. Example of Actions 84
Figure 5-5. Example Of Badges 84
Figure 5-6. Example Level 86
Figure 5-7. Polynomial Function 86
Figure 5-8. Points Required Per Level 87
Figure 6-1. Game Elements 95
Figure 6-2. Positive Affect (PA) 95
Figure 6-3. Cognition 96
Figure 6-4. Page Visits 96
Figure 6-5. Chat Messages 97
Figure 6-6. Socialbility 97
Figure 6-7. Heatmap Treatment Group 98
Figure 6-8. Heatmap Control Group 98
Figure 6-9. Requirements Production 99
Figure 6-10. Accumulated User Stories over Time 100
Figure 6-11. Accumulated Scenarios over Time 100
Figure 6-12. User Story Categories 101
Figure 6-13. Number of User Stories Per Category 102
Figure 6-14. INVEST Evaluation 104
Figure 6-15. Kano Evaluation 105
Figure 6-16. Kano Evaluation within Treatment Group 106
Figure 6-17. Kano Evaluation within Control Group 106
Figure 6-18. Customer Satisfaction Coefficient 107
Figure 6-19. Mean Customer Satisfaction Value 107
Figure 6-20. Creativity 108
Figure 6-21. Mean Creativity of User Stories 108
Figure 6-22. Creativity Over Time 109
Figure 6-23. Creativity Scores Within Kano Categories 110
Figure 7-1. Conceptual Framework 116
Figure 7-2. Proposed Conceptual Model For RE Elicitation 117
Figure 7-3. Proposed Conceptual Model For RE Analysis, Specification and Validation 118
Figure 9-1. Heatmap Example 153
Figure 9-2. Box Plot for Game Elements 158
Figure 9-3. Plot Box for Types of Game Elements 159
Figure 9-4. Box Plot for Emotions 161
Figure 9-5. Box Plot for Cognition 162
Figure 9-6. Correlation between Emotions & Cognition 162
Figure 9-7. Experienced Difficulty 163
Figure 9-8. Box Plot for Severity 164
Figure 9-9. Overall Satisfaction 164
Figure 9-10. Box Plot for Overall Satisfaction 165
Figure 9-11. Box Plot for Page Visits 166
Figure 9-12. Box Plot for Chat Messages 167
Figure 9-13. Message Frequency 167

Figure 9-14. Box Plot for Productivity 168
Figure 9-15. Content Production Frequency 169
Figure 9-16. Box Plot for INVEST 171
Figure 9-17. Box Plot for Customer Satisfaction 175
Figure 9-18. Box Plot for Creativity 176
Figure 9-19. Box Plot for Creativity Per Kano Category 178

List of Tables

Table 2-1. Use Case Description 17
Table 2-2. Scenario Techniques Compared (Rupp et al., 2009) 24
Table 2-3. Characteristics Comparison 34
Table 4-1. User Stories for Artifact Design 64
Table 4-2. Concepts of the RE Platform 69
Table 4-3. Game Elements Portfolio 76
Table 5-1. Subjects Segmentation 80
Table 5-2. Personas Used for Business Case 83
Table 5-3. Actions 83
Table 5-4. Badges 84
Table 5-5. Levels 86
Table 5-6. Satisfaction of Game Elements 88
Table 5-7. Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988) 89
Table 5-8. Short Flow Scale (Rheinberg et al., 2003) 89
Table 5-9. Experienced Difficulty 90
Table 5-10. General Satisfaction 90
Table 5-11. Stakeholder Expertise Survey 92
Table 6-1. Group Comparison 102
Table 6-2. User Story Examples 102
Table 6-3. Scrum Experts 103
Table 9-1. 42 Fun Factors (Radoff & Kidhardt, 2011) 145
Table 9-2. Mapping Game Elements to Fun Factors 148
Table 9-3. Mapping Game Elements to Motivation 149
Table 9-4. Pretest Results 154
Table 9-5. Treatment Group 154
Table 9-6. Control group 155
Table 9-7. Reiss Profile Scores 155
Table 9-8. Independent T-Test for Motivation 157
Table 9-9. Game Elements Statistics 158
Table 9-10. Independent T-Test for Types of Game Elements 159
Table 9-11. Correlation Analysis 160
Table 9-12. Independent T-Test for Emotions 160
Table 9-13. Independent T-Test for Cognition 161
Table 9-14. Correlation between Emotions and Cognition 163
Table 9-15. Independent T-Test for Severity 164
Table 9-16. Independent T-Test for Overall Satisfaction 165
Table 9-17. Independent T-Test for Behavioral Data 165
Table 9-18. Independent T-Test for Chat Messages 166
Table 9-19. Message Frequency 167
Table 9-20. Independent T-Test for Productivity 168
Table 9-21. Content Production Frequency 169
Table 9-22. Independent T-Test for Social Interest 169
Table 9-23. Independent T-Test for INVEST 170
Table 9-24. Correlation between Estimable, Small and Testable 171

Table 9-25. Kano Evaluation Table 174
Table 9-26. Independent T-Test for Customer Satisfaction 175
Table 9-27. Independent T-Test for Creativity 175
Table 9-28. Statistics for Creativity Per User Story 177
Table 9-29. ANOVA Test For Creativity & Kano Categories 177
Table 9-30. Spearman's rho test for Creativity & Kano Categories 178

1 INTRODUCTION

While requirements engineering (RE) is perceived as a crucial activity to deliver high quality software products (Rupp, Simon, & Hocker, 2009), many IT projects nowadays still fail to meet customer's expectations (Charette, 2005). One reason for this is the absence of key stakeholders in elicitation workshops and review meetings, often caused by the lack of interest and time (Kappelman, McKeeman, & Zhang, 2006). To encounter this problem, gamification is a possible mean, whose benefits and strengths have been identified and used by other researchers (Fernandes et al., 2012; Snijders et al., 2014). In general, gamification applies certain game design elements and game mechanics for business purpose to solve real-business problems (Zichermann & Cunningham, 2011). This research examines how stakeholder collaboration, creativity and their motivation can be improved to increase the quality of requirements and the process of requirements engineering.

This chapter is structured as follow: First a brief overview of the occurring problem is described in Section 1.1, followed by a description of the research objective and goals in Section 1.2. The research approach with a list of related research questions and process steps is outlined in Section 1.3 and Section 1.4. Finally, Section 1.5 and Section 1.6 are devoted to the evaluation of the outcomes and social and scientific relevance.

1.1 PROBLEM STATEMENT

Research has shown that changing requirements in later software development phases create significant higher costs than changes in early phases. These late changes can very often lead to a higher risk of project termination (Sharif, Khan, & Bhatti, 2012). Therefore, most software projects start with a requirements engineering phase to gain clear consensus of stakeholders' needs in order to reduce the risk of exceptional changes in later steps of the development progress. This can either be applied to a more straightforward approach as in waterfall, or to an incremental method to enhance agility (Sommerville, 2010). No matter how the process is structured, the quality and performance of the activities in this early phase are important in order to prevent system failure or abandonment.

Although several RE techniques, methods and tools exist to facilitate requirements engineering more efficiently, many software projects nowadays still fail to deliver on time, within cost and expected scope (Charette, 2005). Reasons for software project failures are wrong, unsatisfied or unmet user requirements, often caused by poor stakeholder collaboration and communication, incomplete understanding of needs, insufficient domain knowledge, fluctuating requirements, inadequate documentation and several other factors. Furthermore, the lack of stakeholder participation in requirements elicitation workshops and review meetings are reasons why many IT software projects are never completed (Kappelman et al., 2006). For those with an active role in the company, motivation is often an issue that needs to be addressed. Keeping stakeholders motivated and engaged potentially increases profitability and productivity (McManus, 2004). Since most software products are developed for organizational internal or external clients, it is vital to consider the desires and needs of these parties. The software can only result in a qualitative reliable product when all involved parties show interest in the development process (Baroudi, Olson, & Ives, 1986).

In summary, despite the requirements engineering phase at an early stage of a software project and existing requirements management tools and techniques, it is challenging to involve key stakeholders in the software development lifecycle and keep them engaged in order to improve the overall product quality.

1.2 RESEARCH OBJECTIVE

Stakeholder involvement in the RE process of a project is an important factor to increase the success of a software system. A study performed by Kujala, Kauppinen, Lehtola, & Kojo (2005) shows that there is a strong correlation between user involvement and the success of information system development as well as user satisfaction. Early studies by Baroudi et al. (1986) have shown that a greater user involvement leads to a higher system acceptance and usage.

The aim of this research is **to improve the quality and creativity of requirements by increasing stakeholder engagement in the requirements engineering process**. A possible mean to achieve this objective is with the support of gamification. The idea behind gamification is to use game-design principles and game elements from classical video and board games, for instance points, badges and leaderboards (PBL), and apply them to non-gaming contexts (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). In Section 1.2.1 and 1.2.2 we briefly introduce gamification and scenario-based RE to provide the reader with a rough overview.

1.2.1 GAMIFICATION

Gamification is often used in education for learning and instruction (Kapp, 2012), but the principles may also be applied to other domains (Herger, 2014). Many research papers show that when people are engaged, they are more likely to be productive (Ellis & Sorensen, 2007). Employee motivation and customer loyalty can be increased by gamification from which companies can gain competitive advantage.

The promises of gamification might also be beneficial in the context of requirements engineering. Not much scientific work has yet been conducted in this particular domain. The research approach from Fernandes et al. (2012) was a first step in this direction. The authors developed a gamified software tool to improve collaboration and participation in requirements elicitation workshops. The tool was used for initial requirements elicitation and tested in two case studies with positive results.

1.2.2 SCENARIO-BASED RE

This research tries to apply gamification principles to scenario-based requirements engineering. In general, scenarios are narrative descriptions of contexts to explain how a user can interact with systems containing of complex sequences of events. They allow reasoning about the system behavior and can be presented in different formats (J. Carroll, 2000).

The reason why scenarios are selected for this research is because they apply to initial requirements elicitation and are useful to further analyze and explore logical sequences (A. Sutcliffe, 1998). They enable to focus on how actual users will interact with the product. Scenarios take the user's viewpoint and allow to reason and argument about the system behavior in specific detail. They also enable the validation of implemented requirements against their intended design. Furthermore, they are easy to understand, both by users and requirements engineers, because stakeholders find it easier to express their ideas with real life examples rather than abstract concepts (Glinz, 2000).

Overall, the objective of this research is to make recommendations to improve the quality of scenario-based requirements engineering combined with the theory of gamification. This research project was conducted at MaibornWolff, a software engineering and IT consulting company. They are headquartered in Munich and operating in different industry fields within Germany and Switzerland. The consultants already use scenarios as a documenting technique in form of natural. Furthermore, the management is very interested to improve their consultancy work with the aid of gamification.

The objective of this research study is broken down into following three sub-goals that serves as a guideline for our thesis:

1. Increase stakeholder engagement in requirements engineering
2. Improve the quality of elicited and documented requirements
3. Improve creativity in requirements engineering

1.3 RESEARCH QUESTIONS

From the research objective in the previous chapter the following main initial research question was formulated. In addition, from the main question six sub-questions were derived, each to address a particular problem area.

How can gamification effectively be applied to scenario-based requirements engineering in order to increase stakeholder engagement and improve the quality and creativity in requirements engineering?

RQ 1: *What are the common methods, techniques and tools in scenario-based requirements engineering?*

RQ 2: *What is the potential and the benefits of gamification in the context of requirements engineering?*

RQ 3: *Which activities of scenario-based requirements engineering can be effectively gamified?*

RQ 4: *What game elements do positively support scenario-based requirements engineering?*

RQ 5: *How can the quality of scenario-based RE be measured?*

RQ 6: *How effective is gamified scenario-based requirements engineering in practice?*

The intention behind the first two research questions is to get a current overview of the requirements engineering practices, scenario-based RE techniques and gamification approaches.

While the first two questions are mainly focused on the literature part, the third and fourth questions try to answer how the two topics can be combined. To answer these questions a gamified scenario-based RE solution is developed.

The fifth question tries to figure out what metrics can be used to measure the quality of the collected scenarios and how they can be made comparable.

The purpose of the last research question is to measure the effectiveness of a concrete gamified prototype in scenario-based requirements engineering.

1.4 RESEARCH METHOD

The described issue from the problem statement is occurring in the real world and triggered through a particular situation at MaibornWolff. Therefore, a practice-oriented research is used in order to produce new knowledge and find an applicable solution. With practice-oriented research it is possible to contribute to a successful intervention allowing to change an existing situation (Verschuren, Doorewaard, & Mellion, 2010). Since scenario-based RE is widespread, the proposed solution can be re-used by other industry companies as well (Glinz, 2000).

Because the problem is well known and observed, this project uses design science research (DSR) to develop a new solution. The creation of an artifact during this process plays an important role to meet certain goals (Simon, 1969). According to March & Smith (1995) the outcomes of a construction process under the design science paradigm are constructs, models, methods and instantiations. A prototype based on a conceptual design is developed and evaluated in an experimental condition.

The general methodology of DSR can be represented with a five step process model, which is successively performed throughout this project (Takeda, Veerkamp, & Yoshikawa, 1990). All process steps contribute to the knowledge base and allowed the identification of knowledge gaps. The five-step process model is represented in Figure 1-1 and should be considered as an incremental and iterative cycle. Other researchers, such as Peffers, Tuunanen, Rothenberger, & Chatterjee, (2007) proposed a similar DSR process model.

FIGURE 1-1. DESIGN RESEARCH PROCESS

1.4.1 AWARENESS OF PROBLEM

With the lack of stakeholder participation and motivation in requirements elicitation workshops, new approaches have to be identified to solve the problem of demotivation and disengagement.

An in-depth literature study based on multiple sources is conducted to gain more information about the occurring real-world problem and drawbacks. A conceptual framework is created that serves as a proposal for a design recommendation.

1.4.2 SUGGESTION

Next, creative steps towards a problem solution drawn from the acquired knowledge and theory base are performed. The idea is to reuse existing software solutions and customize them to an individual gamified platform for requirements elicitation. The output of this process step is a recommended solution for a tentative design. Furthermore, a framework is constructed to measure the effectiveness of gamification in requirements elicitation.

1.4.3 DEVELOPMENT

Once a stable and reasonable solution exists, a tentative design is implemented in this phase. The novelty is primarily in the design and not in the construction of the artifact. The appearance and functionality of the prototype has lower prioritization than the design principles of the artifact.

1.4.4 EVALUATION

Once the tentative design is completed, it is evaluated according to a set of criteria that are defined in the conceptual framework. Qualitative as well as a quantitative analysis are used to measure its effectiveness. The data is thoroughly analyzed and interpreted in order to answer our hypotheses. Deviations from expectations are tentatively explained.

1.4.5 CONCLUSION

The last phase of the research cycle addresses the conclusions that are drawn from the research project. The conclusion part comprises a detailed description and interpretation of the evaluated findings. The final results are exposed and anomalous behavior of the artifact was the subject of future research. This part also summarizes the outcomes and is part of the knowledge contribution.

1.5 EVALUATION METHOD

Von Alan, March, Park, & Ram (2004) stated that evaluation is crucial and requires researchers to demonstrate the utility, quality and efficacy of a design artifact using rigorous evaluation methods. By performing evaluation in a real environment and real setting, naturalistic evaluation methods such as case studies, field studies, surveys and action research can be selected (Sun & Kantor, 2006).

The artifact design in this research is tested with a controlled experiment. A two-group experimental design in form of a posttest-only is performed (Bhattacherjee, 2012). Hereby, the first group receives an intervention in form of gamification. In contrast, game mechanics and elements were completely absent within the prototype of the control group. Both groups then have a certain amount of time to gather requirements for a particular business case. The data is compared with statistical methods, such as a two-group independent sample t-test (Scheffe, 1999).

Tremblay, Hevner, & Berndt (2010) proposed to use focus groups for the evaluation of an artifact design in information science (IS). A design researcher not only designs an artifact, but also tries to provide evidence that this artifact solves a real problem. In this research we use focus groups to study the results of the experiment and reason about the requirements quality of the two groups.

The experiment was conducted together with internal employees of MaibornWolff. The intention behind the experiment is to test the prototype with a small group of subjects (6 people per group) and let them interact with the RE platform for several hours.

1.6 RELEVANCE

1.6.1 SCIENTIFIC RELEVANCE

Gamification has been around for a long time, but its awareness has been documented the first time in 2010 (Deterding, Dixon, Khaled, & Nacke, 2011). While several companies already use the principles of gamification, the trend towards a more pleasant environment in software engineering has just started (Dubois & Tamburrelli, 2013). When considering the progress of scenario-based requirements

engineering, it is apparent that there is still a great amount of research to be done in this particular field (Rolland, Achour, et al., 1998).

This research provides a state-of-the-art study on gamification and scenario-based requirements engineering techniques. Furthermore, the artifact design and evaluation demonstrate an example how these two concepts can be combined with respect to scientific value. The results can be used for future research to deeper elaborate in this specific field of interest. This research combines theory and practice and tries to answer the hypothesis that gamification is an enabling factor for enhanced requirements engineering and stakeholder engagement.

1.6.2 SOCIAL RELEVANCE

Gamification has the hidden potential to increase work and learning performance by changing people's behavior in a positive manner. This is mostly done by increasing engagement with the application of game-design elements (Deterding, Sicart, et al., 2011). This phenomenon has the advantage to encourage stakeholders to participate and collaborate more efficiently in requirements engineering workshops (Fernandes et al., 2012). Due to the fact that most software projects stand and fall with stakeholder requirements and expectations, it is very important to get these key people involved in the development process (Charette, 2005). With game design principles, stakeholders can have more fun in RE workshops and deliver valuable insights that otherwise would be missed. Chances are higher that software projects get realized according to project plans and stakeholder expectations (Kappelman et al., 2006).

Furthermore, software companies can gain benefit from a concrete example of how effectively scenario-based requirements engineering is applied in an early phase of a software project. Other companies might try to apply similar approaches for their projects in order to improve requirements engineering and stakeholder engagement. With the combination of the two elements, this research was a first attempt in this direction, resulting in valuable information and lessons learnt that can be reused and applied by academics and practitioners in Information Science (IS).

2 THEORETICAL BACKGROUND

This chapter aims to get a clear picture of the current state of scenario-based requirements engineering and gamification. Different retrieval methods were used for multiple sources. The primary source of information was scientific literature and searches were primarily made with Google Scholar. Aside from these, a number of books pertaining to either gamification or scenario-based requirements engineering were selected. This combination ensured that the research was primarily based on widely accepted and respected works.

Although an initial literature study was conducted at the beginning of the project, the process of selecting the most relevant source material was done in several incremental steps. Forward- and back navigation (snowballing) of references and citations was performed to gain a more complete picture of the scientific work in the particular domain. The starting point of a new topic was always from researchers who gained high reputation and were often cited by others.

The literature review addressed four main topics that were relevant for the artifact design and experimental evaluation. The theoretical background begins with a general introduction and overview of requirements engineering in Section 2.1. Next, Section 2.2 contains methods and techniques to document and analyze scenarios in requirements engineering. Section 2.3 investigates on the topic of requirements verification and validation. Finally, gamification and its principles are introduced and discussed in comprehensive way in Section 2.4.

2.1 REQUIREMENTS ENGINEERING

Requirements engineering is one of the most crucial steps in software development and bridges the gap between business needs and software engineering, or in other words between the problem domain and solution domain (Hull, Jackson, & Dick, 2005). The problem domain is the environment where the system will operate once implemented. When building a new software system or extending an existing one, the first step is always to understand and analyze the current environment.

On the other side, software engineers develop software in the solution domain in order to solve a particular problem. The gathered information (requirements) from the problem domain have to be analyzed and transformed into a solution that satisfies stakeholders' needs.

While the objectives in the problem domain are usually very vague and structured in form of a wish-list, the requirements for the solution domain are well formed and documented. However, their structure and completeness depend on the quality and involvement of the stakeholders who have an interest in the solution. In an ideal world, all requirements would be clearly articulated, without any ambiguity and 100% customer satisfaction. However, reality has shown that this is merely the case and requirements engineers need to possess a broad set of competencies and skills (D. Q. Nguyen, 1998). Figure 2-1 illustrates the bridging function of requirements engineering between the two domains.

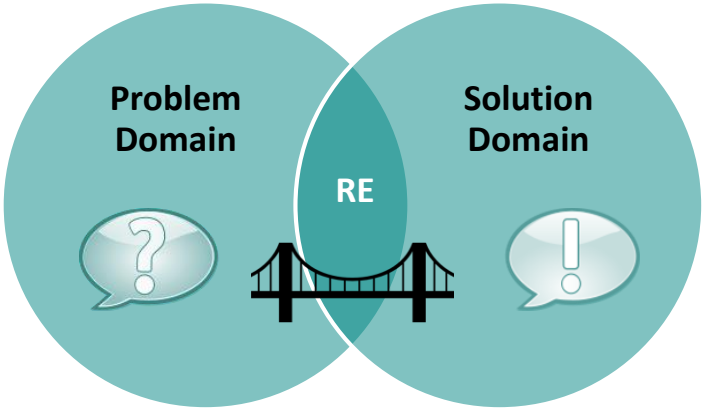

FIGURE 2-1. BRIDGING FUNCTION (ADAPTED FROM GUNTER, GUNTER, JACKSON, & ZAVE, 2000)

The domain of requirements engineering can be split into two disciplines: Requirements development and requirements management (see Figure 2-2) (Wiegers & Beatty, 2013). Requirements development in the Rational Unified Process (RUP) is also known as requirements elaboration (Jacobson, Booch, & Rumbaugh, 1999). Requirements development is further subdivided into elicitation, analysis, specification, and verification and corresponds to the generation of a requirements specification document (RSD) (Dorfman, 1997). Requirements management consists of version control, change control, status tracking and traceability of the individual requirements (Wiegers & Beatty, 2013).

The scope of this research project will be limited to the requirements development part, because the emphasis remains on the requirements acquisition and refinement process. However, the importance of requirements management is acknowledged too.

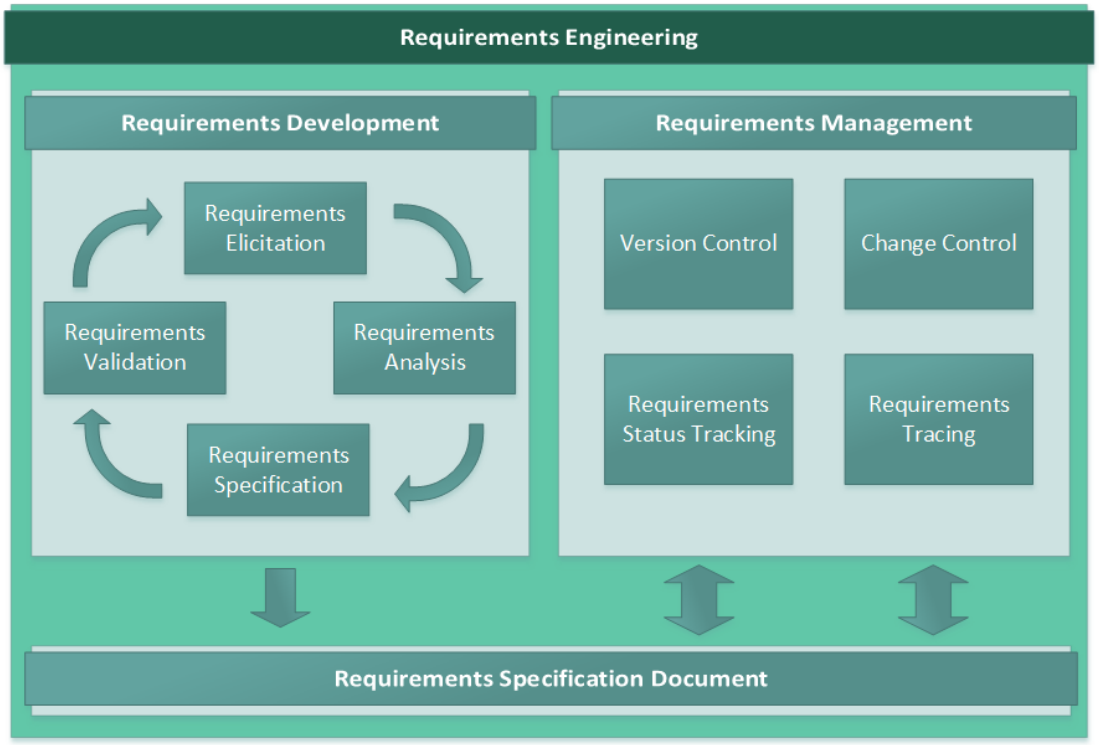

FIGURE 2-2. REQUIREMENTS ENGINEERING DISCIPLINES (ADAPTED FROM WIEGERS & BEATTY, 2013)

2.1.1 REQUIREMENTS

Poor defined requirements are the major cause for software projects to fail. Therefore, requirements engineering is one of the most crucial and complex activities in a software project (Hofmann & Lehner, 2001). Requirements in software engineering are descriptions of actions, behavior and constraints of a system in order to meet stakeholders' needs. Sommerville (2010) classified requirements into functional and non-functional requirements and distinguished between two different levels of abstraction: user requirements, system requirements.

- ***Functional requirements*** define the behavior of the system and what it should do.
- ***Non-functional requirements*** also known as quality attributes (QA) or soft-goals (van Lamsweerde, 2001), describe the system attributes and define the constraints of a system.
- ***User requirements*** define the software functionality from a user perspective. They define what the software has to do to accomplish the user's objectives.
- ***System requirements*** are more detailed description of software functions, services and operational constraints.

2.1.2 RE PROCESS

The process identifying stakeholders and capturing and analyzing their requirements and is called requirements engineering. Pohl (2010) defines requirements engineering as "the process of eliciting individual stakeholder requirements and needs and developing them into detailed, agreed requirements documented, and specified in such a way that they can serve as the basis for all other system development activities". RE is a human-centered process and can be divided into four high-level activities as proposed by Dorfman (1997). An example of the cycle is visualized in Figure 2-3.

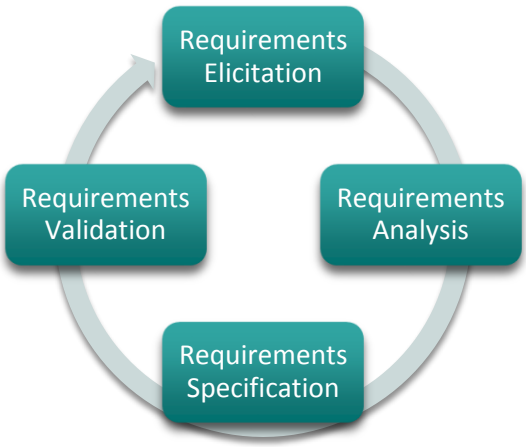

FIGURE 2-3. RE PROCESS CYCLE (DORFMAN, 1997)

REQUIREMENTS ELICITATION

Elicitation is the first step and responsible to identify all relevant key stakeholders and discover what they need. The goal is to gain clear understanding of the problem and identify opportunities to solve it. A stakeholder is defined as an individual, team, organization or classes thereof, having interest in the realization of the system to be developed (ISO/IEC/IEEE, 2011a). Because stakeholders are heterogeneous groups of people their views might often be conflicting. It is fundamental to organize regular stakeholder negations and meetings to reach agreement.

Stakeholders often find it difficult to express their desires and often don't even know what they want. Furthermore, much domain knowledge is implicit and people have different backgrounds and understanding (Christel & Kang, 1992). This makes it very challenging for the analyst to gather the right information and propose a solution that fits everyone's needs.

ELICITATION TECHNIQUES

Many techniques exist to extract tacit knowledge and help the analyst to understand the occurring problem. Van Lamsweerde (2009) differentiates between *stakeholder-driven* and *artifact-driven* elicitation techniques. While the former technique puts the stakeholder into the center, the latter focuses on the system to be developed.

Traditional techniques include generic data gathering such as *interviews*, *questionnaires*, *surveys* and *analysis of existing documents*. Collaborative elicitation techniques usually involve groups of stakeholders to facilitate agreement and buy-in. For instance, *focus groups*, *user groups*, *brainstorming* sessions and *workshops* (Goguen & Linde, 1993).

Cognitive approaches include techniques originally developed for knowledge acquisition, such as *protocol analysis* (Ericsson & Simon, 1984), *laddering* (Reynolds & Gutman, 1988), *card sorting* (Heaton, 1993) and *repertory grids* (Beail, 1985). On the other side, contextual approaches use *ethnography* as an observational technique to understand operational processes and discover implicit requirements (Sommerville, 2010). With *beta customers* a group of users is selected to test a preliminary product version (Regnell & Brinkkemper, 2005).

More advanced techniques include *prototyping* and *model-driven engineering* (MDE). Instead of describing requirements they can be visualized with prototypes. A prototype is a preliminary version of a software product and enables experimentation and exploration at low costs (Smith, 1991). With model-driven engineering abstract models of systems are created as a primary form of expression and later transformed to concrete implementations. With MDE productivity can be increased by simplifying the process of design and promoting communication between individuals and teams working on the system (Schmidt, 2006).

With user-centered design (UCD) end users of the system receive more attention. User requirements are considered from beginning and included in the complete development cycle. (Abras, Maloney-Krichmar, & Preece, 2004). Two modern techniques included in UCD are *storyboarding* (Sutherland & Maiden, 2010) and *storytelling* (Boulila, Hoffmann, & Herrmann, 2011).

This list of techniques is not exhaustive, but rather gives an overview of popular documentation techniques for requirements engineering. The choice of the right technique depends on time, project size, resources and the kind of information that needs to be elicited (Nuseibeh & Easterbrook, 2000).

REQUIREMENTS ANALYSIS

After initial requirements are captured, the next step is to analyze and refine the requirements and constraints. According to IEEE Software Engineering Standards (ISO/IEC/IEEE, 2011b) a documented requirement is qualitative acceptable if it meets the following characteristics:

- ***Necessary***: The requirement is an essential factor for the success of the system
- ***Implementation free***: The requirements are neutral and don't define the solution
- ***Unambiguous***: Every requirement has only one interpretation
- ***Consistent***: No conflict, overlaps or duplications between requirements

- ***Completeness***: Contains all information to define the system
- ***Atomic***: Every statement includes only one requirement and does not contain conjunctions
- ***Feasible***: The implementation of the requirement is technically possible
- ***Verifiable***: The requirement is stated in such a way that it can be tested and evaluated
- ***Traceable***: The life of a requirement can be followed in both forward and backward direction

In practice it is not always easy to fulfill these characteristics. For instance, completeness is very difficult to achieve as long as a system is being developed to meet stakeholders' needs. Furthermore, there is no notion of a requirement to being complete. It is always possible to come up with new requirements (Firesmith, 2005).

REQUIREMENTS ESTIMATION

Requirements engineers should work with developers to discuss the feasibility and efforts of the requirements (Wiegers & Beatty, 2013). This allows to identify risks and conflicts between other requirements. Efforts can be estimated with the support of *use case points* (UCP) (Carroll, 2005). UCP is a forecasting technique to estimate the efforts and costs based on existing system use cases.

Other estimation techniques include *Constructive Cost Model* (COCOMO) (Boehm, 1981), *Function Point Analysis* (Dreger, 1989), *Wideband Delphi* (Boehm, 1981), *Planning Poker* (Grenning, 2002) to name a few.

REQUIREMENTS PRIORITIZATION

Prioritization is a negotiation task between stakeholders and helps to select the right set of requirements for a software release (Berander, 2007). Determining the business impact and prioritizing the elicited requirements are essential during analysis (Sommerville, 2010). In general, requirements are classified into three categories: *must*, *should* and *could*. The first category describes a requirement that must be satisfied in the final solution to be considered a success. The second category represents requirements that add value to the solution, but are not mission critical. The last category describes requirements which are considered desirable but not necessary (Rupp et al., 2009).

To prioritize requirements many different techniques exist. A good and often used technique for software requirements is the *cost-value* method developed by Karlsson & Ryan (1997). A more sophisticated technique is the *Analytic Hierarchy Process* (AHP) (Saaty, 1988). The *100-point* method is a simple approach to prioritize requirements together in groups (Leffingwell & Widrig, 2003). Other prioritization techniques that were developed for requirements engineering are *binary search tree* (BST) (Karlsson, Wohlin, & Regnell, 1998), *planning game* (Beck, 2000), *win-win* (Boehm & Ross, 1989) and *numeral assignment technique* (Karlsson et al., 1998).

REQUIREMENTS MODELING

Requirements modeling is the process of "building abstract descriptions of the requirements that are amenable to interpretation" (Nuseibeh & Easterbrook, 2000). To get a better picture of the system and facilitate communication, Davis (1993) proposed to use models in combination with text. Models bridge language and vocabulary barriers among team members. Furthermore, they can guide elicitation, measure completeness, uncover problems and enhance common understanding (Wiegers & Beatty, 2013).

A well-established and intended modeling language for requirements analysis is the *Unified Modeling Language* (UML) (Booch, Rumbaugh, & Jacobson, 1999). UML was developed by the Object

Management Group (OMG) and accepted as an international standard. The idea behind UML is to simplify and streamline the software development process over the entire engineering lifecycle and enhance the reuse of building blocks. UML offers an efficient framework to model system behavior and interaction of the real world. Furthermore, with the supplement of written text UML can be used to discuss the correctness and completeness of diagrams between different stakeholders. Many types of UML diagrams exist and can be used for different illustration purpose. While some UML diagrams (e.g. use cases and activity diagrams) are more appropriate for RE, other diagrams (e.g. package diagrams) are more convenient for communication in the development phase (Rupp et al., 2009).

Mylopoulos, Chung, & Yu (1999) proposed to move from a more object-oriented to a goal-oriented approach as an alternative concept for requirements analysis. According to the authors, traditional semi-formal modeling techniques like UML have several weaknesses in the elicitation and analysis phase. Most modeling techniques lack to reason about the non-functional requirements. *Goal-oriented modeling* takes a wider perspective. It not only captures the software, but also the environment it operates in. The benefit of goals is that they can be complemented with scenarios. Scenarios describe how the goals are achieved. Moreover, scenarios support the refinement of goals and enable the derivation of requirements. Scenarios are a core part of this research and therefore, discussed in in Section 2.2.

REQUIREMENTS SPECIFICATION

Once the requirements are gathered and thoroughly analyzed, they need to be documented. Sommerville (2010), Wiegers & Beatty (2013) differentiate between two types of documented requirements: *user requirements specification* (USR) and *system requirements specification* (SRS). User requirements are high-level descriptions of the system, whereas system requirements are more concrete statements of the functionality to be performed.

The idea of the document is to decompose the problem into components and serve as an input to design specification. The documentation not only is used by the technical engineers, but also acts as an agreement between stakeholders. The level of detail and content depends on the stakeholders reading the document (Wiegers & Beatty, 2013).

REQUIREMENTS DOCUMENT

ISO/IEC/IEEE (2011b) developed an industry standard for SRS that organizations can use as a basis for their documentation. The document should address at least the following parts:

- ***Introduction***: Purpose of document, project scope, glossary with definitions, abbreviations and acronyms, references and an overview of the remaining document
- ***General description***: product perspective, functions, user characteristics, general constraints, assumptions and dependencies, apportioning of requirements
- ***Specific requirements***: Covering function, non-functional requirements and constraints

DOCUMENTATION FORMATS

To document software requirements there are various formats available (Wieringa & Dubois, 1998). These can either be *formal*, *semi-formal* or *informal*. Formal methods are based on rigorous mathematical reasoning for which the syntax, semantics and rules are explicitly defined. Semi-formal techniques include diagram and tabular techniques that present information in structured form. With informal techniques we use unrestricted natural language to document requirements.

User requirements are often written in natural language and without any reference to a solution, technology or implementation approach (Sommerville, 2010). The advantage of natural language is its simple format and expressivity. However, the downside is that they can easily be misunderstood. Diagrams and tables can help to avoid this drawback. Rupp et al. (2009) suggested using natural language consistently and therefore, structuring it with a template. Another form of structured language is pseudo code. Pseudo code is a mixture between human and programming language and used to specify system requirements (Horowitz, Sahni, & Rajasekaran, 1997). With structured language variability is reduced and requirements are more standardized (Sommerville, 2010).

On the other side, system requirements contain references to technologies and solutions, but should be written so that the functional behavior and non-functional characteristics of the solution are specified without describing how they should be implemented (Wiegers & Beatty, 2013). System requirements not only use natural language, but also semi-formal methods. UML is an enabler to visualize and design the system (Booch et al., 1999). Hereby, notations are structured in way to allow reasoning about the system. Formal mathematical specifications are also used to describe system requirements, but seen less frequently (Wieringa & Dubois, 1998).

REQUIREMENTS VALIDATION

Requirements validation is the fourth activity of the requirements process and contains steps to review the documented requirements. To ensure that the requirements are specified correctly, they need to be assessed by the stakeholders. According to Boehm (1984), validation assures that the characteristics of requirements accurately describe the intended system behavior and properties. E.g. ensuring that a set of requirements is *correct*, *complete*, and *consistent* so that a solution can be built and tested to prove that it satisfies the desired needs.

Validation occurs in incremental steps and includes different techniques. Reviews of requirements artifacts can take a more formal or informal setting. A formal technique includes *inspections*, which is an organized examination of the requirements. The goal of inspection is to check the artifacts in detail and find inconsistencies (Fagan, 2001).

Another successful and more informal technique is called *peer reviews*, sometimes also called *desk-checks*. In peer reviews the artifacts are distributed to a set of stakeholders and feedback is collected. The issues are then discussed within the group (Wiegers, 2002).

Walkthroughs are also less formally organized and are used to check early ideas and often conducted ad-hoc. This activity is considered to be very cumbersome because the requirements and documents are discussed line-by-line to ensure the document competes with everyone's understanding (Wiegers, 2001a).

Prototypes and *simulation models* are useful techniques to validate requirements and ensure help the user understand how the system will behave (Sommerville, 2010). They allow to identify missing or inappropriate requirements during the analysis (Rupp et al., 2009).

When a system or parts of a system is built, it needs to be tested to ensure that stakeholders' needs are met. *User acceptance tests* (UAT) can be performed to evaluate the software's quality (Cimperman, 2006). Hereby, testing is done by the end users before production release.

Another approach to verify that the system was developed correctly is *requirements-based testing* (RBT). In RBT test cases and measurement criteria are derived, designed and executed. The results are verified and defects are tracked and managed (Mogyorodi, 2001).

2.2 SCENARIO-BASED RE

2.2.1 SCENARIOS

While functional requirements are statements describing the expected system services, scenarios describe a condition that could possibly happen in a system process. In software engineering scenarios are described as "an ordered set of interactions between partners, usually between a system and a set of actors external to the system" (Glinz, 2000). Other researchers have similar definitions. For instance, Sutcliffe (2003) defines a scenario as "facts describing an existing system and its environment including the behavior of agents and sufficient context information to allow discovery and validation of system requirements". Van Lamsweerde & Willemet (1998) framed the term to "temporal sequence of interactions among different agents in the restricted context of achieving some implicit purpose".

Scenarios were originally used in theaters for a plot or scene and later for military activities (H. A. Becker, 1983). In business industry scenarios are often used by management for their strategic long-term planning (Mason, 1994). The introduction of scenarios in human-computer interaction (HCI) took place in the late 80s (Young & Barnard, 1986) and gained popularity in requirements engineering with the introduction of use cases (Jacobson, 1992). Scenario-based RE has evolved over time and can be represented and documented with various techniques and formats (Rolland, Achour, et al., 1998).

The main purpose of scenarios is to stimulate thinking about possible events, opportunities and risks in a system. They are often applied to discover and communicate requirements among stakeholders due to their simple comprehensibility (Glinz, 2000). While stakeholders frequently have problems expressing their needs, scenarios are comfortable to tell stories of functional system behavior and expose the underlying problem domain. People find it easier to relate to real-life examples rather than abstract descriptions (Alexander & Maiden, 2005). Scenarios not only facilitate the discovery of new requirements and reason about existing ones, but they also serve as a basis for system and user acceptance testing (Ryser & Glinz, 1999).

2.2.2 REPRESENTING SCENARIOS

The scenario perspective looks at how users interact with the system. System users have goals and scenarios enable the requirements engineer to describe how these goals can be achieved. They ensure that people with different views gain a common understanding (Glinz, 2000). Therefore, scenarios can take many forms and provide various types of information on different levels of abstraction. The adequacy of the representation style depends on stakeholders, the engineering activity and the underlying software system (J. Carroll, 2000). Specification of scenarios can vary between informal descriptions, such as rich narrative texts, or more formal models (Kyng, 1995). Scenarios can be expressed in natural language, diagrams, pictures, mock-ups, prototypes and other many media (A. G. Sutcliffe, Maiden, Minocha, & Manuel, 1998). The spectrum of scenario representations in research and practice includes at least the following (Jarke, Bui, & Carroll, 1998):

- ***Informal representation***: video recordings, literal transcripts, free form text, etc.
- ***Semi-formal representation***: Structured texts, templates and forms, use cases, process maps, message sequence diagrams, activity diagrams, state charts, pseudo code, etc.

- ***Formal representation***: state charts, petri nets, logic of action, etc.

In system development, Alexander & Maiden (2005) defined six common types of scenarios used. These can have different representation style and are defined as follow:

- ***Story***: Narrative description of connected sequence of events, e.g. a user story that is written in plain text as often seen in agile methodologies.
- ***Sequence***: straight-line of interactive steps taken by human or system agents, e.g. List of numbered user actions.
- ***Structure***: More elaborated representation of a scenario, e.g. activity diagram.
- ***Situation***: Snapshot of a future state of the system, e.g. a picture or an example of a user interface of an imagined future state.
- ***Simulation***: Models to explore and animate stories or situations, e.g. animated diagram to show the eventual real effects of alternative conditions and courses of action.
- ***Storyboard***: Drawing or a sequence of drawings to describe a story, e.g. mock-ups of a flow that are linked together.

Scenarios may also be documented on different abstraction levels. They can either take the form of a generic description with abstract classes, or they can be more specific with real examples and concrete objects (instances) (Rolland, Souveyet, & Achour, 1998).

The degree of formality and the right choice of design representation and abstraction level depends on the purpose of the scenarios and the types of stakeholders. Jarke et al. (1998) proposed to specify requirements on a higher level of granularity with formal models when the numbers of scenarios become too many, too broad or too deep. Written free-form textual discussion is handy to explain the initial behavior of a system, but they are not scalable because of the nature of unstructured text. In more complex situations it is better to list functionalities with structured patterns.

2.2.3 DOCUMENTING SCENARIOS

Different scenario modeling and representation techniques are described and compared below and illustrated with an example that was often encountered in the RE literature (Bittner, 2002). Most selected techniques are part of the UML notation language, such as use cases, activity diagrams and state machine diagrams. Furthermore, some scenario-based techniques are applied in user-centered software engineering (UCSE), for instance, user stories, wireframes, storyboards and customer journeys. The reason why these techniques were selected is because of their high acceptance and standardization in software engineering (Abras et al., 2004; Dobing & Parsons, 2006).

PLAIN TEXT

Plain narrative texts are applied when scenarios are initially gathered from stakeholders. The advantage of natural language is that stakeholders do not need to understand any notations to read and understand written texts (Rupp et al., 2009).

However, textual language faces serious quality problems. The sequence of action is not always clear with natural language. This makes it very difficult to draw a borderline between system and environment. Moreover, natural language is not always unambiguous and can easily be misinterpreted wrongly when several people with different backgrounds and knowledge are involved (Glinz, 2000). Nonetheless, scenarios are suitable as a precursor to frame rough requirements. An

adjusted example of a written scenario for an automated teller machine (ATM) could be stated as follow:

> *"A customer arrives at the ATM machine and inserts his bank card. Next, he enters a personal identification number (PIN), which will be sent to the bank for validation. If the PIN is wrong, the machine will eject the bank card. If the PIN is correct the customer will be able to select an amount of money to withdraw cash from the machine. Next, the bank checks the account balance. If the customer has sufficient credit the machine will dispense the cash and eject the bank card. Otherwise, the machine will eject the bank card without giving out any money."*

USE CASE DIAGRAM

A more structured technique include use cases (Jacobson, 1992). Use cases are written from an outside perspective and are generally simple to understand by the stakeholders who possess knowledge about the problem to be solved (Cockburn, 1997). They are an effective mean to model system interactions and have become a UML norm (Jacobson, 1992). Use cases arise when a person (or a system) needs to interact with the system under design in order to achieve a desired goal. The use case is supplemented with information describing the interaction and behavior.

In literature there is no clear distinction between scenarios and use cases. While some people consider that each scenario corresponds to one use case, others define a single thread through a use case as a scenario (Stevens & Pooley, 2006). Cockburn (1997) defines a scenario as sequences of use case steps that represent different paths through a use case. The following example illustrates a use case diagram with associated description of an ATM machine.

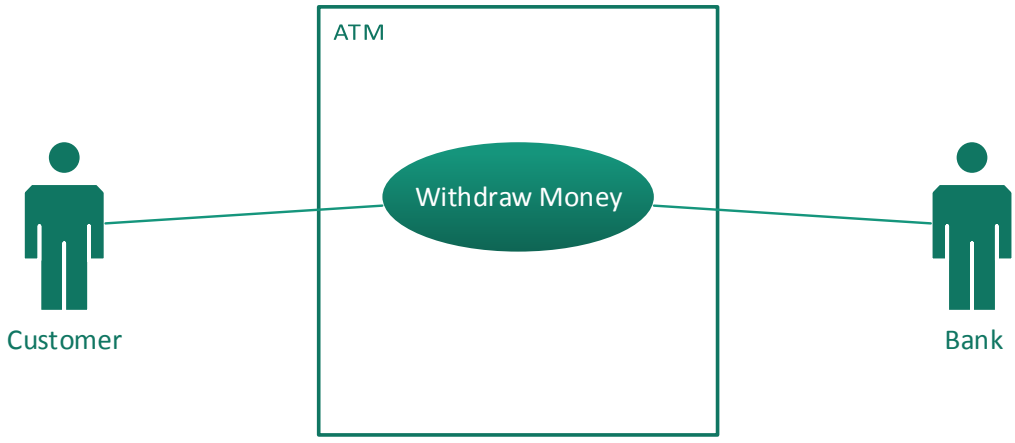

FIGURE 2-4. USE CASE DIAGRAM

USE CASE DESCRIPTION

Use case diagrams are used to sketch the outline of the system, but only supplemented with use case descriptions the sequence becomes logical (Rupp et al., 2009). The following example for withdrawing money illustrates a sequence description in tabular form (see Table 2-1). The advantage of this example is that it is based on natural language and therefore, does not require any notation elements. On the downside, the description of alternative and exceptional flows can become very complex and unclear to the reader and might lead to inconsistencies. Other representation techniques to support use cases are illustrated further below.

Name:	*Withdraw Money From ATM*
Description:	*This purpose of this use case is to describe the steps involved to withdraw cash from an ATM machine.*
Actors:	*Customer, Bank*
Pre-conditions:	*ATM machine is running*
Trigger:	*Bank card is inserted into ATM machine*
Basic flow:	*1. User enters his PIN* *2. Bank authorizes user* *3. User enters desired amount* *4. Bank checks balance* *5. Machine dispenses cash* *6. Machine ejects bank card*
Alternative flow 1:	*1. User enters his PIN* *If PIN is invalid then go to step 6.*
Alternative flow 2:	*4. Bank checks balance* *If balance is smaller than selected amount then go to step*

TABLE 2-1. USE CASE DESCRIPTION

USE CASE 2.0

Due to the great success of use cases in the software industry, Ivar Jacobson and colleagues improved the practice and developed a process for agile development methods (Jacobson, Spence, & Bittner, 2011). This new approach is called Use Case 2.0 and can be used for incremental software projects. The modelling language remains the same as in UML. The changes are more pragmatic and include a process step known as "slicing". Hereby, the use cases with their basic and alternative flows are broken down into smaller pieces of work, making it easier for the software team to tackle. These slices are input for the product and sprint backlogs. An example of how a use case can be sliced is shown in Figure 2-5.

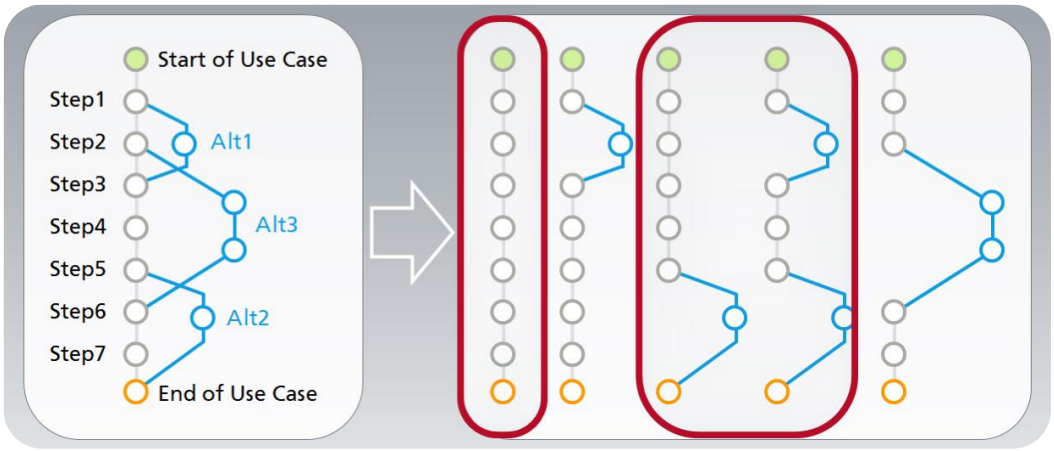

FIGURE 2-5. USE CASE SLICING (JACOBSON ET AL., 2011)

ACTIVITY DIAGRAM

Activity diagrams are related to flowcharts and used to illustrate activities and actions within a business process or system. They describe the functionality of a system from an external view and enable the analyst to present a concrete visualization of a use case (Booch, Jacobson, & Rumbaugh, 2000). Activity diagrams are typically easily comprehensible for both analysts and stakeholders (Podeswa, 2009). They allow to display multiple conditions, iterations, concurrencies and actors within a workflow. The following figure shows an activity diagram of the ATM example.

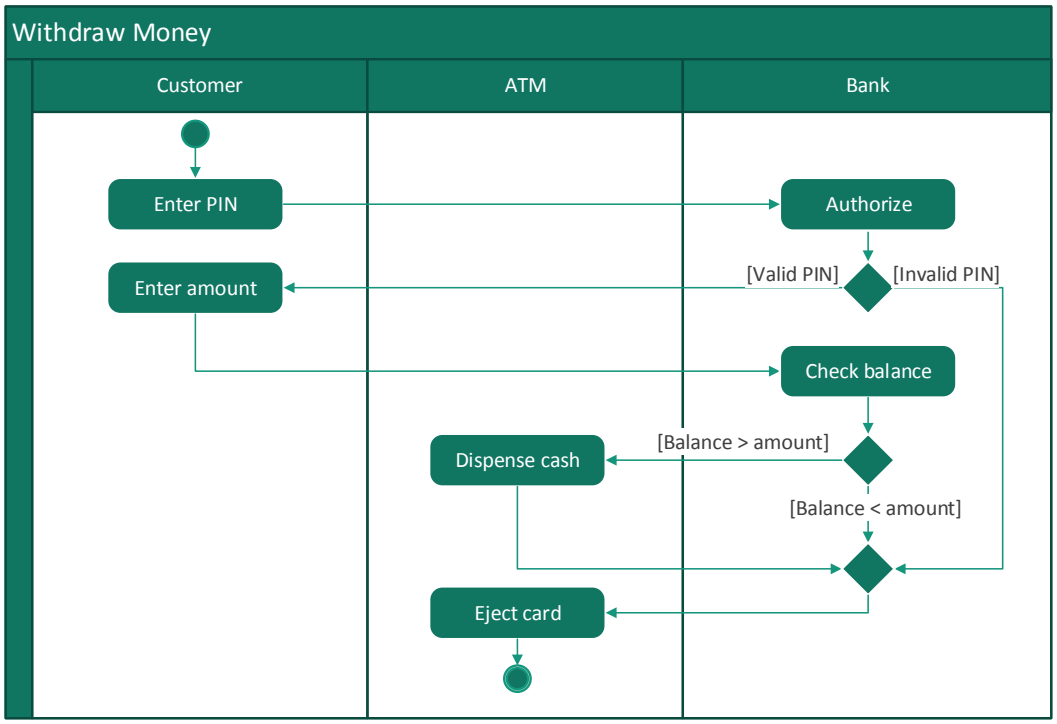

FIGURE 2-6. ACTIVITY DIAGRAM

MESSAGE SEQUENCE DIAGRAM

A message sequence diagram, sometimes also called event diagrams or event scenarios, illustrates the communication and interaction between the different partners in time. These diagrams show the objects and classes involved in the scenario and the sequence of messages being exchanged to carry out the desired functionality. They are read in descending order from left to right.

Sequence diagrams can be very useful for an organization to document how a future system should behave (Booch et al., 1999). During the requirements phase of a project, business analysts can extend existing use cases with a more formal level of refinement. When that occurs, use cases are often refined into one or more sequence diagrams.

Sequence diagrams have many strengths. They allow to demonstrate the interaction between objects and actors in chronological order. Furthermore, the diagram allows to show the responsibility of the different objects. However, sequence diagram tend to get very complex even for simple communication procedures. Furthermore, they can take up a lot of space when modeling (R. C. Martin, 1998). The example with the ATM machine is shown in Figure 2-7 and shows the communication and interaction between the customer, the ATM machine and the bank.

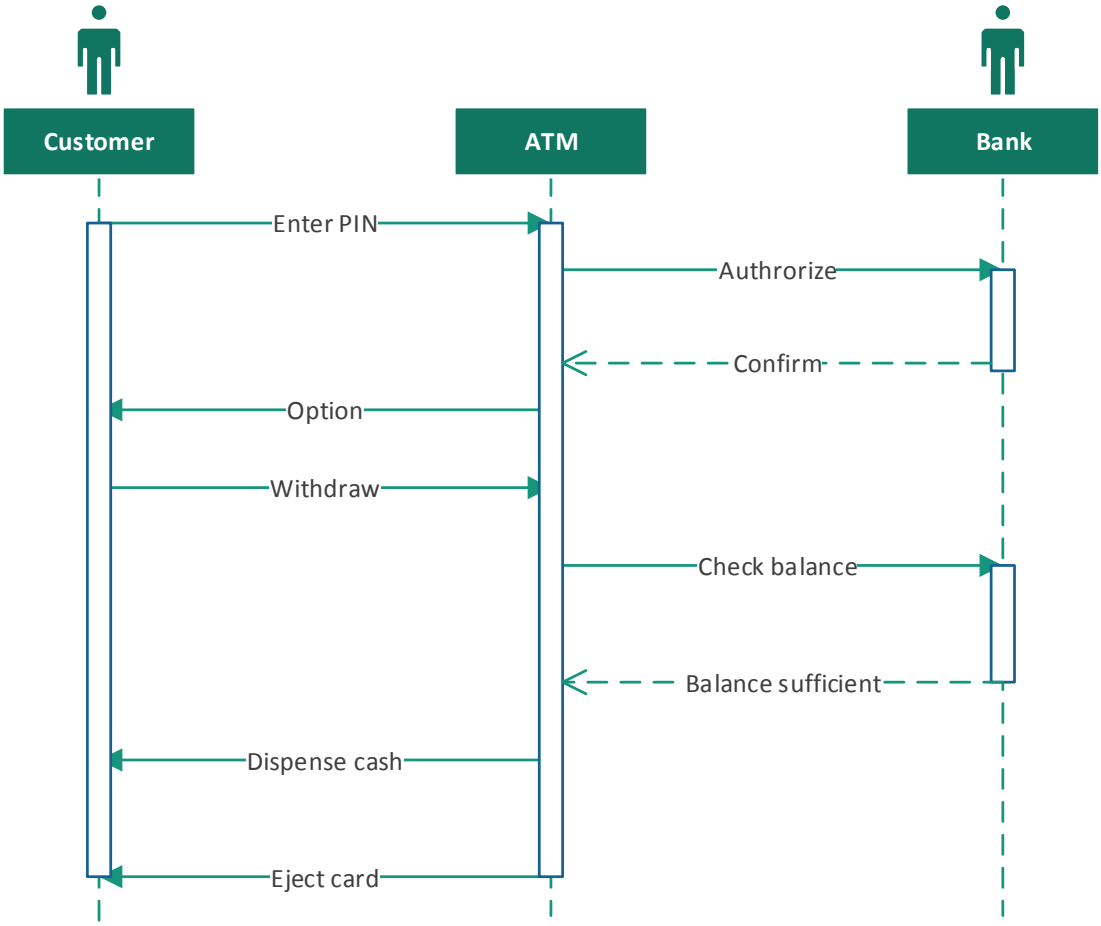

FIGURE 2-7. MESSAGE SEQUENCE DIAGRAM EXAMPLE

COLLABORATION DIAGRAM

A collaboration diagram, sometimes also called communication diagram or interaction diagram, is an illustration of the relationships and interactions among classes and objects (instances), similar to the message sequence diagram. While Sequence diagrams highlight the time, collaboration diagrams aim at demonstrating the order of messages being sent between objects (Booch et al., 2000).

Collaboration diagrams are best suited to the portrayal of simple interactions among relatively small numbers of objects. As the number of objects and messages grows, a collaboration diagram can become difficult to read (Rupp et al., 2009). An example of the ATM machine is shown below. The example shows how objects are statically connected. Icons are used to represent objects and messages are depicted by arrows. Sequence numbers are the numbers allotted to the messages. These numbers show the sequence order in which the messages are passed between the objects.

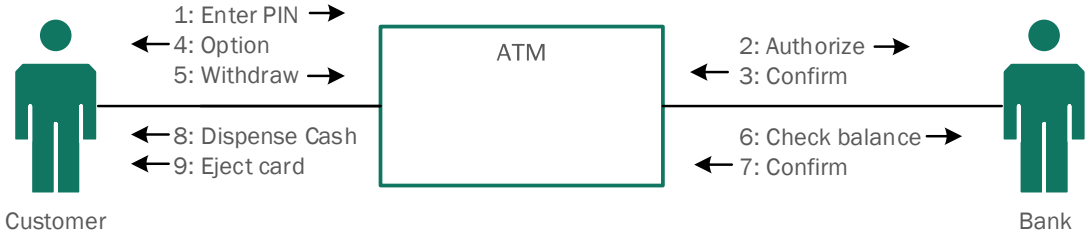

FIGURE 2-8. COLLABORATION DIAGRAM

STATE MACHINE DIAGRAM

State machine diagrams, also just called state diagrams, are used to illustrate the possible states of an object and the causes and effects of those state changes within the object. These diagram are useful to describe classes and understand the behavior of the object through the process. A strength of this modeling notation is that they can be hierarchical: within a state it is possible to define another diagram. This view of modelling is appropriate for real-time systems. Event-based modeling was introduced in real-time design methods such as those proposed by Ward (1986) and Harel (1987). Figure 2-9 shows a simple state machine diagram for the ATM example.

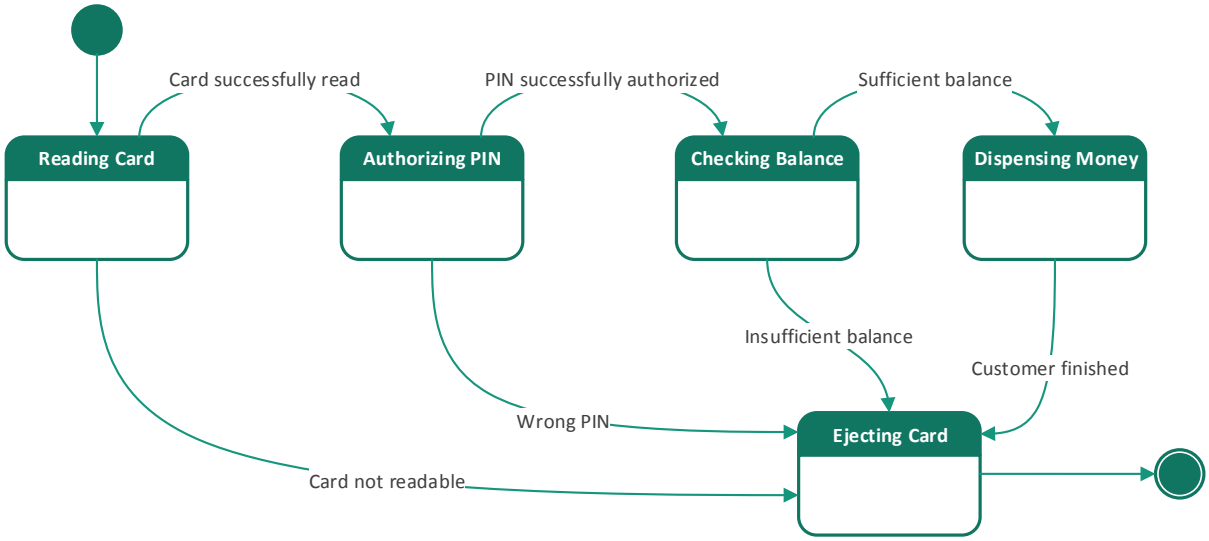

FIGURE 2-9. STATE MACHINE DIAGRAM

GOAL MODEL

Goal-oriented RE (GORE) complements scenario-based RE (SBRE). While SBRE describes *what* a system should do, GORE focuses more on *why* a system should behave in a certain matter (Rolland, Souveyet, et al., 1998). Van Lamsweerde (2000) defined a goal as "an objective that the system should achieve through cooperation of agents in the software-to-be and in the environment". Goals can be presented on different levels of abstraction, where high-level goals describe strategic and organizational wide objectives and low-level goals more technical oriented objectives. Furthermore, goals can be classified into functional (hard) and non-function (soft) goals. Functional goals are concerned with the expected services to be provided; non-functional goals are related to the quality attributes of the service to be delivered. (Mylopoulos, Chung, & Nixon, 1992; van Lamsweerde, 2001).

Many different techniques exist to model goals (Lapouchnian, 2005). Two mainstream GORE methodologies are *Knowledge Acquisition in automated Specification (KAOS)* (van Lamsweerde, 2001) and *i-star (i*)* (Yu, 1993). The i* framework covers both dependencies among actors and goal modeling. On the other hand, KAOS covers goals of all types but is less concerned with the actor's intention. Both frameworks have certain benefits and disadvantages when compared to each other (Werneck, Oliveira, & do Prado Leite, 2009). An example of the ATM machine with a goal model is represented in Figure 2-10. Hard goals are represented with circles and soft goals with clouds.

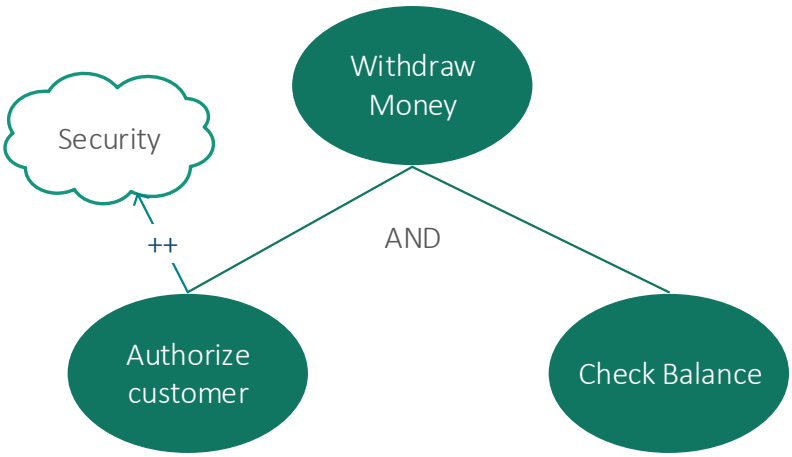

FIGURE 2-10. GOAL MODEL

USER STORY

User stories are narrative descriptions on a high-level of abstraction and express a user's need. (Cohn, 2004). They are mainly applied in agile methodologies to identify the functions a system must provide and to facilitate project management. User stories shift the focus from documenting requirements to discussing them within the project team. They have the advantage of focusing the team on what they are building from an end-user perspective. User stories are less detailed documented than use cases and can range between a use case step and an entire use case (F. Martin, 2003).

Cohn (2004) defined a large user story as an "epic", which barely means a big user story. They are usually written broad in scope, short on details, and require further decomposition before they can be implemented. In addition, the author uses the term "theme" to define a collection of user stories with similar purpose.

While use cases are permanent artifacts and mainly used to document agreement between customer and developers, user stories are written to facilitate agile release and iteration planning. The user story identifies *who* the user is, *what* he needs, and *why* he needs it. For a more formal representation they typically follow a simple template as proposed by Cohn (2004):

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

User stories are written on index cards and arranged on walls or tables to facilitate planning and discussion. There are also digital solutions that support the collection and management of user stories. Each story can be estimated and prioritized by the development team. A great advantage of user stories is that anyone can write them and they are easy to understand by all team members and stakeholders (Alexander & Maiden, 2005).

Personas play an important role in terms of user stories (Cooper, 1999). A persona is a fictional character with a name, picture and other relevant characteristics, which goals have to be achieved by the system. From these personas epics are derived and later turned into concrete user stories (Cohn, 2004). An example of a user story for the ATM machine could be stated as follow:

- *As a customer, I want to withdraw money from my bank so that I don't have to wait in line at the bank.*
- *As a bank, I want to authorize the customer so that fraud can be avoid.*

Stories are an effective and inexpensive way to capture, relate, and explore experiences in the design process. The following list summarizes the benefits of user stories (Quesenbery & Brooks, 2010):

- They help to gather and share information about users, tasks, and goals.
- They put a human face on analytic data.
- They can spark new design concepts and encourage collaboration and innovation.
- They are a way to share ideas and create a sense of shared history and purpose.
- They help to understand the world by giving insight into people who are different.
- They can persuade others of the value of our contribution.

WIREFRAMING

Wireframing is a method used during specification where the identified objects are placed on a user interface. A wireframe represents a simple layer, consisting the structure and navigation of the application as well as the order of the content. It shows the functionality in terms of the relationship with other objects, sets constraints, and includes data type definitions. The wireframe describes the object's expected behavior and exceptional behavior in certain situations (D. M. Brown, 2010).

Wireframes promote communication between stakeholders and the development team by increasing the understanding of functional requirements. Thus, they represent a useful enhancement to the textual and formal scenario approach of use cases with no significant impact on effort. However, wireframes are part of the design process because they describe how the solution needs to behave to meet the requirements. They are already solution-oriented and should be avoided in an early phase of requirements engineering (Garrett, 2010). Figure 2-11 represents an example of a wireframe for the ATM machine.

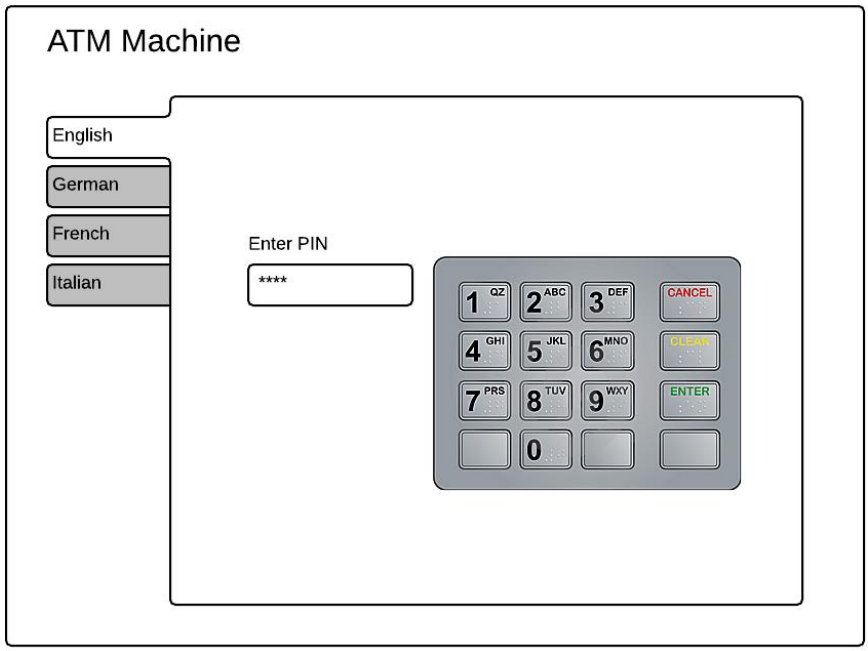

FIGURE 2-11. WIREFRAME

STORYBOARD

Storyboards, also known as user interface-flow diagrams, are part of the UCSE approach (Alexander & Maiden, 2005). They illustrate the interaction required to achieve a user goal. However, instead of using a list of steps in textual form, a storyboard visualizes the interaction of events like a comic book. A storyboard consists of scenes that are related to each other and visualized in sequential steps. They are used to model the interactions that users have with the software system, as in a use case flow. Scenarios become more comprehensible when visual information is added. Storyboards help designers refine their ideas, generate scenarios, and communicate with other people (J. Brown, Lindgaard, & Biddle, 2008). The ATM example in form of a very simple storyboard is illustrated in Figure 2-12.

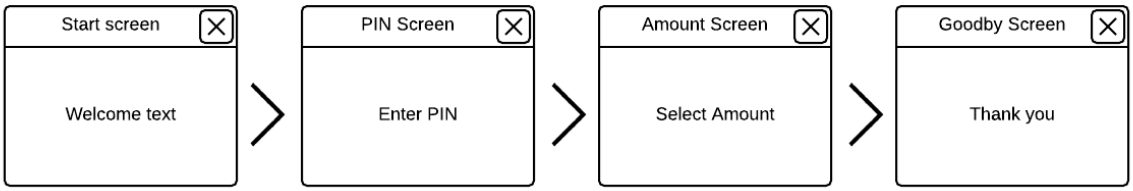

FIGURE 2-12. STORYBOARD

USER JOURNEYS

User journey, sometimes also called customer journeys, is a method derived from marketing to create great experiences for the customer (Nenonen, Rasila, Junnonen, & Kärnä, 2008). They are a design tool to understand product and service interactions from a user point of view. Not much scientific literature was found related to this method. However, from several blogs and reports on the web a trend was identified in the user experience (UX) community (Chris, 2013; Hobbs, 2005).

With this approach a journey of a user or customer is described by representing different touch points. Touch points define the point of contact between a user and the system. A journey is a series of steps which represent a scenario in which a user interacts with the system. They typically are applied at the beginning of a project when new requirements are discovered (Chris, 2013). An example of a user journey showing touch points for the ATM example is visualized in the Figure 2-13.

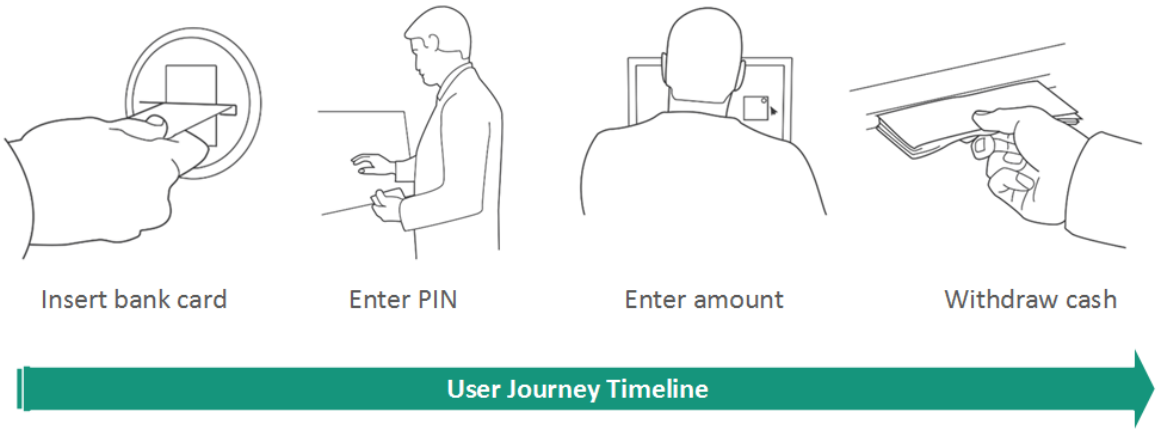

FIGURE 2-13 USER JOURNEY

2.2.4 SCENARIO TECHNIQUES COMPARED

There is no clear and correct answer when it comes to selecting the right method for documenting scenarios. The selection appropriate technique depends on different factors and can be for every project different. According to Rupp et al. (2009) following factors can influence the choice:

- *Acceptance*: Do the majority of stakeholders understand and advocate the technique?
- *Notation skills*: How familiar are stakeholders with the notation of the scenario technique?
- *Specification level*: What level of detail is required to document the scenarios? Rupp defined five different levels of detail:
  - *Level 0*: Rough system description with very abstract descriptions and business goals
  - *Level 1*: Scenarios and user stories to describe business processes
  - *Level 2*: Definition of functional processes that are derived from business processes
  - *Level 3*: Refinement of specification level 2 with more technical details
  - *Level 4*: Technological solutions for implementation are defined on this level
- *Type of system*: What type of system is being developed? More technical or more business oriented?
- *Complexity*: Is the product trivial or highly complex? For very complex products a combination from different techniques might be useful.
- *Consistency*: How difficult is it to maintain changes with the selected scenario technique when requirements change?
- *Unambiguity*: Does the scenario technique allow to document the requirements in detail and without any ambiguity?

The evaluation of the different techniques based on these factors was conducted by Rupp et al. (2009) based on hard facts and heuristics. The following table summarizes the results in a matrix.

Key:

++	very suitable	t	technical oriented
+	suitable	b	business oriented
0	no influence	0-4	specification level
-	not suitable		
--	badly suitable		

	Acceptance	Notation skills	Specification level	Type of system	For complex systems	Consistency	Unambiguity
1. Plain text	++	++	0-1	t, b	- -	- -	- -
2. Use cases diagram	++	+	1	t, b	0	-	-
3. Use case description	++	+	1-2	t, b	0	- -	0
4. Activity diagram	+	-	1-4	t, b	++	0	++
5. Message sequence diagram	-	-	1-4	t, b	-	0	++
6. Collaboration diagram	-	-	1-4	t, b	-	0	++
7. State diagram	-	- -	1-4	t	+	0	++
8. Goal model	++	+	0	t, b	0	0	+
9. User story	++	++	0-1	t, b	-	++	0
10. Wireframing	+	+	1-3	t, b	-	+	++
11. Storyboard	+	+	1-3	t, b	-	+	++
12. Customer Journey	+	+	0-2	b	- -	-	++

TABLE 2-2. SCENARIO TECHNIQUES COMPARED (RUPP ET AL., 2009)

The matrix developed by Rupp et al. does not cover technique 9, 10, 11 and 12. These were separately added and evaluated based on the studied literature.

Because of the simple writing style of user stories and the easy illustration of wireframes, storyboards and user journeys, all four techniques are highly accepted by stakeholders (Quesenbery & Brooks, 2010; Sutherland & Maiden, 2010). User stories are very suitable, because of their quick and easy writing style and simple management.

While user stories only use plain text they don't require any notation skills. On the other side, wireframes, storyboards and user journeys are illustration of user interfaces and objects and therefore, require some notation skills. However, these can be freely selected by the designer.

User stories are positioned on level 0 and 1 for specification level, because they define user goals on an abstract level of detail. Customer journeys are positioned on level 0 to 2, since they can also be used to describe functional processes. Wireframes and storyboards contain more functional information and can cover non-functional requirements, such as usability (S. A. Becker & Berkemeyer, 2002). These two techniques can be used on level 1, 2 and 3.

As industrial projects with agile development teams have shown, all four methods can be applied to either more business or more technical oriented software applications (Cohn, 2004).

For complex systems, all four techniques are less appropriate. They primarily allow to reason about the system behavior from the perspective of a user and not from the system (Abras et al., 2004).

When requirements change it is easy to update and maintain user stories, because they are written on index cards and can easily be shuffled, changed, added or removed (Breitman & Leite, 2002). Wireframes and storyboards are also easy to update and maintain (A. Sutcliffe, 2002) when a computer aided software engineering (CASE) tool is used (Kuhn, 1989). For customer journey it can get more complex to keep the conceptual framework consistent when changes occur.

Unambiguity is considered to be neutral for user stories. Although they are written in plain text and might leave room for interpretation, they have a clear structure and syntax when appropriately applied as proposed by Cohn (2004). Wireframes, storyboards and customer journeys allow to visualize the problem domain with pictures and illustrations. Therefore, ambiguity is not as much a problem, especially when supplemented with textual content (A. Sutcliffe, 2002).

2.2.5 SELECTED TECHNIQUE

We touched upon various scenario documentation techniques, each providing different strengths and weaknesses. While some methods are more suitable for illustration of the interaction between actors and systems (e.g. activity diagrams), other methods allow to better reason about the goals of a user (e.g. user stories). The right choice of the scenario documentation technique depends on several factors as mentioned by Rupp et al.

To conduct an experiment with a gamified requirements engineering tool for elicitation and specification, we decided to select user stories as a mean to document scenarios. User stories are narrative texts and can easily be written by any stakeholder. Clients without any or only minor software engineering skills can write them easily and use them for further analysis. User stories stimulate collaboration among stakeholders in the requirements gathering process and facilitate estimation and prioritization for the project roadmap. An additional benefit of the selected technique is that it gives the development team room for creativity (Cohn, 2004). User stories have gained high

popularity in recent years and are successfully used in agile software development methodologies, such as Scrum or extreme programming (XP) (Ambler, 2014).

This technique makes it is more efficient and effective for the researcher to define metrics and compare the generated stories between the treatment group and the control group. Furthermore, volunteering participants in the experiment don't need to learn any notation and should be able to write effective user stories within a few minutes of practice.

2.3 RE QUALITY

The quality of a software product has a large influence on customer satisfaction and brand value of the product. Quality is the degree to which a product meets customer or user expectations (Radatz, Geraci, & Katki, 1990; Sommerville, 2010). Another definition is provided by Weinberg (1992) who states that "Quality is value to some person". This definition restates that quality is ambiguous as different people ascribe different meanings to the term quality. The strength of this definition is that it allows to raise questions such as "who are the end users of the software product?" and "what is valuable to them?" to be answered by the development teams.

Since tailor-made software products are often driven by stakeholder desires, the management of key stakeholders, as well as the their requirements are an important task (Rupp et al., 2009). Documenting the right requirements with the right level of detail not only impacts the end product quality, but also can reduce the costs and number of changes in later development phases (Boehm, 1981). Therefore, identifying and documenting the right requirements from beginning is a critical success factor for the final software product (Hofmann & Lehner, 2001).

2.3.1 QUALITY ASSURANCE FOR REQUIREMENTS

Quality assurance (QA) is the definition of processes and standards that lead to high-quality end products (Sommerville, 2010). The aim of QA in requirements engineering is to find mistakes and deficiencies in requirements. According to the International Requirements Engineering Board (IREB) (Pohl & Rupp, 2011) a distinction is drawn between three types of quality aspects for requirements:

- ***Content***: Checking for deficiencies regarding their content
- ***Documentation***: Adherence to the documentation rules and guidelines defined for the project
- ***Agreement***: Checking for agreement of all relevant stakeholders

The validation criteria to measure the content are aligned to the ISO standard (ISO/IEC/IEEE, 2011b): However, there are some issues with the content aspect of requirements. For an analyst it is difficult to write complete and unambiguous specifications. Stakeholders may interpret requirements in different ways. In some cases it might even be impossible to reach agreement. Requirements will never be totally complete as long a system evolves to meet the changing needs of its customers and users. Further, it is not always possible to involve all existing stakeholders in the project and very often these requirements are the result of a compromise. Excluded stakeholders may perceive the software system as poor quality because their requirements are not represented (Firesmith, 2005).

REQUIREMENTS METRICS

Rupp et al. (2009) proposed a template-based approach to write requirements to avoid phrasing errors (available in German and English). The authors proposed metrics (e.g. ambiguity, completeness, classification, identification, etc.) to measure their intrinsic quality. These metrics can be used to

measure the syntax structure. Other authors have provided similar metrics, such as the length of requirements in words, the number of subjects, objects and predicates, etc. (Mora M. & Denger C., 2003).

Nowadays, there are also natural language processing (NLP) tools in place to analyze natural written requirements. Significant progress has been made in the past years in NLP, making it possible to extract and analyze information from texts (Miller, 2005). An example of such a tool is QuARS (quality analyzer of requirement specification) (Fabbrini, Fusani, Gnesi, & Lami, 2001). QuARS supports the methodology for the analysis of requirements written in natural language. The tool entails a lexical and syntactical analyzer to evaluate each written sentence and uses a wide set of metrics in order to assess the quality of a requirement specification. The following three features are provided to be evaluated by means of linguistic indicators:

- ***Non-ambiguity***: Detection and correction of linguistic ambiguities in the sentences.
- ***Understandability***: Evaluation of the understandability level of a requirement document and indication of the areas of it needing to be improved.
- ***Specification completion***: Detection and correction of those sentences containing parts to be better specified or needing a more precise formulation.

In automated requirements elicitation recall is more important than precision, because errors of assessment are generally easier to correct than errors of omission (Berry, Gacitua, Sawyer, & Tjong, 2012). The authors claim that "any tool with less than 100% recall is not helpful and the user may be better off doing the task entirely manually". It is simpler for a requirements engineer to evaluate a set of candidate requirements and reject unwanted ones than to browse through an entire document looking for missed ones.

There are metrics to analyze the structure of object-oriented models and establish quality benchmarks to identify potential design problems. Most of them define metrics such as the number of classes, number of use cases, the number of actors or the number of steps to measure size or complexity of requirements. (Abreu & Carapuça, 1994; Chidamber & Kemerer, 1994; W. Li & Henry, 1993). Measuring OO models automatically allows designers to identify weak points in their designs. Uemura, Kusumoto, & Inoue (1999) developed a CASE tool to assess design specifications based on the function point analysis (Albrecht, 1979).

NON-FUNCTIONAL REQUIREMENTS

Requirements quality is not just about whether the functionality has been correctly documented, but also depends on non-functional requirements. In the context of requirements engineering, quality refers to two related but distinct notions (Roger, 2005):

- ***Functional quality***: reflects how well the software complies with a given design, based on functional requirements.
- ***Structural quality***: refers to how the software meets non-functional requirements that support the delivery of the functional requirements.

Non-functional requirements are often called *quality attributes* of a system. Other terms for non-functional requirements are *constraints*, *quality goals*, *quality of service requirements* and *non-behavioral requirements*. In contrast to functional requirements, non-functional requirements define how a system should behave. They significantly influence the product quality of the final software system (Sommerville, 2010).

In literature there are two basic approaches defined to formally treat non-functional requirements, which are product-oriented and process-oriented. The first approach assesses the product after development. The second approach integrates non-functional requirements in the design process (Chung, Nixon, Yu, & Mylopoulos, 2000).

Boehm, Brown, & Kaspar (1978) defined a hierarchical model of software qualities. In the same decade McCall Richards, & Walters (1977) defined a similar set of quality factors. Today, many of these software quality attributes have become an ISO standard (ISO/IEC/IEEE, 2011c).

Quality attributes are hard to specify and are usually stated informally. It is hard to make them measurable and most of the time trade-offs between them are required (Barbacci, Klein, Longstaff, & Weinstock, 1995). For instance, increasing security might negatively impact performance.

Bass (2007) uses a framework with following six steps to specify quality attributes (see Figure 2-14 for visual representation):

1. ***Source of stimulus***: An entity (human, computer system, or any other actor) that generated the stimulus.
2. ***Stimulus:*** The stimulus is a condition that requires a response when it arrives at a system.
3. ***Environment:*** The stimulus occurs under certain conditions. The system may be in an overload condition or in normal operation, or some other relevant state.
4. ***Artifact:*** Some artifact is stimulated. This may be a collection of systems, the whole system, or some piece or pieces of it.
5. ***Response:*** The response is the activity undertaken as the result of the arrival of the stimulus.
6. ***Response measure:*** When the response occurs, it should be measurable in some fashion so that the requirement can be tested.

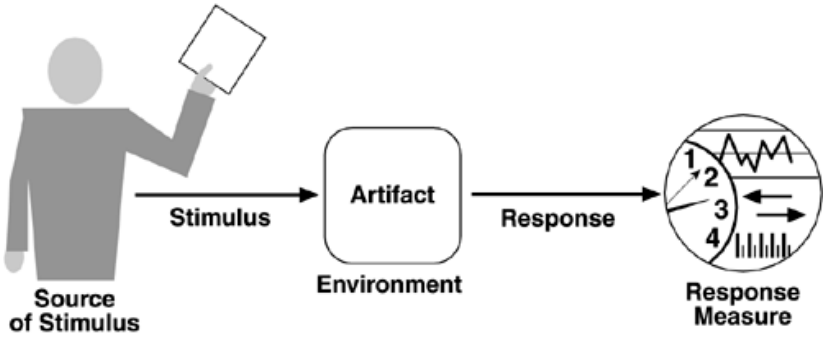

FIGURE 2-14. QUALITY ATTRIBUTE SCENARIO (BASS, 2007)

In GORE non-functional requirements are represented as soft goals, whose satisfaction cannot be established in a clear-cut sense. The main objective of GORE is to iteratively refine higher-level requirements until concrete system requirements are obtained (Yu, 1993). The framework developed by Mylopoulos et al. (1992) is a goal- and process-oriented approach to represent non-functional requirements.

A compositional modeling language to capture and express NFRs was introduced by a research team at the University of Trento, Italy (F.-L. Li et al., 2014). The researchers proposed a goal-oriented method to recursively transform early and informal NFRs to formal requirements specification.

Writing good requirements is not sufficient enough in order to deliver the right software product, but also capturing the right set of requirements is important too (Adrion, Branstad, & Cherniavsky, 1982).

Both documenting the requirements right and documenting the right requirements are critical success factors, which is referred to verification and validation (V&V).

2.3.2 VERIFICATION & VALIDATION

The intension behind verification and validation is to determine that the software works as designed. Finding errors and conflicts in the requirements specification document can avoid system changes and high expenses in later phases. The terms validation and verification are common in software engineering and have following meaning (Boehm, 1984):

- ***Verification*** checks that the software system meets specifications.
- ***Validation*** checks the software system meets the needs of its users.

Verification and validation are closely intertwined. On the one side, verification ensures that the system is being built right. On the other side, validation is the process to check that the right system is being built.

Extending these definitions to requirements engineering, verification ensures that the system will meet its defined requirements, while validation checks that the system will meet stakeholders' needs. In other words, verification determines if the requirements have been specified right by formal proofs (completeness, correctness, consistency, etc.), whereas validation checks if the right requirements have been specified (Wiegers & Beatty, 2013).

This part only covers validation, because the topic of verification was already approached in the previous chapter. However, the review techniques contribute to both verification and validation.

VALIDATION PRINCIPLES

In IREB validation of requirements is based on various principles (Pohl & Rupp, 2011). The following principles ensure that validation effectively identifies errors in requirements:

1. Involvement of the correct stakeholders
2. Separating the diagnosis and the correction of errors
3. Validation from different views
4. Adequate change of documentation type
5. Construction of development artifacts that are based on requirements
6. Repeated validation

VALIDATION TECHNIQUES

Validation is performed by involving relevant stakeholders, other requirement sources (e.g. standards, laws, etc.) as well as external reviewers. There are several techniques for systematic validation of requirements that can also be used in succession.

PEER REVIEW

One of the most core practices in validation to achieve high quality requirements is the review process. The aim of requirements review is to let relevant stakeholders check the requirements documentation and to ensure that the document represents everyone's complete understanding of what is to be accomplished. Several reports confirm its usefulness and many authors propose to use reviews for requirements validation (Ow & Yaacob, 1997; Parnas & Weiss, 1985).

Wiegers (2001b) formulated a spectrum of common used review techniques ranging from formal methods (left side) to informal ones (right side). The different techniques are shown in Figure 2-15 and listed below using Wiegers' definitions.

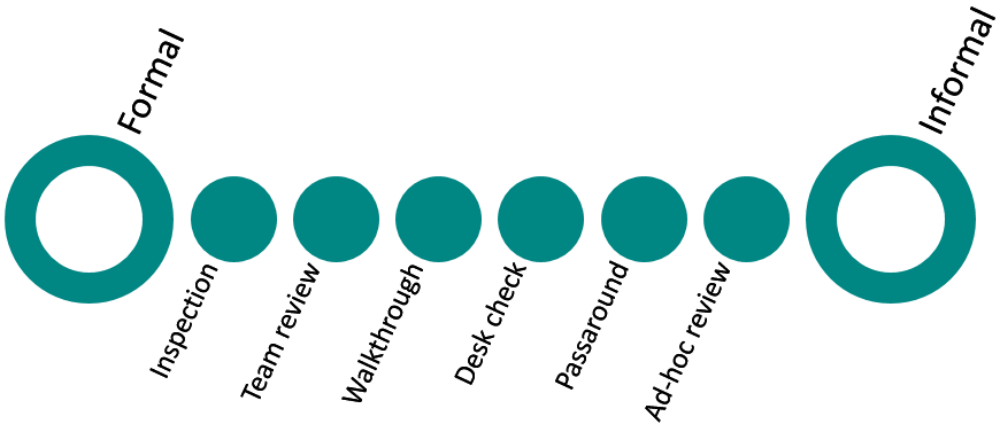

FIGURE 2-15. VALIDATION FORMALITY SPECTRUM (WIEGERS, 2001B)

- ***Inspection***: An organized examination process of the requirements.
- ***Team review***: Similar to inspection, but less formal and less comprehensive.
- ***Walkthrough***: Author leads stakeholders through the requirements document and asks for feedback.
- ***Desk check***: Only one stakeholder examines the document and gives feedback.
- ***Passaround***: Requirements document is sent to stakeholders and feedback is collected.
- ***Ad-hoc review: Author asking a stakeholder for help.***

Informal reviews have no defined process or participant roles. They are usually ad hoc, rather than planned. On the other side, formal reviews define clear objectives and follow a review process. Participants' roles are defined and checklists, rules and methods are used to find defects. Chances are higher that with more rigorous techniques more defects can be found, but this also comes with a higher price to pay (Wiegers & Beatty, 2013).

INSPECTION

Inspections have emerged as one of the most effective quality assurance techniques in software engineering (Laitenberger & DeBaud, 2000). In the meantime different types of inspection techniques emerged, for instance *N-fold inspection*, which uses multiple independent teams (J. Martin & Tsai, 1990). Another inspection technique is called *perspective based reading* (PBR), which tells reviewers exactly what to look for. Each reviewers is offered a viewpoint, from which the requirements should be inspected. This enables the stakeholders to find more defects in less time (Shull, Rus, & Basili, 2000).

In the context of scenarios, do Prado Leite, Doorn, Hadad, & Kaplan (2005) developed a scenario-based reading technique to improve the qualities of scenarios. The technique allows to identify missing, incorrect, ambiguous, contradicting and overlapping information in scenarios. The output is a list of discrepancies, errors, and omissions (DEO). The inspection process is divided into four steps: *plan*, *prepare*, *meet* and *rework*. Their method was applied in 9 different case studies and returned positive results.

Do Prado Leite et al. (2005) developed a taxonomy for different types of reading techniques as shown in Figure 2-16. These techniques can be applied for inspections or peer reviews. Providing reviewers with a reading technique can increase effectiveness of error finding (Rupp et al., 2009).

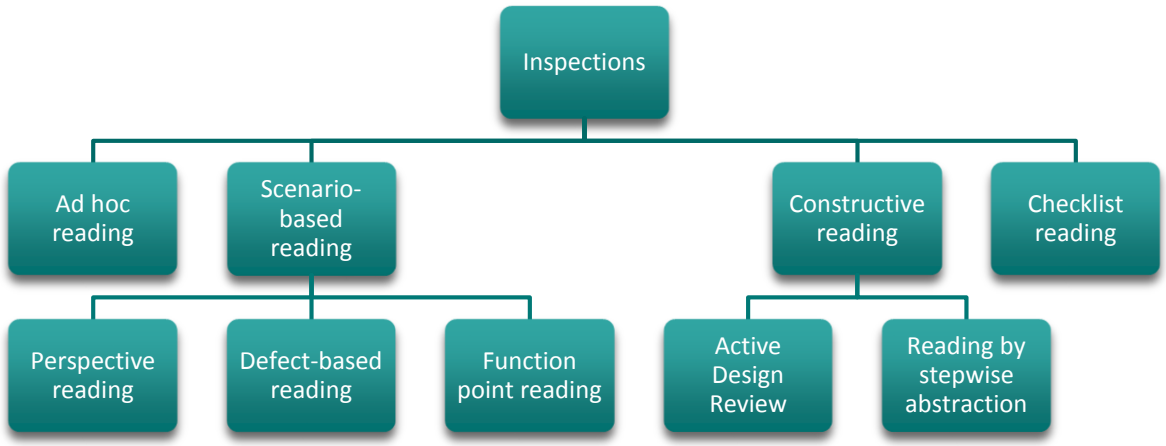

FIGURE 2-16. TAXONOMY OF READING TECHNIQUES (DO PRADO LEITE ET AL., 2005)

PROTOTYPING

Prototyping is a good mean to validate requirements (Smith, 1991). A prototype is an early sample of a product that allows stakeholders to test requirements. They are effective to detect inconsistencies and proof feasibility in an early development phase. Prototypes are excellent for validation by users and customers, because they are more accessible than specifications. They come in different shapes and sizes and can range from paper prototypes to more formal executable specifications (Sommerville & Kotonya, 1998). According to Davis (1993) throw-away and evolutionary prototyping exist.

Sommerville & Kotonya (1998) defined four main activities which should be performed to validate requirements with prototypes:

- Choose prototype testers
- Develop test scenarios
- Execute scenarios
- Document problems

FUNCTIONAL TEST DESIGN

Test cases are a systematic approach to consider each requirement and derive a set of tests for it. Creating test cases is an effective validation technique, because missing or ambiguous information in the requirements description can be identified. They describe the system behavior under specific conditions of inputs, system state, and actions (Jacobson et al., 1999). Test cases not only help to validate requirements, but also to verify that the system functions as specified.

Some software development processes, for instance test-driven development (TDD) even begin with writing test cases before programming (Beck, 2003). Writing test cases first requires the users to consider the expected behavior of the system.

MODEL-BASED VALIDATION

Another form of validation is to transform written requirements into one or more models. For instance, requirements can be transformed into UML class diagrams. With this approach it is possible

to demonstrate that each model is self-consistent and accurately reflects the real requirements of the stakeholders (Sommerville & Kotonya, 1998). Modeling requirements allows to find missing or wrong information in the documentation. However, this technique requires the reader to understand modelling notation skills. Paraphrasing the model as suggested by Rupp et al. (2009) is an effective method to increase comprehensibility. Hereby, (semi-)formal requirements models are transformed into natural language to facilitate understanding.

USER MANUAL DEVELOPMENT

Early availability of a (preliminary) user manual can support the validation of requirements from the perspective of the end user. Writing an instruction manual before it is released requires a detailed analysis and therefore, can reveal requirements inconsistencies (Sommerville & Kotonya, 1998). However, in reality this technique should be avoided. Don Normal, an expert in the field of design and usability stated that "the best designed products won't even need manuals" (Heller, 2001).

2.3.3 CUSTOMER SATISFACTION

Selecting the right set of requirements for product development and implementation largely depends on how successfully the requirement prioritization is done (Berander, 2007). One of the more difficult tasks is helping the stakeholders determine the varying degrees of importance of each requirement. A practical approach to facilitate this process is with the aid of the Kano model, developed in the 1980s and named after Professor Noriaki Kano (Kano, Seraku, Takahashi, & Tsuji, 1984). The model is used to measure the perceived product quality and thus customer satisfaction.

The Kano prioritization approach has gained great attention in agile development to estimate and prioritize the product backlog (Cohn, 2005). It is primarily used for product development and classifies customer preferences into five categories:

1. ***Must-Be Quality***: These requirements are taken for granted when fulfilled, but result in dissatisfaction when not fulfilled.
2. ***One-Dimensional Quality***: These attributes result in satisfaction when fulfilled and dissatisfaction when not fulfilled.
3. ***Attractive Quality***: These attributes provide satisfaction when achieved fully, but do not cause dissatisfaction when not fulfilled.
4. ***Indifferent Quality***: These attributes do not result in either customer satisfaction or customer dissatisfaction.
5. ***Reverse Quality***: A high degree of achievement of these attributes result in dissatisfaction.

Classification occurs by asking the customer a pair of questions for each requirement – one functional and the other dysfunctional – to which the customer selects one of 5 possible answers. The results are then interpreted and evaluated with a frequency table. Eric Klopp, a psychologist at Saarland University, proposed following three calculations for the interpretation of the outcomes (Klopp, 2014):

- ***Category strength*** defines how unambiguous a classification in categories is (Löfgren & Witell, 2005).
- ***Total strength*** defines if more than 50% of the customers consider a product feature as necessary (Löfgren & Witell, 2005)
- ***Fong test*** defines the statistical significance of the classification assignment (Fong, 1996).

Over time a delightful and innovative requirement (Attractive) will shift from exciting to performance (One-Dimensional) and then to an essential basic need (Must-be). Figure 2-17 shows an image of the

Kano model. The horizontal axis represents customer satisfaction and the vertical axis to which degree the requirement is fulfilled.

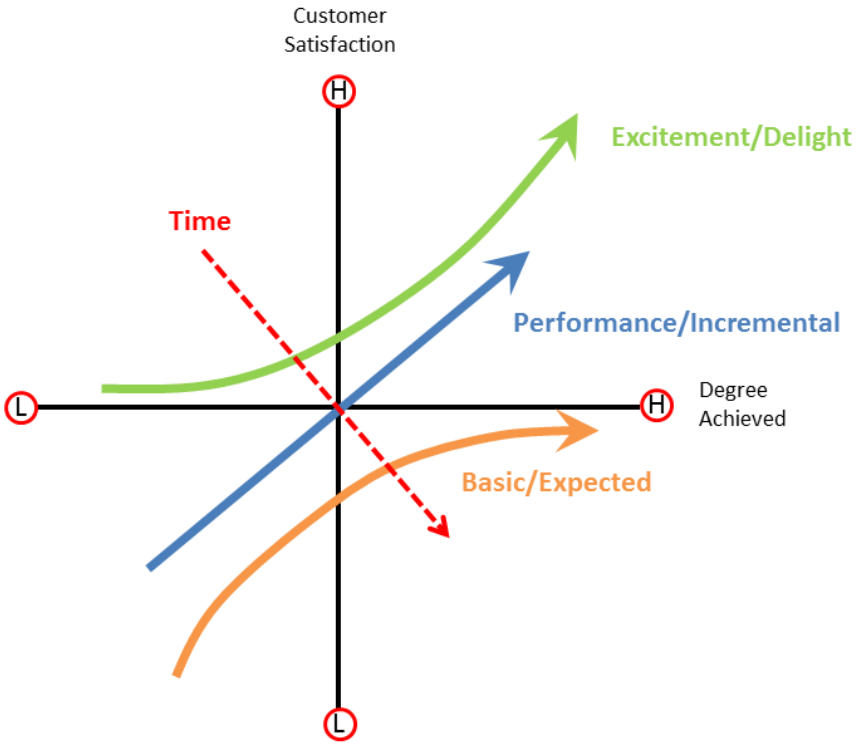

FIGURE 2-17. KANO MODEL (KANO ET AL., 1984)

CUSTOMER SATISFACTION COEFFICIENT

An important indicator to measure the perceived quality of a requirement is the customer satisfaction coefficient (CS coefficient). The coefficient states how strong satisfaction can be increased with the presence of a product requirement or to what extent dissatisfaction can be avoided (Berger et al., 1993). A positive coefficient (CS+) ranges from 0 to 1 and the negative coefficient (CS-) from -1 to 0. CS+ values close to 1 have a high influence on customer satisfaction. CS+ values close to 0 imply that there is very little influence. On the other hand, CS- values close to -1 influence customer dissatisfaction if the product requirement is not fulfilled. CS- values close to 0 signify that the feature does not cause dissatisfaction when not met.

CS+ is calculated by dividing the sum of attractive (A) and one-dimensional (O) column by the total number of attractive, one-dimensional, must-be (M) and indifferent (I) responses. For the calculation of CS-, the must-be and one-dimensional columns have to be added and divided by the same denominator. The result must be multiplied with the factor -1 for a negative value (Berger et al., 1993).

- ***Extent of satisfaction:*** $$CS^{+} = \frac{A + O}{A + O + I + M}$$
- ***Extent of dissatisfaction:*** $$CS^{-} = \frac{O + M}{(-1)\cdot(A + O + I + M)}$$

2.3.4 QUALITY OF USER STORIES

Due to the fact that this research selected user stories as a mean to document scenarios, this sub-chapter elaborates in more detail on the quality and characteristics of user stories.

STRUCTURE OF A STORY

User stories are the building blocks of an agile software project. They show the goals of a project in a user-centered way, making them accessible to both stakeholders and team members. Cohn (2004) defined a tripartite structured format that all user stories should follow. These include the user's role, the user's goal and the user's motivation. In practices, user stories adhere to the following syntax:

- "As a [**role**] I want to [**goal**] so that I can [**motivation**]"

The *goal* defines a feature to be implemented. *Motivation* is the value that will be returned. The *role* defines a person who will directly benefit from the feature.

CHARACTERISTICS

Bill Wake (2003), a senior consultant at Industrial Logic, created a mnemonic as a reminder of the characteristics of a good quality user story. The model is known as *INVEST* and defines the following characteristic:

- ***Independent***: User stories should be self-contained with no dependency to other user stories.
- ***Negotiable***: User stories can always be changed and they capture the essence, not the details.
- ***Valuable***: User stories must deliver value to the end user.
- ***Estimable***: It must be able to estimate the size of user stories.
- ***Scalable***: User stories should not be so big as to become impossible to plan/task/prioritize.
- ***Testable***: User stories must provide the necessary information to make testing possible.

Comparing the characteristics of traditional requirements (ISO/IEC/IEEE, 2011b) to those from user stories (Wake, 2003), it is apparent that most characteristics overlap, but with different names. However, there is a major difference between requirements and user stories. While requirements are supposed to be complete, user stories must remain negotiable. This difference exists, because user stories originated from agile methodologies, where requirements are expected to change. The comparison between the two types are summarized in Table 2-3.

Requirements Characteristics		User Story Characteristics
Necessary	=	Valuable
Implementation free		
Unambiguous		
Consistent	=	Independent
Completeness		
Atomic		
Feasible	=	Estimatable
Verifiable	=	Testable
		Negotiable

TABLE 2-3. CHARACTERISTICS COMPARISON

Next to these characteristics, traditional requirements are usually focused on the capabilities and constraints of the system. Good requirements describe how the system should behave and are

therefore, written from the system perspective (Wiegers & Beatty, 2013). In contrast, user stories are focused on stakeholder goals rather than system functionalities. They are written from a user perspective (Cohn, 2005). Lucassen, Dalpiaz, Brinkkemper, & van der Werf (2015) proposed a Quality User Story Framework with 14 different criteria to assess the quality of user stories. Furthermore, the researchers developed a conceptual model to improve the quality of raw user stories by exploring defects and deviations in user stories.

ACCEPTANCE CRITERIA

When writing a user story, it is advised to include acceptance criteria. Acceptance criteria complement the user story, state when it is complete, and ensure that it is testable. They are a set of statements that specify how the system should behave to meet the user's expectations (Cohn, 2004). Jeffries (2001) wrote about three critical aspects of user stories that are known as the three C's:

- ***Card***: First, the user story is written on cards that has the purpose to remind everyone what the story is about.
- ***Conversation***: Next, the user story itself is communicated from customer to programmers through conversation.
- ***Confirmation***: Finally, to ensure that the user stories are finished and working as intended, they require some acceptance criteria.

Writing acceptance criteria at the beginning of each iteration helps the team to define how a feature or piece of functionality will work from the user's perspective and remove existing ambiguity. Jackson (2013) defined in his online article seven characteristics that define a good acceptance criteria:

- ***Clear***: Shared understanding between team members about the meaning of a criteria.
- ***Simple***: Criteria is stated in language that everyone understands.
- ***Unambiguous***: The expected outcome clearly states what is expected and what not.
- ***Actionable***: Acceptance criteria must be easily translated into test cases.
- ***Implementation free***: The criteria must state the intent and not the solution.
- ***Rough***: They should be relatively described on a high level of detail.
- ***Independent***: The criteria should be independent of the implementation.

An approach used in behavior-driven development (BBD) – which originated from test-driven development (TDD) (Beck, 2003) – to write plausible acceptance criteria was introduced by (North, 2007). Dan North, a coach and consultant in agile development, provided a scenario template to define acceptance criteria. The template allows stakeholders to define the scope of the story by determining which scenarios matter and which are less useful. Once all scenarios have been fulfilled, the story is considered to be finished. The following guideline is provided to formulate scenarios:

- A scenario starts by specifying the initial condition that is assumed to be true at the beginning.
- It then states which event triggers the start of the scenario.
- Finally, it states the expected outcome, in one or more clauses.

The scenario template can be basically used by any stakeholder and has the following form:

```
[User Story]
    Scenario [x]: [Title]
        Given [context] and [more context]...
        When [event occurs]
        Then [outcome] and [another outcome]...
```

Given describes the state of the system when the user starts interacting with it. ***When*** describes the actions the user performs and ***then*** addresses the output of the system. For the ATM example the following scenario could be stated for the withdraw user story:

> As a customer, I want to withdraw money from my bank so that I don't have to wait in line at the bank.
>
> **Scenario** 1: Negative Account balance
>
> **Given** the account's balance is 0 or below
>
> **When** the account owner attempts to withdraw money
>
> **Then** the bank will deny the withdraw
>
> **and** return the account owner a warning message

2.3.5 RE PROCESS QUALITY

El Emam & Madhavji (1995) conducted 30 interviews with companies performing requirements engineering. The goal of the case study was to identify factors to measure the success of the RE process. Overall, 34 criteria were identified and grouped with a cluster analysis. The analysis returned three dimensions of RE success which are:

- ***Cost effectiveness of the RE process***: This dimension is concerned about the whether the resources used in the process is reasonable.
- ***Quality of RE products***: This dimension covers the quality of requirements documents that are produced during the RE phase.
- ***Quality of RE service***: This dimension is concerned with user satisfaction and the extent to which the recommended solution fits the organization.

The results indicate that not only well written requirements contribute to higher software quality, but also the quality of delivered services can have a significant impact.

A newer study on success factors for requirements engineering was conducted by Hofmann & Lehner (2001). The authors compared different factors in requirements engineering and software literature and then performed a cluster analysis. Furthermore, the authors conducted three case studies using action research. Their conclusion of their work reports that the success of an organization-wide adoption of an RE process depends on the following seven main factors:

1. Motivation, commitment and enthusiasm of personnel
2. Usefulness of the RE process
3. Practicality of RE process
4. Training
5. Support
6. Implementation strategy
7. Improvement activities (incremental improvement)

Overall, the researchers identified two characteristics for a successful RE program. First, the process must be considered useful by its stakeholders. If they see the benefits, they will be more likely to become motivated and committed to applying it in practice (Zahran, 1998). The second characteristic is the practical aspect of implementation, meaning that a process must be both simple and flexible to use (Armour, 2001).

2.4 GAMIFICATION

2.4.1 CONCEPT OF GAME

Games have been around for thousands of years and scientists attest that games are a universal part of human experience and present in all cultures (Finkel, 2007). “Senet” for instance, is one of the oldest known board games from the ancient Egypt and dates back to 3100 BC (Piccione, 1980).

Many definitions exist that describe the concept of a game. An earlier definition of game was given by the French sociologist Caillos (1958). He defined a game as “an activity which is essentially free (voluntary), separate (in time and space), uncertain, unproductive, governed by rules, make-believe. More recent definition states that "a game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" (Salen & Zimmerman, 2004).

There is still a big debate in literature what games are, but there are many commonalities in these definitions. The researcher Juul (2003) compared various meanings and proposed the following definition: “A game is a rule-based formal system with a variable and quantifiable outcome, where different outcomes are assigned different values, the player exerts effort in order to influence the outcome, the player feels attached to the outcome, and the consequences of the activity are optional and negotiable”. In his research he found six common features that are necessary for something to be a game. These are:

- ***Rules***: Determine actions, the rights and responsibilities of the players, and each player’s goals.
- ***Variables and quantifiable outcome***: The rules of a game must provide different possible outcomes.
- ***Valorization of outcomes***: Some of the possible outcomes must be better than others.
- ***Player effort***: Games are challenging, contain a conflict or are interactive.
- ***Player attached to outcome***: There is a convention by which the player is attached to specific aspects of the outcome.
- ***Negotiable consequences***: A game can optionally be assigned real-life consequences

For this research we use this definition of games. because of the elaborative research and reasonable arguments provided by Juul (2003).

To distinguish between game and play this research follows the concepts described by Caillos (1958), which are defined as *paidia* and *ludus*. While paidia (playing) describes free-form, expressive, unstructured and spontaneous activities, ludus (gaming), on the other side, structured activities with explicit rules and determined goals.

GAMERS

To understand the effect of games and how playing is anchored in our DNA, we first take a closer look at the population who are playing video games. With the invention of the digital computer in the 30s, video games were moments later born and gained recognition with the release of Pong in the 70s (Kent, 2010). Today people of all ages play video games. There is no longer a stereotype of game player. The average gamer today is 31 years old and has been playing video games for 14 years. 48% of all players are women and are one of the industry's fastest growing population. Furthermore, with the evolution of the internet and the invention of tablets and smartphones the number of casual and social gamers has significantly increased over the past year (ESA, 2014). According to McGonigal

(2011) more than half a billion people worldwide are playing computer and video games at least one hour a day. This results in 3.5 billion hours a week.

Gaming has also started to turn into a sport, known as e-sport (electronic sports), allowing professional gamers compete against each other in organized video game events (Hutchins, 2008). Tournaments such as League of Legends World Championship, Evolution Championship Series, and Intel Extreme Masters provide live broadcasts and cash prizes. The 2014 League of Legends World Championship in South Korea was held in a stadium with more than 40,000 cheering fans. In total, 32 million people were following the spectacle worldwide (Segal, 2014).

These are just some statistics that show the power of games. They can be highly addictive and have a strong pull factor (Griffiths, 2008). This chapter is concerned about the hidden potential behind games and its effect on people's motivation and behavior.

2.4.2 CONCEPT OF GAMIFICATION

The principles behind gamification have been around forever, but the term itself originated in 2002. At that time gamification was "applying game-like accelerated user interface design to make electronic transactions both enjoyable and fast". The idea was that the more an interface feels like a fun game, the more likely users would embrace it (Marczewski, 2013b).

Only since 2010 gamification has started to become mainstream and is today defined as "the application of game-design elements in non-gaming contexts" (Deterding, Sicart, et al., 2011). Other experts enhanced this definition to "the use of game-thinking and game mechanics in non-game contexts in order to engage users and solve problems" (Zichermann & Cunningham, 2011). According to Burke this definition of gamification is still very vague and loosely defined, leading to confusion and misconceptions. Therefore, he specified the definition to "the use of game mechanics and experience design to digitally engage and motivate people to achieve their goals". The key elements of Burke's (2014) definition are:

- ***Game mechanics*:** The use of game elements such as rules, design and tools intended to produce a specific way in which players interact with a game (gameplay).
- ***Experience design:*** The journey players take with elements during the game.
- ***Digitally engage:*** Players interact with computers, smartphones, wearable monitors or other digital devices, rather than engaging with a person.
- ***Motivate people***: The goal of gamification is to motivate people to change behaviors or develop skills, or to drive innovation.
- ***Achieve goals***: Gamification enables players to achieve their goals. When organizational goals are aligned with player goals, the organization will achieve its goals as a consequence of players achieving their goals.

In this research the definition proposed by Burke is preferably used, because it not only describes the intention behind gamification, but also includes the behavioral aspect.

To separate gamification from games, the following diagram illustrates the difference between these two concepts. The two-dimensional matrix consists of different types of game thinking mapped against design purpose and gameplay (Marczewski, 2013a).

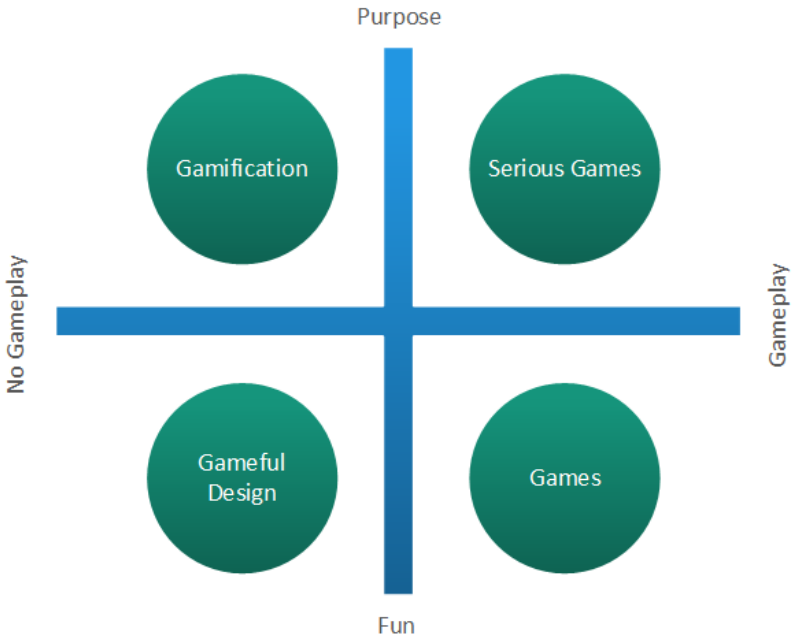

FIGURE 2-18. TYPES OF GAME THINKING

Games have a certain gameplay and have the primary goal to be engaging for the player. On the other hand, *gamification* has no real gameplay, but its purpose is used to solve real problems. *Gameful design* is the use of game thinking in user experience to make the interface easier to use. Aesthetics receive high attention in order to increase usability, rather than the addition of game elements. Games whose primary purpose goes beyond entertainment are called *serious games*. These types of games are like real games, but designed to solve real problems (Schreiber, 2009).

BENEFITS

Gamification has become mainstream (Gartner Inc., 2012) to increase customer loyalty, but also improve employee engagement (Werbach & Hunter, 2012). The reason for this trend is because games are very powerful and have a strong pull factor (Lazzaro, 2004). Games are entertaining and can address specific problems or teach a certain skill (Rosser et al., 2007). Although playing video games increases the risk of becoming addictive (Griffiths, 2008), scientists have proven that video games also come with some advantages: Gaming enables to produce positive emotions, relieve stress, create stronger social relationships, give a sense of accomplishment and improve cognitive skills (Granic, Lobel, & Engels, 2013).

An example of a well-known gamified system is the frequent flyer program. The frequent flyer program is a loyalty program offered by many airlines. Hereby customers collect miles (points) for each distance they travel by plane and later redeem these miles for free air travel or other goodies and services (Kearney, 1990).

A less complex example of a successful game element is the progress bar in e-commerce. Progress bars are effective to give people an indicator of the process completeness. They encourage people to work towards a goal, rather than cancelling their order. The reason is because people are driven by goals and they want to accomplish them (Myers, 1985). An example of the Amazon progress bar is shown in Figure 2-19.

FIGURE 2-19. AMAZON PROGRESS BAR ("AMAZON," 2014)

Gamification is booming across a variety of sectors, including education (e.g. Zondle), healthcare (e.g. Ayogo), task management (e.g. HabitRPG), sustainability (e.g. Nissan Leaf), crowdsourcing (e.g. Fold.it), and user-generated content (Stack Overflow) (Deterding, Sicart, et al., 2011). A proper designed gamified environment has the potential to enhance organizational business value by addressing problems related to human behavior (Herger, 2014).

2.4.3 GAME ELEMENTS

Gamification does not create games, but rather takes certain game elements from classical video games and applies them to solve a business problem related to human factors (Burke, 2014). The Amazon progress bar is one of many successful examples of game components that can be used to influence people's behavior.

In game literature and on the web there is still a lot of disagreement about the correct definition of game mechanics and dynamics. This research will use the framework provided by Werbach & Hunter, (2012). This decision is justified with the reason that their framework was specific designed for gamification. Furthermore, it is very often referenced in literature and highly accepted in the gamification community.

The framework is structured in form of a pyramid with three levels of abstraction (see Figure 2-20). Lower level elements are tied to one or more higher-level elements.

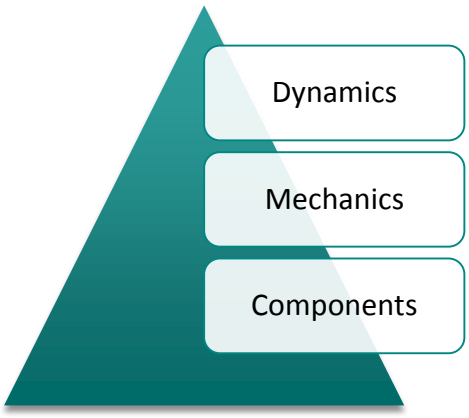

FIGURE 2-20. PYRAMID OF GAMIFICATION ELEMENTS (WERBACH & HUNTER, 2012)

- ***Dynamics*** are on the highest level of abstraction. They are results from the use of mechanics in game play and try to satisfy player's basic desires. The can be seen as hidden 'grammar' in gamification. Examples of dynamics are constraints, emotions, narrative, progression, relationships, etc.
- ***Mechanics*** are compound of components and move the action forward to create pleasure. These can be understood as 'verbs' of gamification. Examples of mechanics are challenges, chance, competition, cooperation, feedback, rewards, transactions, etc.

- ***Components*** are placed on the bottom level of the pyramid and are the specific instantiations of mechanics and dynamics. They are the 'nouns' that make up a game. Examples are points, badges, leaderboards, levels, quests, teams, etc.

POINTS, BADGES AND LEADERBOARDS

The classic triad of game elements used in gamification is called Points, Badges and Leaderboards (PBL). Most gamification programs apply these components because they are very effective, easy to implement and help to achieve business objectives (Werbach & Hunter, 2012; Zichermann & Cunningham, 2011).

- ***Points*** are tangible, measurable evidence of accomplishment. They encourage people to perform actions by collecting them. They help to keep track of score, provide feedback, and display progress. An example of scores are the experience points (XP) that gamers receive while they progress.
- ***Badges*** are a more comprehensive version of accumulated points. They are a visual representations of achievements and are important for social status, as for instance, in the military or Boy Scouts.
- ***Leaderboards*** allow players to compare themselves on a high-score list an ordered ranking. Leaderboards should only be used when competition is essential (Herger, 2014).

In a literature study performed by Hamari, Koivisto, & Sarsa, (2014) a variety of other game elements were encountered. The list was extended with more game elements found by our own literature study ("Game Mechanics," 2014; Werbach & Hunter, 2012). The following list summarizes common types of game elements:

- ***Levels***: help to make tasks more challenging as the user gains new skills and experiences.
- ***Storytelling***: Background narratives, responsible for the dramatics and arouse emotions.
- ***Chance***: Elements of randomness.
- ***Goals***: A clear objective to be achieved by the player in order to receive rewards.
- ***Feedback***: Information about how players are performing.
- ***Rewards***: Items or virtual currencies or goods that are presented to the player when a certain task is achieved.
- ***Progress***: A dynamic in which success is granularly displayed and measured through the process of completing itemized tasks.
- ***Challenge***: Also known as quests, challenges usually implies a time limit or competition whereas a player must overcome a journey of obstacles.
- ***Epic Quest***: An epic-quest is more challenging quest and can only be won within a team.
- ***Avatar***: Allows the user to create an individual virtual character with different types of attributes that can be personalized to his interests.
- ***Status***: The rank or level of a player. Players are often motivated by trying to reach a higher level or status.

All these dynamics, mechanics and components allow for a more compelling user experience, motivating players to take specific actions (Zichermann & Cunningham, 2011). Understanding the effects of game elements on players' behavior and feelings requires us to elaborate more on the topic of motivation.

2.4.4 MOTIVATION

Gamification can leverage motivation in all of us and has the power to let people and organizations achieve their goals (Deterding, 2012). The word motivation originally comes from the Latin term *motivus*, meaning "a moving cause". The Medical Dictionary for the Health Professions and Nursing defines motivation as a "psychological force that moves a person to act to meet a need or achieve a goal" (Stedman, 2011).

Gamification is linked with a number of psychological concepts, especially regarding motivation. Deep fluency and understanding of this concepts is one of the most important key to proper gamification implementation (Herger, 2014; Loftus & Loftus, 1983).

DOPAMINE

To understand the intentions and incentives behind people's behavior we first need to trace the source of motivation to its origins, namely our brain. One reason why so many people like to play games is because they experience positive emotions (e.g. happiness). Scientists have proven with the positron emission tomography (PET) scans that dopamine (happiness hormone) is released in the human striatum when people play video games (Koepp et al., 1998). Dopamine is naturally produced by our brain and makes us feel good. Insufficient dopamine gives us the feeling of depression; an excess deficiency is associated with Parkinson (Pratt, 2011).

Dopamine plays a major role in reward-motivated behavior. Pleasurable activities increase the level of dopamine in the brain (Nestler & Carlezon Jr, 2006). Former research claimed that dopamine is released after obtaining something that satisfies us. However, latest scientists have provided evidence that dopamine is released before we obtain a reward (Salamone & Correa, 2012). Therefore, dopamine is released in order to encourage us to achieve something good or avoid something bad. A study from the Vanderbilt University showed that motivated people have a higher dopamine level in brain regions that impact motivation and reward (striatum and ventromedial prefrontal cortex). In contrast, people who were less motivated showed increased dopamine response in the insular cortex, a brain region responsible for perception, motor control and self-awareness (Treadway et al., 2012).

Now that we know that dopamine is linked to motivation, the next step elaborates more on the concept of psychological needs of humans.

PSYCHOLOGICAL NEEDS

People have needs and desires that motivate them to take action in order to satisfy them (Pink, 2011). Maslow's pyramid of needs is one of the earliest theories describing people's motivation (Maslow, 1943). The theory focuses on what motivates people and defines five levels to maximize satisfaction needs. These levels are known as (1) physiological needs, (2) safety needs, (3) love and belongingness needs, (4) esteem needs and (5) self-actualization needs.

Professor Steven Reiss at Ohio State University conducted a study with more than 6,000 people to identify human needs (Reiss & Havercamp, 1998). With his research results he identified 16 basic desires that guide nearly all human behavior and define our personality. These are:

- ***Acceptance***: the need for approval
- ***Curiosity***: the need to learn
- ***Eating***: the need for food
- ***Family***: the need to raise children

- ***Honor***: the need to be loyal to the traditional values of one's clan/ethnic group
- ***Idealism***: the need for social justice
- ***Independence***: the need for individuality
- ***Order***: the need for organized, stable, predictable environments
- ***Physical activity***: the need for exercise
- ***Power***: the need for influence of will
- ***Romance***: the need for sex, elegance or aesthetic
- ***Saving***: the need to collect
- ***Social contact***: the need for friends (peer relationships)
- ***Social status***: the need for social standing/importance
- ***Tranquility***: the need to be safe
- ***Vengeance***: the need to strike back/to win

The database today includes more than 40,000 people from North America, Europe, and Asia ("Reiss Motivation Profile," 2015).

FUN FACTORS

Social game designer Jon Radoff described in his book ("Game on") a list of 42 trustworthy activities that people enjoy doing (Radoff & Kidhardt, 2011). These fun interactions are cross-tabbed to Steven Reiss' 16 human desires (see Appendix A1: Jon Radoff's 42 Fun Factors). While the list is not primarily focused on games, it is still a useful framework when designing game games. Nevertheless, the drawbacks are that the author does not provide any scientific evidence considering validation and reliability. A sufficient explanation why exactly these 42 factors were chosen is also lacking in his book.

SELF-DETERMINATION THEORY

Theories in cognitive psychology are known as self-determination theory (SDT). SDT is concerned with people's inherent tendencies and psychological needs. It focuses on the degree to which an individual behavior is self-determined and self-motivated, without any external interference (E. L. Deci & Ryan, 1985). SDT identified three intrinsic psychological needs to be satisfied in order to energize motivation (Connell & Wellborn, 1991):

- ***Autonomy*** gives us the freedom of choice and allows us to act according to our own values and beliefs. Autonomy is high when individuals feel they are doing something because they choose to do so, not because they feel pressured (E. L. Deci & Ryan, 1980).
- ***Competence*** refers to seek control of the outcome and experience. Successfully meeting an optimum challenge will directly impact competence (E. Deci, 1975).
- ***Relatedness*** is the need to feel connected to others and to feel being part of a group. Motivation can improve in people if they feel a warm, accepting atmosphere (La Guardia, Ryan, Couchman, & Deci, 2000).

Ryan, Rigby, & Przybylski (2006) analyzed the impact of video games on these three psychological needs. Their results support the hypothesis that video games can effectively satisfy people's desire. Experimental subjects felt that the enjoyment of games was caused by psychological need satisfaction.

World of Warcraft (WoW) is the most popular massively multiplayer online role-playing game (MMORPG) (Byrne, 2014). Comparing WoW with the three psychological needs, we can see that all three are strongly satisfied by the game. This might be the reason for its great success. Competence is achieved by acquiring new skills, being optimally challenged and receiving constant feedback. The

interaction with other people and the relationships with friends and enemies in the game gives a high sense of relatedness. The need for autonomy is thoroughly satisfied too, because players free choice to act and behave.

PLAYERS' MOTIVES IN GAMING

To analyze the different needs of player's motives in video games, Yee (2006) conducted an empirical study with a factor analysis in MMORPG. In this research the author identified 3 factors that motivate people to play video games:

- ***Achievement***: The desire to gain more power and status, progress rapidly and challenge against one another.
- ***Social***: Having an interest in helping friends, find and give support, collaborate and achieve tasks in groups.
- ***Immersion***: Discover and explore new things, creating a personal artificial character and escaping real life.

The author concludes from the research that "people choose to play games for very different reasons, and thus, the same video game may have very different meanings or consequences for different players".

PLAYER TYPES

A well-known player taxonomy of gamers was proposed by Bartle (1996). The Bartle test allows to categorize gamers based on their psychographic characteristics. The model is divided into four categories. The X axis represents preference for interacting with other players vs. exploring the world and the Y axis represents preference for interaction vs. unilateral action.

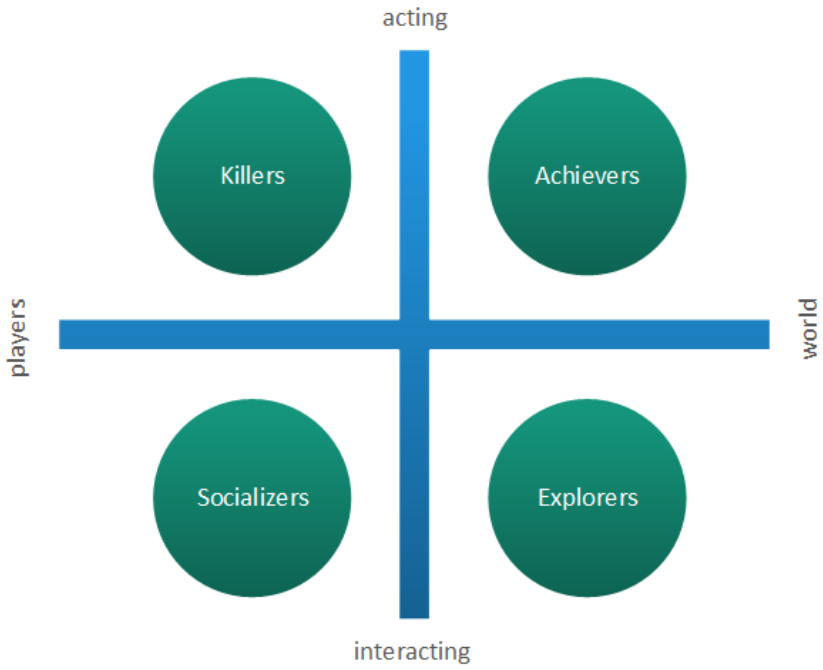

FIGURE 2-21. BARTLE'S PLAYER TYPES MODEL

Killers prefer competition with other players and like revenge. *Achievers* prefer to gain points, levels, gear and other collectable items. *Socializers* play games for the social aspect, rather than the actual

game itself. *Explorers* are players who prefer discovering areas, creating maps and learning about hidden places. The different types of gamers offers useful insight into how people will react in different situations, and also provides a starting point for developing a game or a gamified system (Bartle, 1996).

INTRINSIC AND EXTRINSIC MOTIVATION

SDT distinguishes intrinsic and extrinsic motivation. (Ryan & Deci, 2000). Intrinsic motivation describes the actual activity itself that we enjoy. Intrinsic motivation originates on the inside of an individual and is driven by internal rewards, like enjoyment, positive feelings and happiness. It describes the actual activity itself that we enjoy (Ryan & Deci, 1985). Coon & Mitterer (2008) describe intrinsic motivation when people simply enjoy an activity or see it as an opportunity to explore, learn, and actualize their potentials without any external rewards.

An example of intrinsic motivation is when people go running. They don't get any external rewards for their achievements (except for professional runners). They simply perform the exercise because of health benefits or other internal drivers (Greist et al., 1979).

Extrinsic motivation on the other hand, is what describes the external to keep us motivated (Ryan & Deci, 1985). An external rewards is something coming from the outside of an individual. It describes the effect that we experience when we get rewarded for the fulfillment of a certain activity. For instance, being rewarded with money, bonuses, praise or other tangibles for completing a task. For extrinsically-motivated people, it is not the action or task they like, but rather the outcome.

An example of motivation coming from the exterior is the Nike+ app ("Nike+," 2014). The app allows to collect information about runner's performance. Players can then check their progression and scores and compare themselves with others on a leaderboard. Points, badges and other information being collected are some examples of external rewards.

At its core, the implementation of gamification depends mainly on extrinsic incentives, which is materialized through its points, badges and leaderboards (Zichermann & Cunningham, 2011). Psychological researchers have reached a general consensus that extrinsic motivation can lead to weak and positive short-term effects, and potentially detrimental effects in the long-term (Benabou & Tirole, 2003; E. Deci, 1975). With gamification it is harder to affect intrinsic motivation compared to games that have a holistic game play (Chou, 2015). Great efforts and customization are required to make a game intrinsically motivating.

ACTIVITY CYCLES

To motivate the audience with gamification, a study on reward structures and feedback loops is essential. Based on the theory of reinforcement (Ferster & Skinner, 1957; Pavlov & Gantt, 1928) and the theory of flow (Csikszentmihalyi, 1991), Werbach & Hunter (2012) proposed to implement two types of cycles to increase motivation in gamified solutions. These are called *engagement loops* and *progression stairs*.

An ***engagement loop*** is the constant process of giving the user some reason to be motivated (e.g. a challenge). When this motivation is strong enough, the user will take action (e.g. solve a quest). After accomplishing the task the user receives feedback (e.g. points, badges). When this feedback becomes a motivator, the cycle is closed which in turn will lead to more actions. There should be a reward for every successfully completed action. Even an unexcpected positive message on the screen can create

a positive-reinforcement loop. The model shown in Figure 2-22 has powerful implications for any kind of gamified businesses (Zichermann & Cunningham, 2011).

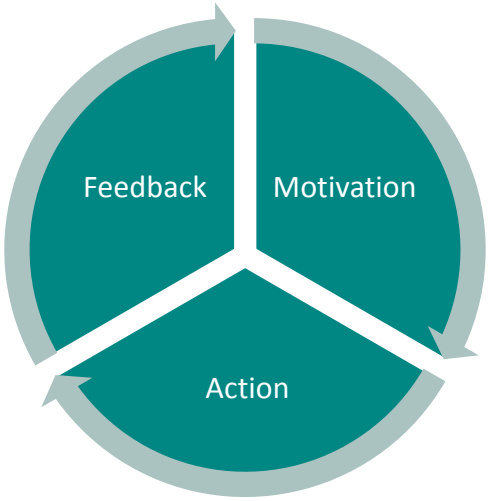

FIGURE 2-22. ENGAGEMENT LOOP

The ***progression stair*** can be seen as the representation of the player journey from novice to master. The concept is based on the theory of flow (Csikszentmihalyi, 1991). The idea is to keep the level of difficulty in alignment with the player's skills. If the challenge is too difficult, the player will feel anxious and lose interest in the game. On the other side, if the game is too easy, the player will be bored and stop playing the game (Chen, 2007). An example of the progression loop is presented in Figure 2-23.

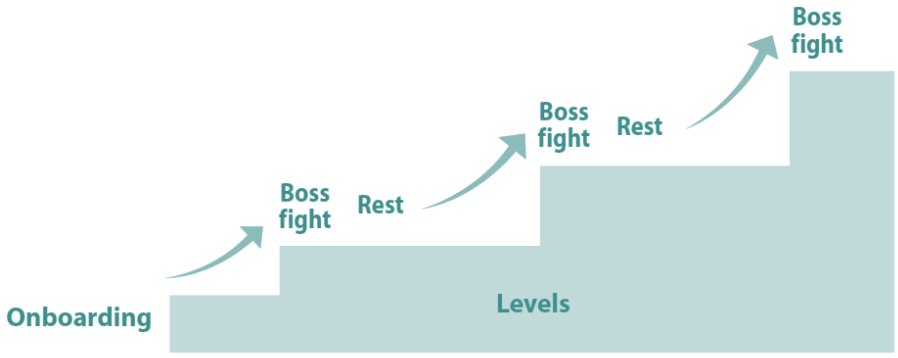

FIGURE 2-23. PROGRESSION STAIR (WERBACH & HUNTER, 2012)

2.4.5 ENGAGEMENT

The aim of gamification is not only to motivate people, but also to engage them (Zichermann & Cunningham, 2011). However, one of the first challenges presented by the literature is the lack of a universal definition of engagement. Existing literature and empirical evidence on engagement is limited because its novelty in the business and academic world (Macey & Schneider, 2008). For this research two major concepts will be compared, which are employee engagement and user engagement. Both are relevant when requirements must be gathered with an online digital platform.

EMPLOYEE ENGAGEMENT

The concept of employee engagement has diverse theoretical roots. Kahn (1990) first used the term personal engagement as being mentally, emotionally and physically invested in one's work. Later, Schaufeli, Salanova, González-Romá, & Bakker (2002) reviewed definitions of engagement in the business context and in academia. The authors define engagement as "a positive, fulfilling, work-related state of mind that is characterized by vigor, dedication, and absorption".

- ***Vigor*** refers to high levels of energy and resilience, the willingness to invest effort in one's job, not being easily fatigued, and persistence in the face of difficulties.
- ***Dedication*** refers to a strong involvement in one's work, accompanied by feelings of enthusiasm and significance, and by a sense of pride and inspiration.
- ***Absorption*** refers to a pleasant state of total immersion in one's work which is characterized by time passing quickly and being unable to detaching oneself from the job.

The concept of engagement gets more complex when related to motivation. While some studies claim that employee engagement is a "high internal motivational state" (Colbert, Mount, Harter, Witt, & Barrick, 2004), other studies suggest that engagement is "the illusive force that motivates employees to higher (or lower) levels of performance" (Wellins & Concelman, 2005). Macey & Schneider (2008) defined employee engagement as an umbrella term for a set of various concepts of psychological states, traits, and behavioral actions. Engagement has also been deemed as "a subset of flow", "flow in a more passive state" and "flow without user control" (Webster & Ahuja, 2006).

Nonetheless, the difference definitions, one thing on which most experts agree on is that engagement is beneficial for performance. Practitioners and researchers have found a strong and significant relationship between employee engagement and organizational outcomes, such as employee retention, productivity, profitability, customer loyalty and safety (Attridge, 2009; Harter, Schmidt, & Keyes, 2003; Kompaso & Sridevi, 2010; Robinson, Perryman, & Hayday, 2004).

USER ENGAGEMENT

In information science there is also a great ambiguity with the term engagement (Chapman, 1997; Laurel, 2013; Quesenbery, 2003). While employee engagement addresses the study of feelings held by employees about their jobs and organization (Macey & Schneider, 2008), user engagement covers the study of people's experience with technology (O'Brien & Toms, 2008).

An early definition of user engagement in information science was given by Leon (1991), which is composed of *user involvement* and *user participation* (Barki & Hartwick, 1989). User involvement is defined as a need-based mental psychological state (attitude) of users. User participation is an observable behavior of users' activities in system development and implementation.

A more recent definitions was proposed by Attfield, Kazai, Lalmas, & Piwowarski (2011). According to the researchers user engagement is "the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource".

USER ENGAGEMENT CHARACTERISTICS

The term user engagement is an abstract construct and closely related to the theories of flow (Csikszentmihalyi, 1991), aesthetic (Beardsley, 1982) and play (Stephenson, 1964). The following list contains characteristics of user engagement summarized from literature by Attfield et al. (2011):

- ***Focused attention***: Users must be focused in order to be engaged (O'Brien & Toms, 2008; Webster & Ho, 1997).
- ***Positive affect***: Emotions experienced by users are motivating (O'Brien & Toms, 2008).
- ***Aesthetics***: Sensory, visual appeal of user interface stimulates and promotes engagement (Jacques, 1995; O'Brien & Toms, 2008).
- ***Endurability***: People remember enjoyable, useful, engaging experiences and want to repeat them (O'Brien & Toms, 2008; Read, MacFarlane, & Casey, 2002).
- ***Novelty***: Unexpected surprise, unfamiliarity and the unexpected appeal to user's curiosity (O'Brien & Toms, 2008; Webster & Ho, 1997).
- ***Richness and control***: Richness captures the growth potential of an activity. Control captures the extent to which a person is able to achieve this growth potential (Jacques, 1995; Webster & Ho, 1997).
- ***Reputation, trust and expectation***: An implicit contract among people and entities which is more than technological (Attfield et al., 2011).
- ***User context***: User's motivation, incentives and benefits that effect the experience and engage users (Jacques, 1995; O'Brien & Toms, 2008).

USER ENGAGEMENT METRICS

Forrester research identified a qualitative and quantitative framework to measure engagement (Haven, Vittal, Spivey, Favier, & Cokeh, 2008). The framework includes four metrics, also known as the four I's:

- ***Involvement***: Presence of a user; Measured by e.g. the number of visitors, time spent, etc.
- ***Interaction***: Action of a user; Measured by e.g. click-through-rate (CTR), online transactions, videos played, comments written, etc.
- ***Intimacy***: Affection or aversion of a user; Measured by e.g. satisfaction rating, sentiment analysis in user-generated content, comments, surveys, etc.
- ***Influence***: The likelihood a user advocates on behalf of the brand; Measured by forwarded and sharing content, invitation to joint, etc.

Attfield et al. (2011) suggest three types of methods to measure engagement. These are:

- ***Self-reported engagement***: Is measured with questionnaires, interviews, reports, think-aloud and other similar methods that allow to get subjective answers. E.g. measuring distorted perception of time.
- ***Cognitive engagement***: This is an objective approach and used to measures task-based, neurological and physiological interactions. E.g. measuring eye tracking, heart rate or blood pressure.
- ***Interaction engagement***: Interaction engagement is also an objective way to measure engagement based on metrics and models. E.g. measuring number of returning users, number of clicks, and number of returning visitors.

To measure interaction engagement Lehmann et al. (2012) defined three types of metrics:

- ***Popularity***: How much a site is used, e.g. total number of users
- ***Activity***: How a site is used, e.g. average number of clicks per visit among all users
- ***Loyalty***: How often users return to a site, e.g. return rate frequency

CONCLUSION

Since a gamified program is most often a technological solution to encounter a human-related problem (Zichermann & Cunningham, 2011), both concepts, employee engagement and user engagement, are important to understand and leverage the organizational performance. Both have very similar characteristics. While the former defines the relationship between an employee and his work environment, the latter describes the relationship between the system user and the technological resource.

This thesis will merely rely on the concept of user performance (O'Brien & Toms, 2008) with the definition provided by Attfield et al. (2011). The measurement of people's relatedness to the designed artifact is more manageable and informative within the scope of this research. Based on the definitions and characteristics of user engagement, the concept of motivation is part user engagement (O'Brien & Toms, 2008).

2.4.6 DESIGN PROCESS

A popular and often referenced gamification implementation process are the six steps mentioned by Werbach & Hunter (2012). The following figure illustrates the complete gamification implementation process.

FIGURE 2-24. GAMIFICATION PROCESS

The way to approach a gamification program is to start with an exhaustive description of the business objectives (*Define*). The next step attends to figure out what target behavior should be promoted by the system (*Delineate*). Thereby, metrics should be defined allowing to measure the success once implemented. Afterwards, the players that will interact with the system must be described (*Describe*). The player model from Bartle (1996) is a good starting point to describe them. Next, repetitive and recursive structures must be identified (Devise). These were earlier introduced and are known as engagement loops and progressive loops. The fifth step of the process is about making the game engaging (Don't forget the fun). However, the question arising here is what makes a game fun? Hunicke, LeBlanc, & Zubek (2004) suggested to use a taxonomy, which are called "the 8 kinds of fun". The following factors can arouse a sense of pleasure when humans play games:

- ***Sensation***: Game as sense-pleasure
- ***Fantasy***: Game as make-believe
- ***Narrative***: Game as unfolding story
- ***Challenge***: Game as obstacle course
- ***Fellowship***: Game as social framework
- ***Discovery***: Game as uncharted territory
- ***Expression***: Game as self-discovery
- ***Submission***: Game as mindless pastime

Finally, the solution is implemented and the success is measured according to the previous defined metrics (Deploy). Kumar (2013) suggested to using brainstorming to identify metrics.

2.4.7 GAMIFICATION IN RE

INNOVATION MANAGEMENT

The requirements engineering phase of a software project requires a great amount of creativity and innovation by the stakeholders to find new solutions or improve existing ideas (D. Q. Nguyen, 1998; Snijders et al., 2014). One way to improve creativity is with an innovation management tool (Trott, 2008). Many tools exist to motivate players to share their creative ideas and allow for collaboration. For instance, *Be-novative* is a solution where players can read and evaluate ideas and track the implementation progress ("Be-novative," 2015). The main goal of is to inspire creative and anonymous idea-generation within large enterprises. Similar tools are *SpigitEngage* and *Venture Spirit* that also motivate employees to spread their ideas ("SpigitEngage," 2015, "Venture Spirit," 2015).

iTHINK

A gamified requirements engineering tool with a more scientific value is known as *iThink* and was developed by Fernandes et al. (2012). iThink is a gamified approach to improve the collaboration and participation in the phase of requirements definition. The authors analyzed several requirement elicitation practices and then developed a system with appropriate game elements. The tool is based on the theory of the six thinking hats (De Bono, 1999). The idea behind this method is to stimulate parallel thinking in teams in order to be more effective and avoid team conflicts. Each hat has a different style of thinking and is represented by a unique color. This approach can be used to discover new requirements and discus existing ones.

The main game elements used for iThink are points and progress bars. Points can be acquired through different user actions, which are mapped to one of the hats. For instance, a user can win points by submitting a new requirement, rating a requirement, commenting on a requirement and completing a discussion of a requirement. In Figure 2-25 an example screen can be seen, in which the user is asked to reflect on a requirement in 4 different ways, each awarding a certain amount of points. The points are then reflected on a progress bar. A prototype of this tool was built and tested in two case studies. The results demonstrated that collaboration was improved and people enjoyed interacting with the system. Overall, project managers were very satisfied with the generated requirements.

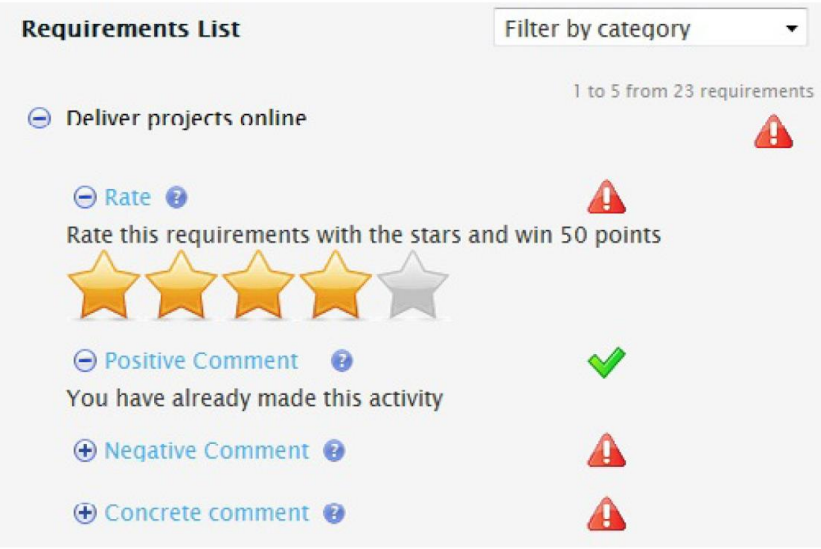

FIGURE 2-25. ITHINK SCREENSHOT (FERNANDES ET AL., 2012)

This research is an examples how game elements can be applied to a non-gaming context, such as requirements engineering. Nonetheless, there were only two types of game elements used to increase user participation and motivation. The game still leaves much potential to further gamify requirements elicitation and analysis.

RE*fine*

An elicitation tool, called RE*fine*, for crowd-centric RE was developed at Utrecht University to open participation to all types of stakeholders and incentivize them with gamification in the long-term (Snijders et al., 2015). The platform consists of six game elements that should socially engage stakeholders. These are roles, resources & points, leaderboards, group forming, exploration and endorsement. The tool was tested in a case study at KPMG with 19 stakeholders. The evaluation of the results was then performed with RE*fine* users and product management experts. Users felt motivated and were more engaged compared to previous experiences. Product management experts agreed that there were only little incentives for stakeholders to return to the platform. A screenshot of the RE*fine* tool is shown in the image below.

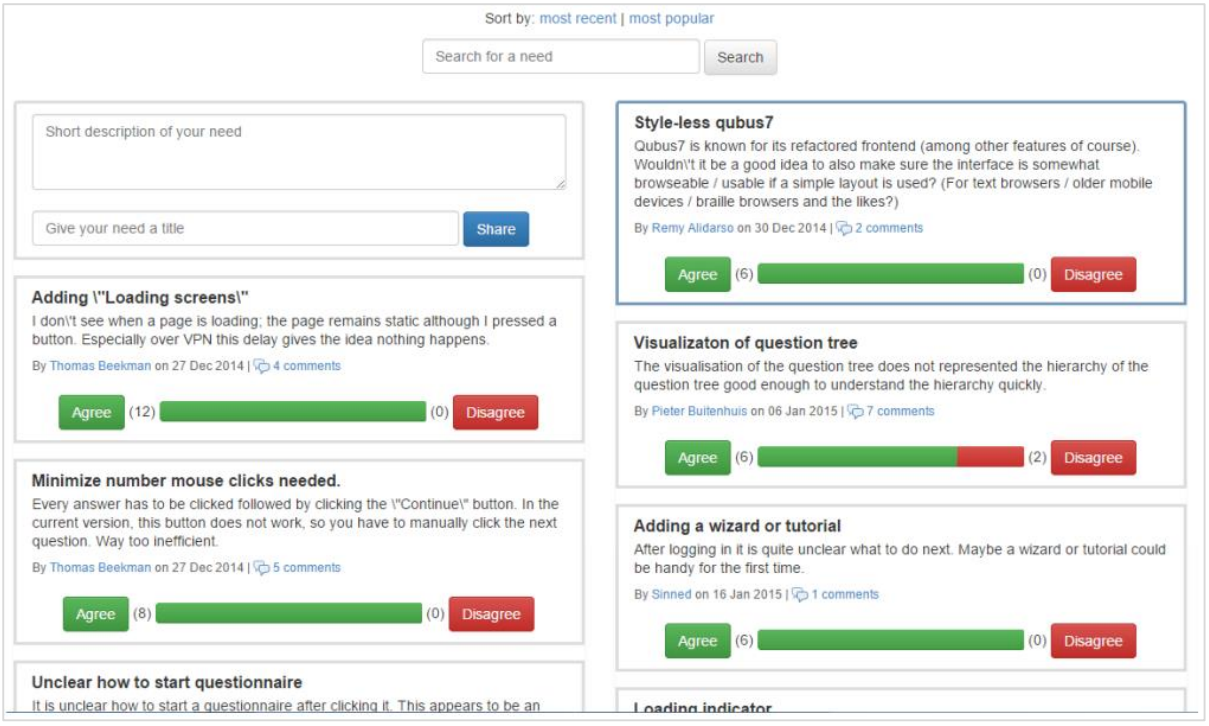

FIGURE 2-26. REFINE OVERVIEW (SNIJDERS ET AL., 2015)

PLANNING POKER

A common method for group estimation of user stories in agile software projects is planning poker, also called Scrum poker (Grenning, 2002). The goal behind this method is to make the estimation task feel a lot more like a poker game and therefore, achieve better group consensus. In this “game” Every group member selects a card with an estimation number and places it face down on the table. The cards are then simultaneously flipped. The highest and lowest bidder discuss the estimation gap. This task is repeated until a consensus is reached.

VIRTUAL WORLDS

Russell & Creighton (2011) proposed to establish a virtual world with a three-dimensional graphical user interface to intrinsically motivate stakeholders when requirements are derived and evaluated. A virtual world is “a synchronous, persistent network of people, represented as avatars, facilitated by networked computers” (Bell, 2008). Having a digital platform with authentic physics, voice chat and character ownership might also facilitate requirements engineering, since it has proven success in marketing, meetings and education (Boulos, Hetherington, & Wheeler, 2007; Lui, Piccoli, & Ives, 2007). Second Life is an example of a popular 3D virtual world in which users can select an avatar and connect with thousands of people simultaneously over the internet (“Second Life,” 2015).

From an empirical study on virtual worlds and online collaboration, Nicholson, Nicholson, Coyle, Hardin, & Krishen (2014) claim that groups using a simulated environment for idea generation discovered significantly more unique ideas compared to a second group only using an environment with a chat feature. Furthermore, the former group experienced a significant higher level of enjoyment compared to the second one.

GAMESTORMING

In 2010 Gray, Brown, & Macanufo wrote a book called *gamestorming*, where brainstorming techniques are turned into games. The authors described more than 80 games to facilitate creative idea generation, starting from many of the standards games such as affinity diagrams implemented with sticky notes, to more imaginary exercises such as forced rankings and storyboarding. The three core principles (as shown in Figure 2-27) behind each game are always the same. Each games has an opening, an exploring, and finally a closing stage.

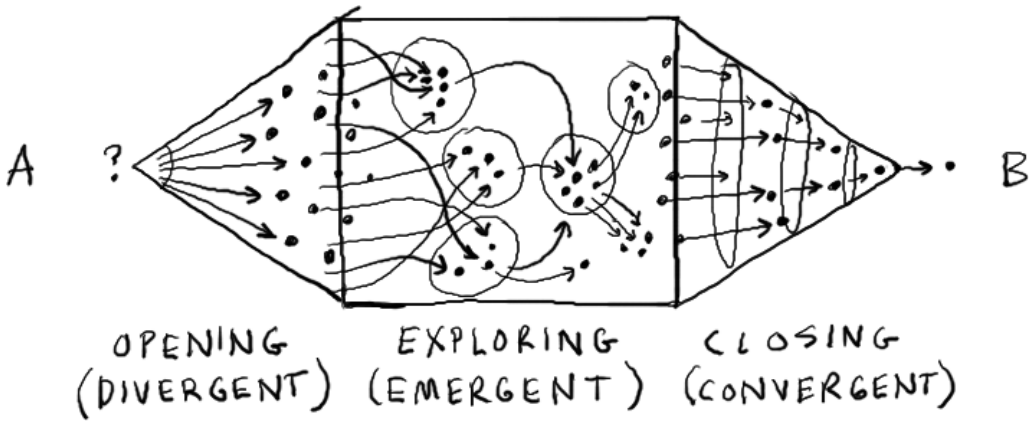

FIGURE 2-27. GAMESTORMING CORE PRINCIPLES (GRAY ET AL., 2010)

Pisarek (2011), the founder of Dynamic Owl Consulting and a Microsoft SharePoint MVP (Most Valuable Player), successfully uses the Cover Story game for his requirements engineering workshops for SharePoint. He uses the game because it not only encourages people to think outside the box, but also allows to detect requirements that perhaps weren’t discovered before the game was played.

3 CONCEPTUAL FRAMEWORK

After thoroughly reviewing the literature, only two studies examining the relationship between gamification and requirements engineering were found (Fernandes et al., 2012; Snijders et al., 2015). The artifact developed for *iThink* was evaluated within two case studies to measure the influence of gamification on the quality of requirements and stakeholder participation. Both case studies involved potential system users for a requirements elicitation workshop. Afterwards, a questionnaire was distributed to measure participants' motivation. Project managers received a questionnaire to rate the requirements' quality. In conclusion, the tool helped to improve collaborative requirements elicitation and enhanced stakeholder participation rate. RE*fine* was developed to increase crowd-centric RE participation and evaluated in a case study at KPMG (Snijders et al., 2015). Stakeholders were more motivated and engaged in the elicitation process compared to previous experiences.

Despite the novelty of these two researches, there are still some limitations: The researchers only evaluated their tool in the context of a case study, making it difficult to generalize the results (Wohlin et al., 2012). Another limitation was related to the lean questionnaire in the iThink project. Five questions were asked to measure motivation and three simple questions to measure quality. Furthermore, the impact of alternative causes, such as usability, design and participant's background were omitted. These last two points are considered to be a threat to internal validity (Bhattacherjee, 2012).

Other studies measuring the direct relationships between gamification and requirements engineering activities were not encountered in the literature study. Nonetheless, previous studies investigated the relationships between gamification and engagement, as well as between engagement and performance.

In Section 3.1 we first have a closer look at the relationships between the different concepts, followed by a description of the conceptual model in Section 3.2. Finally, we introduce our two hypotheses that are measured in an experiment in Section 3.3.

3.1 RELATIONSHIPS

3.1.1 GAMIFICATION AND USER ENGAGEMENT

Considering that future employees will have grown up in a world where video games are common (ESA, 2014), using gamification as a mean to motivate and engage people seems reasonable. A very often cited example of a successful gamified business solution is the *Language Quality Game* (LQG) from Microsoft. The company developed a web based platform to enable and engage native speaking citizens to help assess and improve the linguistic quality of Windows 7. A company-wide policy together with management support allowed employees worldwide to play the game by reviewing translations of the Windows dialog boxes. The game rewarded users with points and levels and ranked them on a leaderboard. The scoring system compared individual players, but also aggregated scores on country level. This game setting created a competitive and dynamic environment, followed by an increase number of users. More than 4,000 employees reviewed over 500,000 Windows 7 dialog boxes in one month. Approximately 7,000 defects across 36 different languages were reported (J. Williams & Ross, 2010).

Jennifer Thom, a research scientists on Collaborative User Experience at IBM, examined the impact of the removal of extrinsic rewarding elements, such as points and badges that were installed on a social networking system (SNS) (Thom, Millen, & DiMicco, 2012). A large enterprise that was already using a gamified SNS was selected. Employees received points when performing certain actions on the social network, and were then ranked on a leaderboard. In an experiment performed over 6 months, Thom et al. (2012) discovered that the absence of all game features had a significant negative impact on the content contribution and user participation. For instance, the amount of profile comments was significantly reduced by more than half. Other actions, such as uploading photos, addition of lists and commenting on these two actions was nearly halved.

Mekler, Brühlmann, Opwis, & Tuch (2013a) reported similar results. The researchers examined the effect of points on players' motivation. Participants had to annotate pictures with tags and were then rewarded with a certain amount of points. The findings of their study concluded that rewarding points increased tagging. Other studies concluded that that feedback mechanisms (engagement loops) and challenges (progression stairs) increased group collaboration (Jung, Schneider, & Valacich, 2010).

Overall, the current empirical research on gamification largely supports the popular view that gamification has the ability to influence people's behavior. Hamari, Koivisto, & Sarsa (2014) performed a literature review based on 24 empirical gamification studies. The authors concluded that the majority of the results had positive psychological and behavioral outcomes. However, they pointed out that the success of gamification also depends on other factors. For instance, the context in which the gamification is being implemented or the users interacting with the system. Furthermore, the research indicates that the effect of gamification is mainly short-term, and could be caused by its novelty.

3.1.2 USER ENGAGEMENT AND PERFORMANCE

The effect of engaged users on performance in requirements engineering is twofold. On the one hand, increased stakeholder engagement in requirements engineering leads to higher user satisfaction and greater system acceptance (Harris & Weistroffer, 2008). On the other hand, engaged and motivated people generally perform better and with higher quality.

Studies measuring the relationship between engagement and performance in requirements engineering were only encountered in the *iThink* (Fernandes et al., 2012) and RE*fine* research projects (Snijders et al., 2015). However, various literature and studies between employee engagement and job performance do exist. At a general level, engaged employees not only better accomplish their goals, but they are also physically, cognitively and emotionally closer connected to their endeavors (Kahn, 1990).

Results from research organizations demonstrated that there may be a strong link between engagement, performance and business outcomes (Harter, Schmidt, & Hayes, 2002). Gallup's research group developed a survey with 12 actionable workplace elements that were linked to employee engagement. A meta-analysis was used to examine 192 business organizations with nearly 40,000 business units. The results indicated a strong correlation between employee engagement and organizational performance outcomes (Harter, Schmidt, Agrawal, & Plowman, 2013).

Furthermore, Gallup research group has shown that employee engagement are strongly related to innovation (Krueger & Killham, 2007). Based on telephone interviews with more than 1,000 people, 66% who responded to be engaged at work also indicated that their current job brings out their most

creative ideas. Only 3% of disengaged employees strongly agreed that their job brings out their most creative ideas.

Flow can also be seen as a motivating force for excellence as recommended by literature (Nakamura & Csikszentmihalyi, 2002). Individuals experiencing flow are more motivated to carry out further activities. To adapt this theory to requirements engineering, we assume that motivated and engaged stakeholders with the right level of flow are more likely to deliver creative requirements and with higher quality. Consequently, engaged stakeholders outperform disengaged ones.

3.2 CONCEPTUAL MODEL

The conceptual framework is constructed with the theories derived from gamification, engagement and performance. The links between these three factors are depicted in a conceptual model in Figure 3-1, which serves to measure our hypotheses. The model recognizes three main variables with two relationships. Furthermore, two control variables are added to avoid threats to internal validity. These two control variables are stakeholder expertise and motivation.

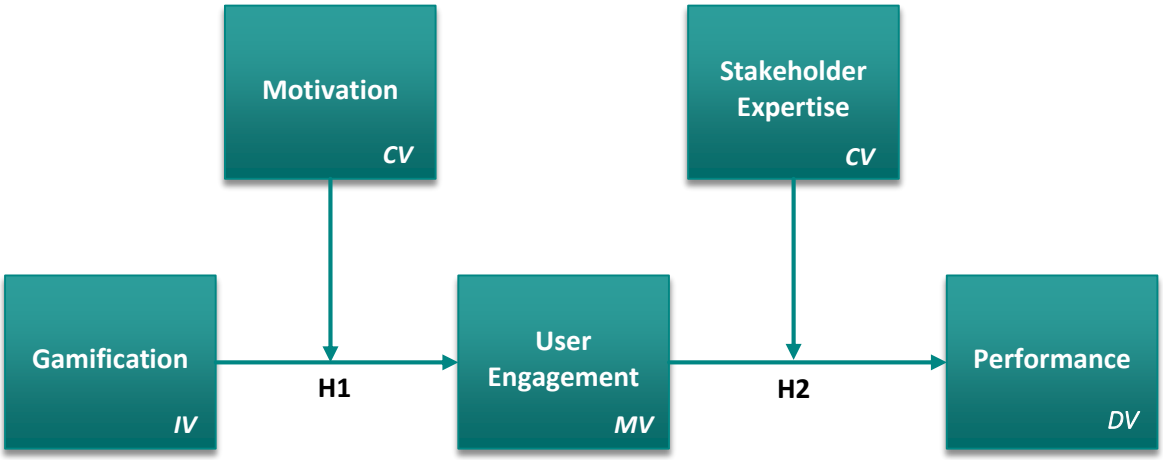

FIGURE 3-1. CONCEPTUAL MODEL

Among these variables, “Gamification” represents the independent variable (IV). We examine the effect of different game elements on “User Engagement”, which in turn serves as a mediating variable (MV). As “User Engagement” is suggested to have a positive effect on performance, there is a dependent variable (DV) described as “Performance”.

Furthermore, “Stakeholder Expertise” and “Motivation” are introduced as control variables (CV) to reduce noise and therefore, improve internal validity. Literature suggested that work experience can facilitate job performance (Quiñones, Ford, & Teachout, 1995). Therefore, it was important to consider this variable to ensure that the measured effect is caused by gamification.

3.2.1 CONCEPTS

GAMIFICATION

Among the multiple definitions in the literature, we decided to use following definition for gamification: “the use of game mechanics and experience design to digitally engage and motivate people to achieve their goals” (Burke, 2014). This definition is more specific and not only describes the intention behind gamification, but also includes the behavioral aspect.

This concept is measured with a dichotomous variable by dividing a sample into two groups. In an experimental setting participants are separated into a treatment group and a control group based on their personal motives and work experience. Both groups then receive a digital online platform with the same basic features that enable elicitation of user stories.

To ensure that the effect is originally caused by gamification, both platforms must have the same features and same user interface. Gamification is then used as a treatment to increase stakeholder engagement for the one group.

From psychological researchers we know that extrinsic motivation may work in the short-term, but like other bribes they will lose their effect over time (Benabou & Tirole, 2003; E. Deci, 1975). To mitigate this risk, this experiment uses a combined set of game elements to also incentivizes intrinsic motivation (E. L. Deci & Ryan, 1980). With respect to the platform's capabilities and researcher's programming skills, the following game components were selected and applied to the gamified artifact design:

- ***Points*** are the basis to reward users for their activities on the online platform.
- ***Badges*** are visualizations of achievements and should give a certain surprise effect.
- ***Leaderboard*** is a ranking that shows all signed up players in an order from top-down.
- ***Levels*** are the difficulty phase in a game to enable linear progression.
- ***Challenges*** are journeys towards a goal, which are rewarded with badges and extra points. The purpose of this element is to affect cognition and keep flow on an optimum level.
- ***Activity feed*** shows a stream of all the recent actions in the community.
- ***Avatar*** is a graphical representation of the player that can be individually selected.
- ***Onboarding*** is the process of getting familiar with the platform.
- ***Game master*** is responsible to moderate the game, has the opportunity to interact with the players via this game element to answer questions and influence their behavior.
- ***Storytelling*** is the background narrative of the onboarding program. The purpose with a story is to arouse positive emotions.
- ***Video*** is an element to display visual and audio media about user stories and the business case.
- ***Facial animation*** is used for motion and audio-driven characters that represent the personas.
- ***Progress bar*** allows to show a current state of the player in a process.
- ***Quiz*** is an element to let players check their knowledge after completing the onboarding program.
- ***Timer*** is a clock that shows the remaining game time and should generate some sense of pressure.
- ***Liking*** is a feature where the user can express what they like, enjoy or support certain content.
- ***Prize*** is the reward that the winner will receive after the game is finished.

The reason for this broad selection is to have a large portfolio of game elements, each satisfying different intrinsic needs. Since not all stakeholders react equally to a particular stimulation, the objective is to implement various game elements and mechanics that were often encountered in literature in order to increase user engagement (Herger, 2014; Werbach & Hunter, 2012; Zichermann & Cunningham, 2011). However, this broad selection bears the risk of losing control of the effect that individual game elements or subsets of game elements have on human behavior.

Since collaboration is an important success factor in RE (Kappelman et al., 2006), both platforms are equipped with a chat feature where users can communicate and interact with each other. Omitting this feature for the one platform could have serious threats to internal validity. However, the gamified platform includes a game master who is responsible to answer questions and moderate the game (Tychsen, Hitchens, Brolund, & Kavakli, 2005). He is able to interact with players via chat and motivate them with verbal feedback when necessary.

MOTIVATION

Motivation is the "psychological force that moves a person to act, to meet a need or achieve a goal" (Stedman, 2011). Literally, motivation is the desire to do things and is a crucial element in setting and attaining goals. In the context of this experiment motivation is important to know stakeholders desires and therefore, select appropriate game elements to influence their behavior. For instance, some stakeholders might have a higher demand for power and status in their community, while others care less. In this case, it is assumed that a leaderboard can motivate people with a high demand for power and status.

A possible approach to quantify motivation is with the aid of the Reiss profile test (Reiss, 2004). The Reiss profile test is a standardized assessment instrument often used by Human Resources for talent recruiting and coaching. The 16 identified human desires were solidly explored with a factor analysis and have been proven to be statistically reliable with high validity (Havercamp & Steven Reiss, 2003). Due to the fact that all employees working at MaibornWolff had to fill out the Reiss profile, it is a logical consequence to use Steven Reiss' theory for this research.

IMPACT OF GAMIFICATION ON MOTIVATION

It is important to know to which degree the selected game elements could affect stakeholder motivation. To answer this question an indirect mapping between the game elements and the 16 human desires was performed in two steps. First, the game elements were mapped to Jon Radoff's 42 fun factors (Radoff & Kidhardt, 2011) (see Appendix A1: Jon Radoff's 42 Fun Factors). Gamification communities have elaborated more on these 42 fun factors from a gamification perspective (Whatley, 2011). Next, the frequency of each desire in Jon Radoff's matrix was summed up for every game element that was present in the fun factor matrix (see Appendix A2: Mapping Game Elements to Fun Factors). For instance, "Badges" include both "Collecting" and "Finding Random Treasures", which have an impact on "Power". The number of crosses were then added up for "Power" and entered into a new sheet with "Badge" on the opposite side. Consequently, this task resulted in a two-dimensional matrix with a mapping between 17 game elements and 16 human desires (see Appendix A3: Mapping Game Elements to Motivation).

To improve visualization, the table was transformed into a bubble graph (see Figure 3-2). The size of the bubbles represent relatively how strong the human desires could have been affected by the game elements. The construction of this graph was performed rather exploratory and with room for improvement. Nonetheless, we still are able to conclude that the combination of these game elements most notably influence people with high desire for order, status, curiosity, power, tranquility and independence.

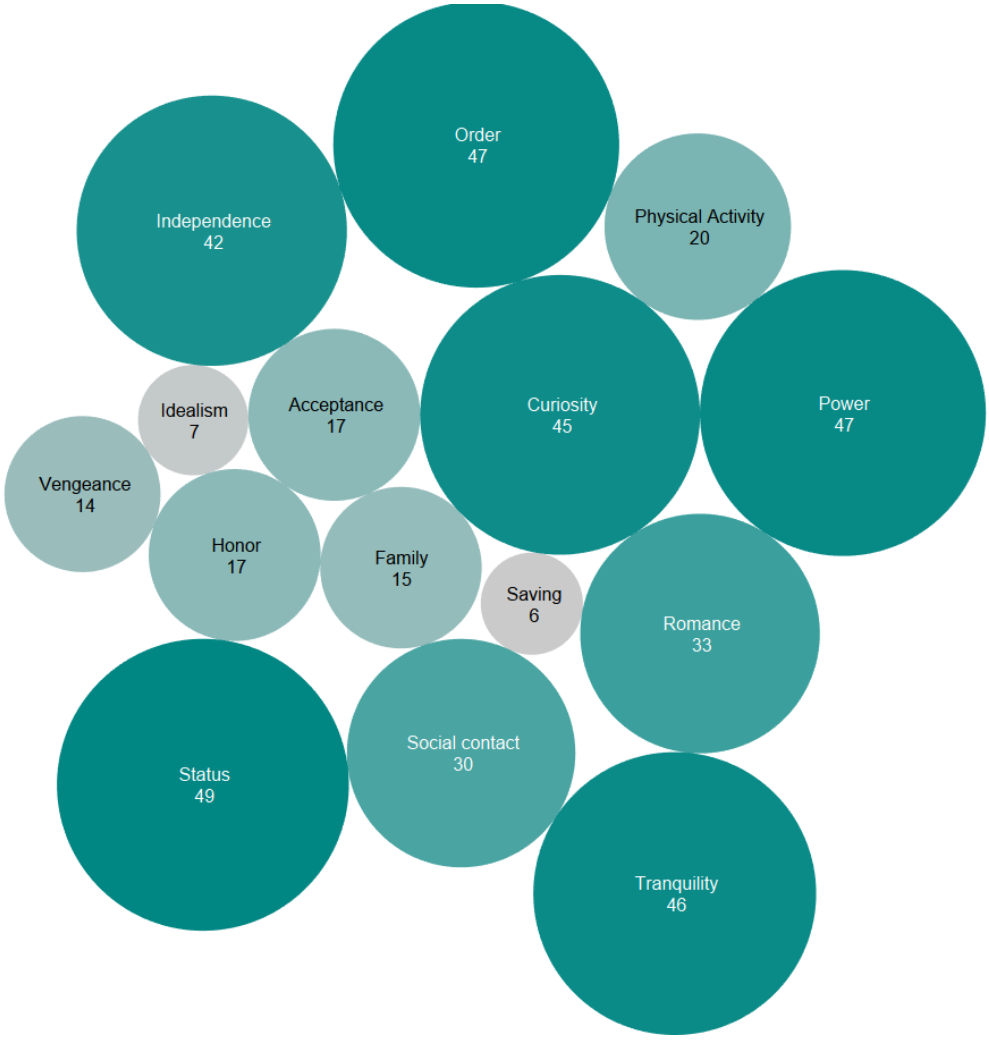

FIGURE 3-2. IMPACT SIZE OF GAMIFICATION ON MOTIVATION

USER ENGAGEMENT

User engagement is a very complex concept that is much debated in literature. In the context of this thesis we rely on the definition provided by Attfield et al. (2011) that resulted from an extensive research: "User engagement is the emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource". This multi-dimensional construct consists of several underlying concepts and is decomposed as shown in Figure 3-3.

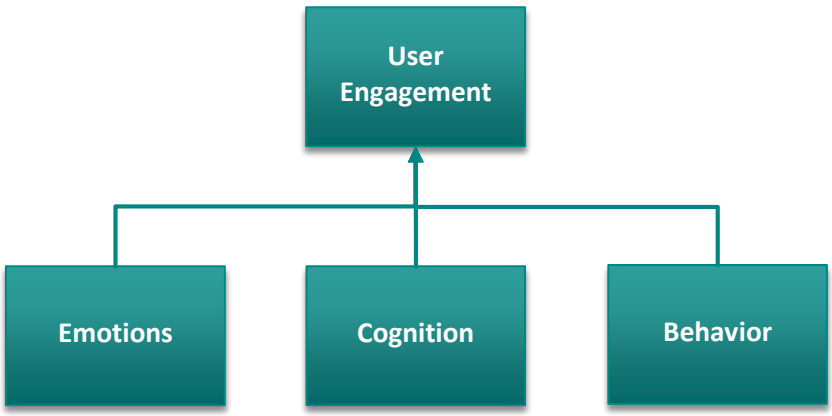

FIGURE 3-3. CONSTRUCT OF USER ENGAGEMENT

- ***Emotions*** are the user's feelings, e.g. happy, sad, excited, bored, etc. (Attfield et al., 2011)
- ***Cognition*** are the user's mental states, e.g. concentrated, challenged, lost, interested, etc. (Attfield et al., 2011)
- ***Behavior*** are the user's interactions, e.g. click, visit, comment, vote, etc. (Attfield et al., 2011)

Within the scope and setting of this research, user engagement is measured in two different ways as proposed by Attfield et al. (2011). First, self-reported engagement is used to collect qualitative and subjective information from participants by means of a survey. Questions are formulated to measure emotional and cognitive connections. Secondly, interaction engagement is used to measure user behavior on the RE platform.

Emotions can either have positive or negative affectivity and are both independent of each other (Watson, Clark, & Tellegen, 1988). Several scientific methods exist to measure self-reported human emotions. This can be done with a verbal or non-verbal survey. Examples of a verbal questionnaire include the *Positive and Negative Affect Schedule* (PANAS) (Watson et al., 1988), *Differential Emotions Scale* (DES) (Izard, 1993) or the *Affect Grid* (J. A. Russell, Weiss, & Mendelsohn, 1989). Examples to measure emotions with a non-verbal approach are the *Self-Assessment Manikin* (SAM) instrument (Lang, 1980) or the *PrEmo* tool developed by the Susa Group ("PrEmo Tool," 2015).

Past research studies developed different scales to measure cognition. In game design cognition is related to the theory of flow, for which several instruments exist. The *Flow Short Scale* consists of 10 items that allows to measure all components of the flow experience with a 7-point scale. The scale has been tested several times in different domains with a Cronbach α above 0.9 (Rheinberg, Vollmeyer, & Engeser, 2003).

Interaction engagement is used to quantitatively and objectively measure the behavioral connection that existed between the user and the digital platform. With analytical tools it is possible to keep track of users in the background when they interacted with the online platforms.

PERFORMANCE

Performance is an abstract, multidimensional construct that cannot be pointed to directly (Viswesvaran, 2001). A widely endorsed definition of performance is that of Viswesvaran & Ones (2000). The authors defined performance as "scalable actions, behavior and outcomes that employees engage in or bring about that are linked with and contribute to organizational goals".

To conceptualize and operationalize performance, the dimensions and indicators had to be identified. From a systematic literature review, Koopmans et al. (2011) identified 10 different sub-dimensions of performance. Three sub-concepts were perceived useful for this research and selected to effectively measure performance in requirements engineering. These are defined as follows:

- ***Productivity***: The average amount of requirements that a stakeholder produced in a given amount of time.
- ***Quality***: Quality is the degree to which a requirement complies with the needs of the stakeholders.
- ***Creativity:*** The phenomenon whereby something new and in some way valuable is created.

These three chosen dimensions enable the measurement of performance in RE and are visualized in below figure.

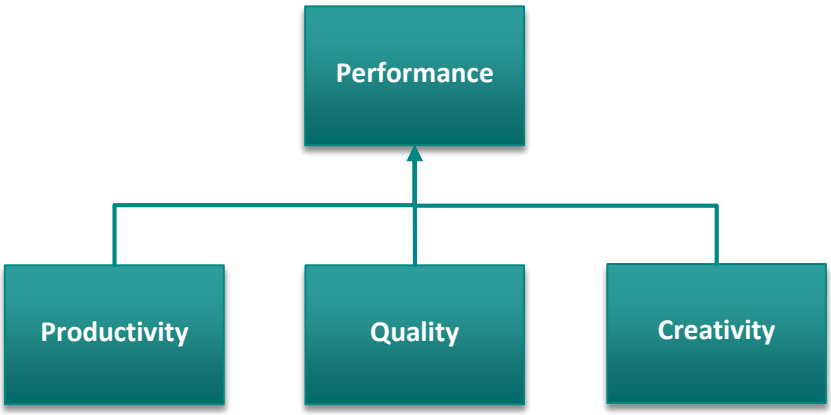

FIGURE 3-4. CONSTRUCT OF PERFORMANCE

One narrower concept of performance is productivity, which can be related to the number of produced requirements over time (Kemppilä & Lönnqvist, 2003). This is an objective indicator that is quantitatively measured. Examples for RE include the number submitted user stories and scenarios.

Another important indicator to measure performance is quality (Campbell, 2008). The quality of requirements can be analyzed by several objective and subjective means. The *INVEST* model is well known and used for IT projects at MaibornWolff (Cohn, 2004). A further quality measurement tool includes the *Kano* model (Kano et al., 1984).

The third sub-dimension is creativity (Fluegge, 2008). Creativity is an important success factor in requirements engineering that allows stakeholders to discover new requirements and develop innovative solutions (L. Nguyen & Shanks, 2009). Creative performance in organizations is defined by cognitive psychology as "the ability to produce work that is both novel (i.e. original, unexpected) and appropriate (i.e. useful, adaptive concerning task and constraints)" (Sternberg & Lubart, 1999).

STAKEHOLDER EXPERTISE

Stakeholders who are involved in the requirements engineering process have different backgrounds and experiences. Depending on their skills and capabilities they can affect the concept of performance and threaten internal validity. For instance, broad domain knowledge and extensive competences are two dimensions that can significantly lead to higher performance (Viswesvaran, 1993). According to Engeser & Rheinberg (2008) "flow is related to performance, but does not necessarily cause it. The correlation arises simply because expertise leads to more flow". To overcome this limitation, we have to control the differences in expertise.

To get an overview of participants involved in the experiment, the concept of stakeholder expertise was defined to collect demographical and work related information. Demographics allow to measure quantifiable characteristics of a given population (Andrew, 1998). Work experience was important to create two equal balanced groups. It is assumed that people with more work experience have the capability to provide better results.

3.3 HYPOTHESES

For the purposes of this research, the conceptual model can be interpreted as follow: The introduction of gamification in an online digital RE platform leads to higher user engagement. Consequently, higher engaged users in requirements elicitation perform better than disengaged ones.

From the links between gamification, user engagement, and performance, two hypotheses were derived. The objective of this study is to investigate these relationships and find evidence to reject the null hypothesis.

- ***H1*.** If a diversified gamification RE platform aligned with stakeholder motivation is deployed, then user engagement is significantly increased.
  - **H_0: $\mu 1 = \mu 2$**
    Stakeholders show no difference in user engagement when gamification is deployed to RE elicitation.
  - **H_1: $\mu 1 > \mu 2$**
    Stakeholders exhibit a higher rate of user engagement when gamification is deployed to RE elicitation.
- ***H2*.** If stakeholders are more engaged in requirements engineering with respect to their expertise, then the overall performance of the process and outcomes is significantly increased.
  - **H_0: $\mu 1 = \mu 2$**
    Stakeholders show no difference in performance when being more engaged in RE elicitation.
  - **H_1: $\mu 1 > \mu 2$**
    Stakeholders exhibit a higher rate of performance when being more engaged in RE elicitation.

Hypothesis 1 aims at proving that the relationship between gamification and user engagement has a significant impact. When game elements, such as points, badges, leaderboards, etc. are implemented for requirements elicitation workshops in order to influence stakeholders' behavior, then it is assumed that people are more likely to be involved in the gathering process. Overall, a diversified gamification platform allows to influence people's behavior and positively affect their emotions and cognition.

With Hypothesis 2 this research project tries to find evidence that when users are more engaged in requirements elicitation workshops, they are also more likely to perform better. Engaged users might not only come up with more and higher quality requirements, but their ideas could be of more novelty compared to disengaged users.

To test these two hypotheses, a digital prototype that enables online requirements elicitation with a digital platform is developed and evaluated in an experimental condition. The next chapter describes the design and construct of the artifact design.

4 ARTIFACT DESIGN

Design science research (DSR) involves the design of a novel or innovative artifact and the evaluation of the use of such an artifacts in information systems. According to Van Aken (2005) the main purpose of DSR is to achieve domain knowledge by building and deploying an artifact. DSR artifact can include models, methods, constructs and instantiations (von Alan et al., 2004). In this chapter we describe the design of the software artifact that was developed to measure the two hypothetical research statements from previous chapter.

To ensure that gamification has a significant impact on user engagement and performance, an online digital platform for requirements elicitation in form of user stories was built. A prerequisite for such an online system was to motivate stakeholders in interdisciplinary teams to express their requirements, irrespective of location and time. Face-to-face meetings are no longer the sole communication method to discuss problems and make decisions (Damian, 2007; Desanctis & Gallupe, 1987). An online requirements elicitation system enables stakeholders in different time zones to participate in the requirements engineering process.

A further benefit of having a centralized online solution is that all ideas are made explicitly available, allowing the system to turn into a knowledge base. The entered user stories can be archived and traced back to its originator. Furthermore, with this type of solution there are no media breaks. All requirements are immediately stored in a central database and don't need to be manually transcribed after an RE workshop. This can save the Requirements Engineer a great amount of time.

There are still some down-sides with such a tool. Having people express their requirements with an online system threatens collaboration. Research has shown, that knowledge and idea extension in teams is important for innovation and productivity (Paulus & Yang, 2000). Therefore, it was an essential requirement to provide functionalities so that stakeholders can communicate with each other.

To develop a design artifact, a set of prioritized requirements were first written in form of user stories from a researcher's perspective. This user story list is documented in Section 4.1. Next, a documentation of the software architecture is provided in Section **Error! Reference source not found.** to get a clear picture of the technical implementation. Finally, the graphical user interface with examples and screenshots of the system features is described in Section 4.3.

A second prototype had to be built as well to measure the effect of gamification. This prototype was built exactly equal, but with the absence of all game elements. The differences between the two prototypes are explicitly highlighted where necessary.

4.1 SYSTEM REQUIREMENTS

Since we are experimenting with user stories, we decided to use this technique to document requirements for our own research project. The following table contains a list of key requirements. These requirements were prioritized with the MoSCoW approach to determine their importance (Brennan, 2009). The prioritized user stories were then used as a task list for implementation. User stories with a prioritization of 4 were not implemented, but can be considered for future enhancements.

As a	I would like to	So that	Prio.	Game Version
researcher	allow stakeholders to add user stories	they can express their system goals	1	
researcher	allow stakeholders to add acceptance criteria (scenarios)	they can state the boundaries of a user story	1	
researcher	implement game elements	stakeholders become more engaged	1	✓
researcher	provide points	stakeholders are incentivized	1	✓
researcher	provide badges	stakeholders are positively reinforced	1	✓
researcher	provide a leaderboard	stakeholders can compare themselves	1	✓
researcher	provide levels	stakeholders can progress	1	✓
researcher	provide quests	stakeholders feel challenged	1	✓
researcher	provide an avatar	stakeholders have a visual representations of their character	1	✓
researcher	allow stakeholders to change their personal data	their profile is up to date	2	
researcher	provide an activity feed with recent actions	stakeholders stay informed about the activities in their community	2	✓
researcher	allow stakeholders to like user stories	features can be prioritized	2	✓
researcher	provide a chat feature	stakeholders can collaborate together	2	
researcher	provide a game master	this role can communicate with the players and provide verbal feedback	2	✓
researcher	provide progress bars	stakeholders feel like working towards a goal	2	✓
researcher	provide a tutorial	stakeholders get on boarded quickly and know how to play	2	
researcher	show the remaining game time	stakeholders feel some pressure	3	✓
researcher	provide a support function	stakeholders can get help	3	
researcher	use personas	stakeholders think from different perspectives	1	
researcher	use facial animations for personas	stakeholders feel more connection with the potential system users	3	✓
researcher	integrate an animation video to describe the business case	stakeholders get more familiar with the problem	3	✓
researcher	show top user stories	the most desired requirements are highlighted	2	
researcher	monitor users' behavior	I can measure user engagement	1	
researcher	provide a search function	stakeholders can find information more quickly	3	

researcher	provide a mandatory login function	unauthorized access can be prevented	1	
researcher	prevent stakeholders from cheating	the game stays fair for all	1	✓
researcher	provide a mobile version	stakeholders can play on their smartphones and tablets	3	
researcher	show user stories on a user story map	readability is increased	4	
researcher	have a bot that takes the role of the game master	the researcher can observe other activities in the background	4	✓
researcher	Use regular expression validation	stakeholders follow the syntax structure of user stories	4	
researcher	allow release planning	stakeholders can define the product backlog for an iteration	4	
researcher	allow versioning	I can keep track of incrementally different versions	4	
researcher	integrate a backlog system	Project manager can keep track of progress	4	

Priority key:

1 = ***Must have***: A requirement that must be satisfied in the final solution to be considered a success.
2 = ***Should have***: A high-priority item that should be included in the solution if it is possible.
3 = ***Could have***: A requirement which is considered desirable but not necessary.
4 = ***Won't have***: A requirement that will not be implemented, but may be considered for the future. (Brennan, 2009)

TABLE 4-1. USER STORIES FOR ARTIFACT DESIGN

4.2 ARCHITECTURAL DESIGN

The intention behind a target software architecture is to fully define the logical and physical application and data components. Bass (2007) defined software architecture as "the set of structures needed to reason about the system, which comprise software elements, relations among them, and properties of both". Architectural design shows that all stakeholder concerns are addressed by modeling and describing the architecture from separate points of view (Kruchten, 1995).

First, an overview of the technical architecture, which is separated on different layers, is presented. Next, a description of the external web interfaces used is provided, followed by a documentation of the different system components and their relationships in form of a class diagram.

4.2.1 TECHNICAL ARCHITECTURE

For the very reason that the artifact design is used and tested by interdisciplinary and locally distributed teams, the software is hosted online on an Apache Webserver, which is installed on Linux. Data is stored and retrieved in a MySQL database system and PHP 5 is used as a server-side programming language.

On top, Wordpress is installed as a content management system (CMS), because the software is freely available, open-source and based on PHP and MySQL (Mullenweg, Boren, Jaquith, Ozz, & Westwood,

2011). The CMS basic functionalities are used for requirements elicitation by simply offering users a form with various input fields.

Wordpress is the most popular blogging and content management system ("CMS Usage Statistics," 2015) and comes with many reliable modules than can be reused. This gives us the advantage to quickly develop an applicable and convenient prototype with the support of external modules. A core plugin which was successfully integrated to Wordpress is the gamification API (Application Programming Interface) from *Captain Up* ("Captain Up," 2015). With the gamification API it is possible to reward users with points, badges and levels and display them on a leaderboard.

For the visual representation of the content, several state-of-the-art web technologies, such as HTML, CSS and JavaScript are used. With a steady increase of mobile devices (Rivera & der Meulen, 2013), our prototype was not only developed for desktop browsers, but also for smartphones and tablets.

Figure 4-1 outlines an overview of the technological architecture with relevant components represented by the prototype. The non-gamified prototype is equally composed with the same layers. However, this version does not utilize the external gamification API from Captain Up.

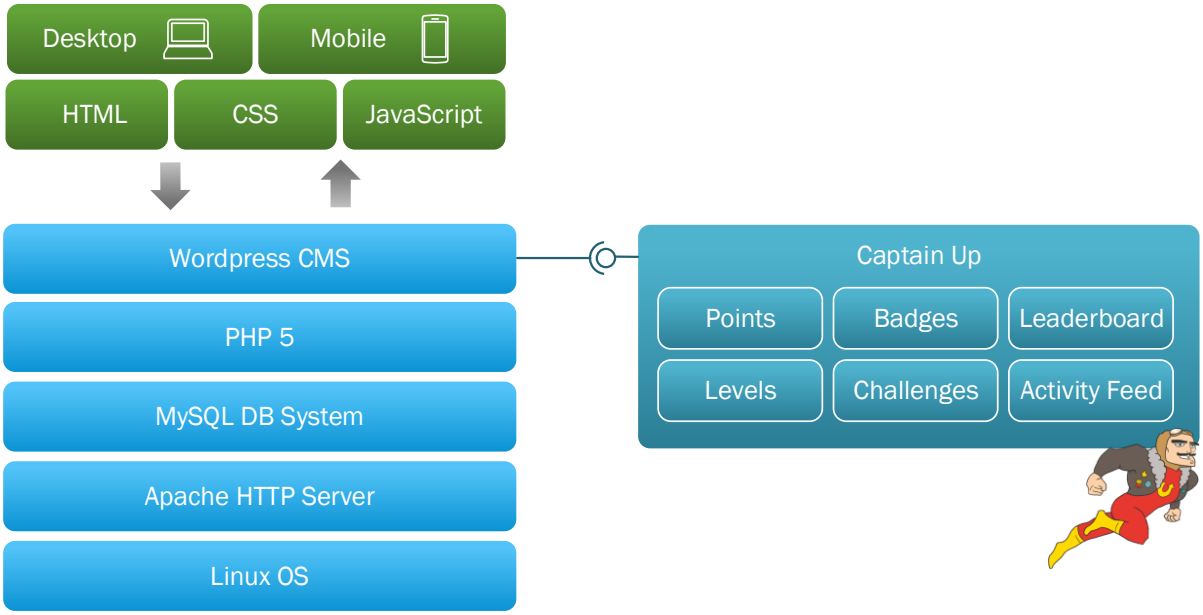

FIGURE 4-1. TECHNICAL ARCHITECTURE

4.2.2 GAMIFICATION API

Captain up is a very simple web interface allowing to turn any website into game like experience with challenges, rewards and leaderboards ("Captain Up," 2015). The software comes out of the box and can be managed and customized on the company's platform. The digital platform is offered with a freemium pricing strategy, where the basic features are provided free of charge, but a premium is charged for proprietary features (Pujol, 2010). Furthermore, Captain Up offers deep analytics on user behavior and is well trusted by many larger organizations such as Dell, MacAffee and HP.

The API is considered as an engagement layer that can be installed on top of any website and starts rewarding users for engaging with the site, using points, levels, badges and leaderboards. Every user action can be individually defined and rewarded with a certain amount of points. For instance, adding a requirement is rewarded with 30 points. Levels and badges can also be customized, including how they look and what users have to do to achieve them. Levels basically define the minimum amount of

points a user must possess. Badges are rewarded when a challenge has been accomplished. A challenge is composed of as a single action or a set of actions. For instance, writing 10 requirements results into a badge plus some bonus points.

From the literature we know that a positive reinforcement and a progression stair can keep players challenged and engaged (Werbach & Hunter, 2012). Positive reinforcement occurs through a small floating widget, called Heads-Up Display (HUD), which appears on the border-side of the screen. The HUD shows the player's avatar and the amount of missing points for the next higher level. Every time a player is rewarded with points or a badge, the HUD returns a striking notification and reinforced the user to continue. By clicking on the HUD, the game center opens and a player can see additional information about his current status in the game.

Figure 4-2 shows an example of the HUD (left side) and the game center pop-up with an overview of the player's profile (right side).

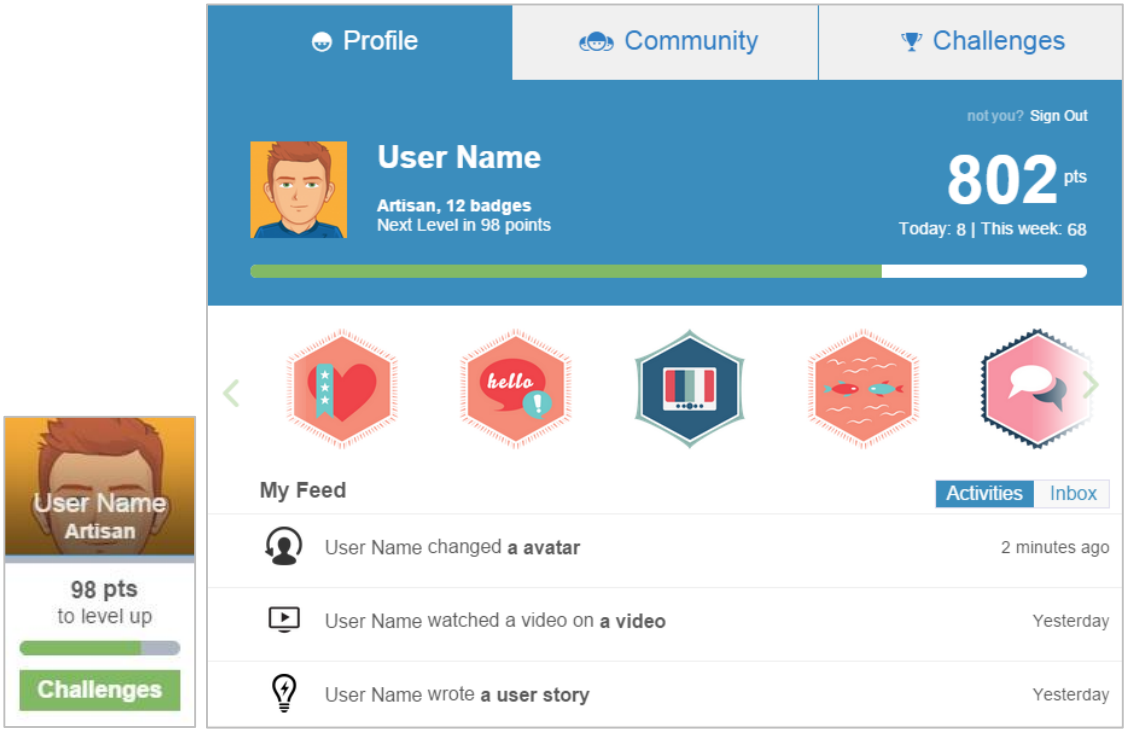

FIGURE 4-2. CAPTAIN UP HUD AND GAME CENTER

Levels and badges are used to provide a progression stair. While players at the beginning can quickly level up, this becomes more difficult as they gain more experience points. The same mechanic applies to badges: some badges can be achieved easily, others require more brain-power and skills.

Managing levels, badges, actions and points take place in the backend of the Captain Up platform. On the other hand, the management of the requirements gathering process functions is performed in the backend of Wordpress.

The advantage of using an existing solution is that it lets us focus on the game design, rather the development of a rewarding system. While most of the basic actions are already implemented and supported by the gamification plugin, some interface connections had to be additionally configured with JavaScript.

4.2.3 UML CLASS DIAGRAM

Class diagrams designed in UML are useful to show the static structure describing the functions of a system (Booch et al., 2000). The diagram allows us to illustrate the system's classes, their attributes and the relationships among objects. To highlight the main difference between the gamified and the non-gamified solution, two class diagram were designed and visualized with different colors.

GAMIFIED VERSION

Figure 4-3 illustrates a UML class diagram that is designed for the gamified prototype. The diagram consists of nine classes, showing their attributes and relationships. The four classes on the first row represent the gamification elements, which are marked in green color. Each user stands on a certain game level, depending on the total amount of collected experience points. Users are ranked on a central leaderboard based on their total number of accumulated points. To gain experience points, users can repetitively perform single actions or unlock badges, which are composed of a set of action conditions.

The core functions for the requirements gathering process are illustrated in blue color. Users can add user stories to the knowledge base and attach several scenarios to them. Furthermore, for each user story a persona must be selected who represents a certain role or function.

NON-GAMIFIED VERSION

For the control group, who was not incentivized by gamification, a similar class diagram exists, but excluding all game elements. The same class diagram is used, but with the removal of all green class elements and green class attributes. These are avatar, actions, badges, leaderboard, levels, likes and points. Figure 4-3 illustrates the static design for the non-gamified prototype.

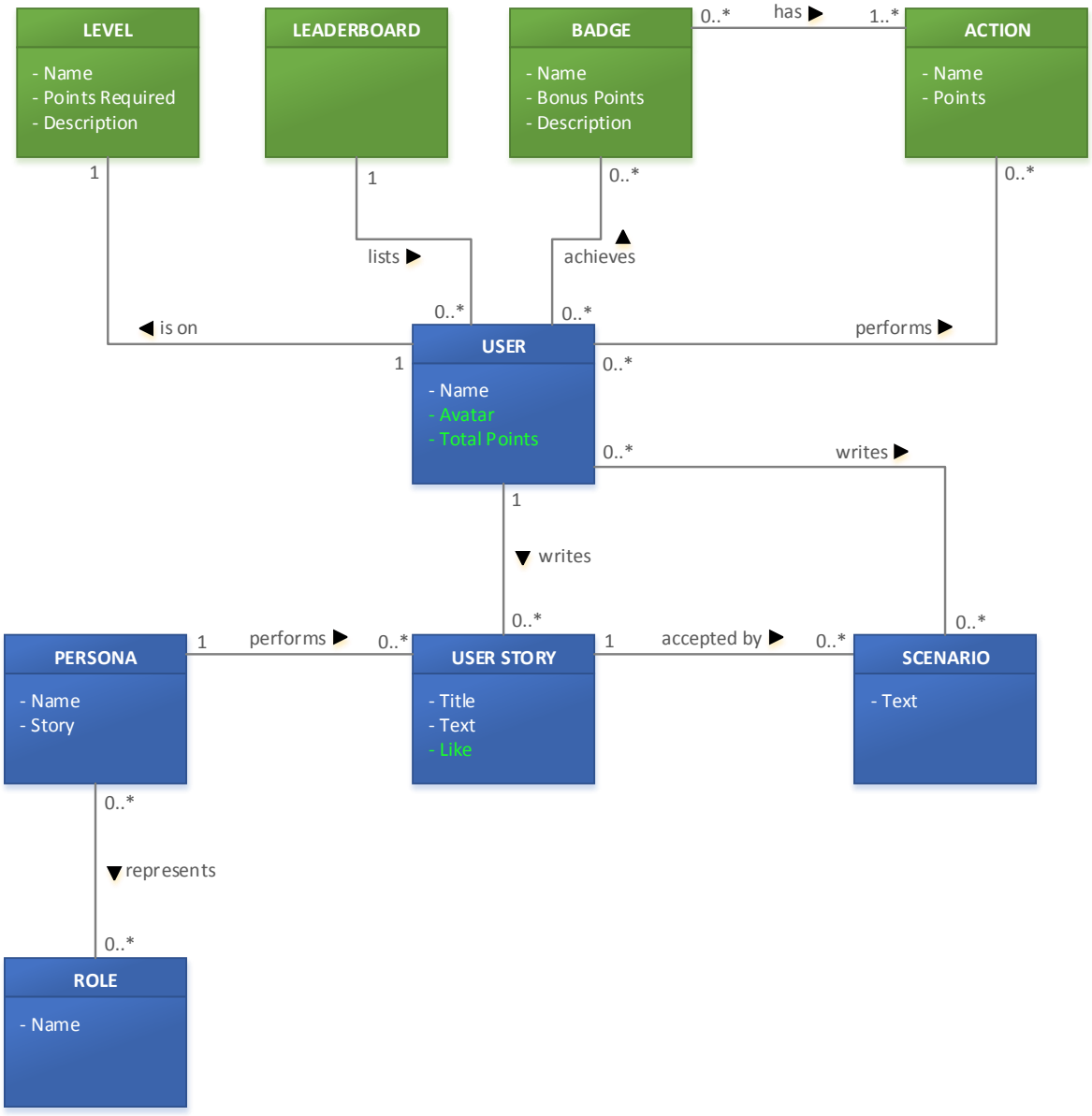

FIGURE 4-3. UML CLASS DIAGRAM

The following table provides a description of the concepts included in the class diagram.

Concept	Description
ACTION	The fact or process of doing something, typically to achieve an aim (Stevenson, 2010). ACTIONs attributed with points allowed to create behavior-based engagement for USERs. For instance, writing 1 USER STORY rewarded the USER with 30 points.
BADGE	A visual representations of achievements (Werbach & Hunter, 2012). A BADGE has a name and a description. Furthermore, bonus points can be individually assigned and USERs can achieve them when certain ACTION conditions are fulfilled. For instance, writing 10 USER STORIES rewarded the USER with a special BADGE plus 200 bonus points.
LEADERBOARD	A feature that shows how USERs stack up alongside one another (Werbach & Hunter, 2012).

LEVEL	Defined steps in player progression (Werbach & Hunter, 2012). This helps to make tasks more challenging as USERs gains new skills and experiences. USERs need a certain amount of points to rise on a new level. For instance, a user levels up when he receives 300 points.
PERSONA	A fictional character with a name, picture and other relevant characteristics, which goals have to be achieved by the system (Cooper, 1999). For every defined PERSONA a name and a brief background story is given.
ROLE	A prescribed or expected behavior associated with a particular position or status in a group or organization ("Role," n.d.). A PERSONA can represent several ROLES.
SCENARIO	An acceptance criteria complementing the USER STORY, stating when the USER STORY is complete, and ensuring that it is testable (Cohn, 2004). USERs can write SCENARIOS for existing USER STORIES. They then get rewarded with points or BADGES.
USER	A person who uses or operates a system or a software product without the technical expertise required (Abras et al., 2004). USERs have a name, an avatar, accumulated number of achieved points and are placed on a certain game level.
USER STORY	Narrative texts that describe an interaction of the user and the system, focusing on the value a user gains from the system (Cohn, 2004). USERs can write user stories for a certain PERSONA. They can then get rewarded with points or BADGES. Furthermore, USERs can like a USER STORY.

TABLE 4-2. CONCEPTS OF THE RE PLATFORM

4.3 GRAPHICAL USER INTERFACE

The visual appeal of a user interface is an important factor for effective user engagement (O'Brien & Toms, 2008). An interface should serve as connector between the user and the system. It must allow to accomplish business objectives faster and more efficient than previously used tools or methods. Furthermore, the interface should be easy to use, and evoke a sense of pleasure and accomplishment (Galitz, 2007).

For an aesthetic visual design, the prototype relies on the *Hueman* interface theme, which is available at free disposal. The theme was developed by Alexander Agnarson (2015), a front-end developer and designer from Vänersborg, Sweden. The theme can be individually customized with many extensive styling features. Furthermore, the high resolution theme automatically adjusts itself to both desktop and mobile browsers. Since its launch for Wordpress, it had been downloaded nearly 500,000 times and was rated with an average of 4.9 out of 5 stars ("Hueman Theme," 2015).

4.3.1 FRONT PAGE

The front page is divided into several areas. On top of the page lies the navigation menu that scrolls with the page. The page title together with a button containing a link to add user stories is located on the second level. Below that, a three-column layout is displayed. The first column comprises a summary of the top five rated user stories. The middle column is responsible to show the main content of the current page. The third column visualizes a stream of last activities in the game community, which was also provided by the Captain Up API. The bottom part consists of a chat function that pops up when clicked on. The purpose of this features is to allow people to collaborate and work together on requirements. Users can either interact privately or as a team in a public chatroom. The game master is also present in the chat to moderate the game. A timer is located in the footer to put some sense of pressure on the player (Moretón, 2014). Figure 4-4 shows a screenshot of the graphical user interface.

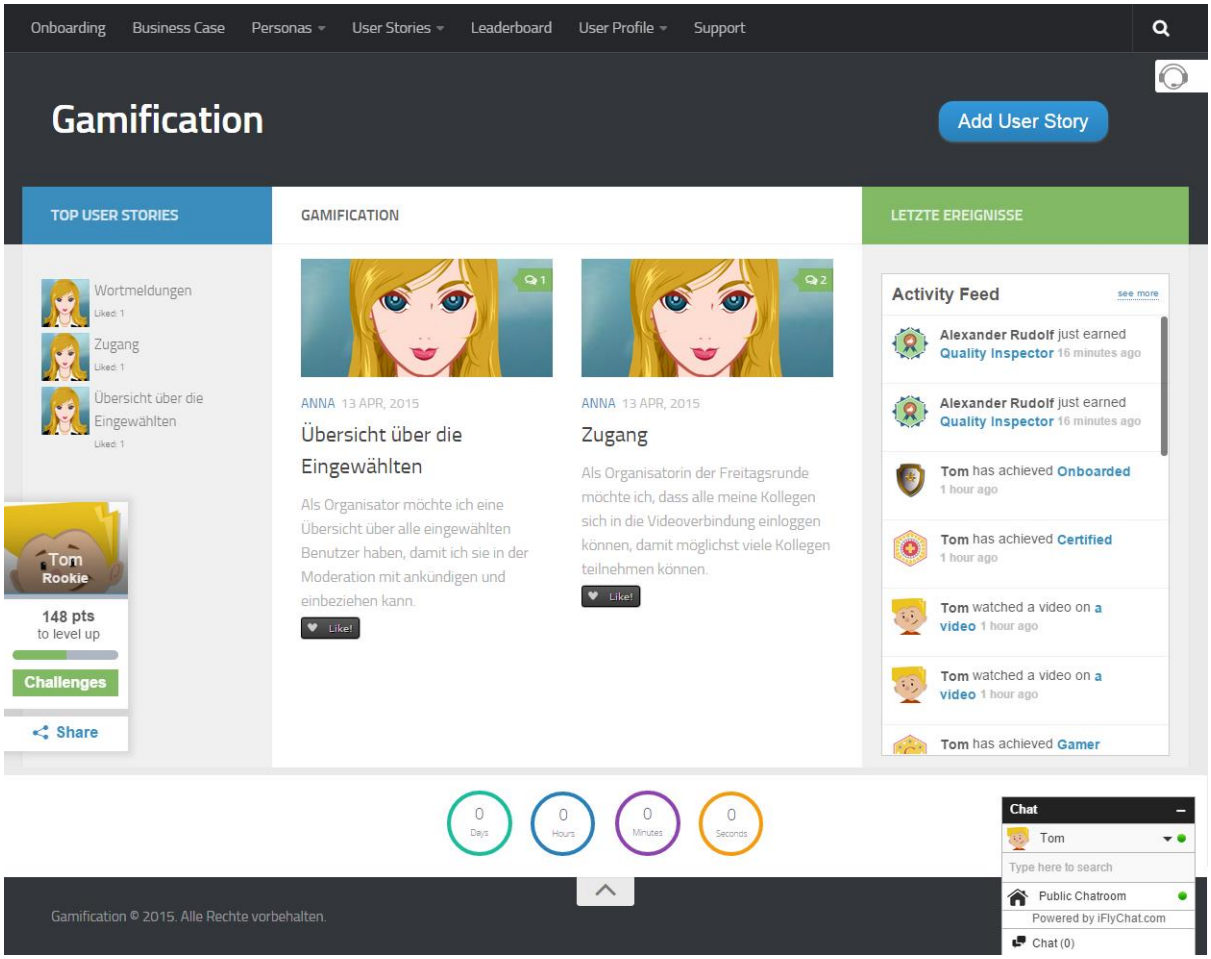

FIGURE 4-4. FRONT PAGE SCREEN

The HUD is positioned on the left side of the browser and always visible on top of the website content. By clicking on the HUD the game center opens as shown in Figure 4-5. The rest of the content in the background is shaded until the game center is closed again.

Four functions that are mentioned on the front page screen are disabled for the non-gamification version of the prototype. These are:

- Heads-Up Display
- Countdown timer
- Activity feed
- Game Master

Furthermore, the non-gamified prototype does not contain a link to a leaderboard nor to the avatar page that can be reached via a sub-menu in the user profile page.

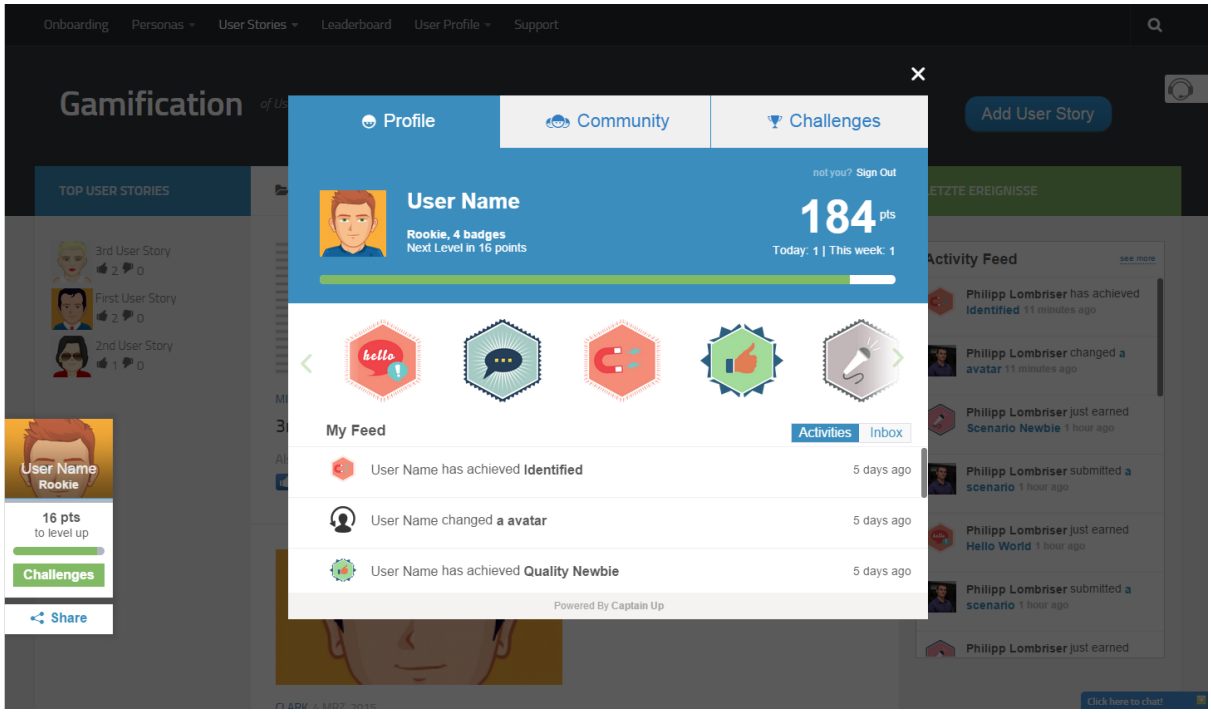

FIGURE 4-5. GAME CENTER SCREEN

4.3.2 ONBOARDING PROGRAM

Onboarding is an important task in human resource for retention management. The primary goal of an onboarding process is to help new employees increase performance levels and create higher levels of fit within both the job and the organization (Bauer & Erdogan, 2011). In this thesis the term onboarding referred to guiding a new user in a requirements development context, where he was trained on RE topics such.

Both prototypes come with an implemented onboarding program. However, while the non-gamified prototype mainly is composed of textual descriptions, the gamified version contains more joyful components, making the onboarding process more entertaining. The onboarding program in this case is not a particular game component, but rather a game mechanic (Werbach & Hunter, 2012). For instance, every page has a progress bar at the bottom, showing the current onboarding state. This should encourage users to complete the onboarding program (Myers, 1985). Users can also watch a short video about user stories and receive a reward for this action. At the end of the tutorial a short quiz is provided to the gamification team.

We assume that a diversified combination of different game elements can contribute to a significant increase of the user engagement and performance. Figure 4-6 shows an example screen of the onboarding process. The image shows an embedded YouTube video and a progress bar. In this case, the user would receive a special badge and a number of extra points for watching the video.

FIGURE 4-6. ONBOARDING EXAMPLE PAGE

4.3.3 VIDEO-BASED CHARACTER ANIMATION

Part of the onboarding program also contained an explanation of the occurring business problem. For this, the artifact design included a video-based character animation element. The use of animation is increasingly common in multimedia teaching and learning. Animations are assumed to increase interest and motivation to direct attention and explain how things work (Karppinen, 2005).

To explain the business case in an entertaining and fun way, a video was created with PowToon ("PowToon," 2015). PowToon is a cloud-based software for creating animated presentations and explainer videos. Characters and objects can be dynamically moved around the canvas with a human hand. An individual voice can be embedded in the background describing the occurrence.

This animation video is part of the onboarding program to get users familiar with the business problem. A fictional character, called Tom, was developed to guide the user throughout the entire tutorial. The main goal of the video is to get users hooked and create a fun and exciting experience. With this approach we try to provoke curiosity as an internal motivator.

An example screen of the animated video is shown in Figure 4-7. For the non-gamified version the video was replaced with a textual description of the business case and complemented with images.

FIGURE 4-7. ANIMATION VIDEO

4.3.4 ANIMATED PERSONAS

Personas play an important role when formulating user stories. A persona is an imaginary representation of a user group with specific attributes and characteristics. The objective behind personas is to not only to have stakeholders express their personal desires, but to write user stories from other perspectives (Cooper, 1999).

For the experimental part several personas were identified and developed. In the gamified artifact these personas are visualized with the free *Voki* service to make them more vivid (Otway, 2012). Voki is a web-based application that allows us to create personalized speaking and moving computer facial animations. The tool has an integrated text-to-speech feature and can be embedded on any Wordpress page. Therefore, these facial animations were customized to resemble personas. While the primary goal is to get stakeholders familiar with the personas, the idea behind Voki is to provoke some fun and novel experience.

These personas are introduced in the onboarding program where people could listen to their story. The embedded Voki layer is complemented with a profile description that contained additional details and interests. An example of a facial animation is shown in Figure 4-8.

FIGURE 4-8. FACIAL ANIMATION

For the non-gamified version the Voki element is replaced with a static image. The persona's speech is added as written text to the profile description.

4.3.5 ADD USER STORY FORM

As soon as the onboarding program is completed, users can start writing their own user stories. Hereby, a form with following three input fields is presented:

- Persona (dropdown list)
- Title (text field)
- User Story (text area)

Personas are selected with a dropdown list that contains different fictional characters. Next, the user can enter a title to label the user story. The title of the story should always briefly describe an activity that the user wants to carry out. In the text area the user story is written. The story should include a

role, a feature and a benefit. The provided template has several benefits. By specifying a role the analyst knows who to talk to about the feature. By specifying the benefit we cause the story writer to consider why they want a certain feature (Cohn, 2005).

A validity check is only performed to test if input fields were left blank. There is no regular expression validator that checked if the user adhered to the user story template. We assume that the template syntax in the placeholder of the text area would be sufficient.

The same form is applied to both prototypes. A screenshot visualizing the three input fields is shown in Figure 4-9.

FIGURE 4-9. ADD USER STORY SCREEN

The artifact also allows stakeholders to mutually rate their user stories. This rating feature is known as liking or voting and often used by social networks (Lerman & Ghosh, 2010). Liking gives users some feedback for their generated content. Displaying content with visual ratings in the web is very common today (Cheng & Vassileva, 2006). After a requirement is added to the system other users can like the user story.

4.3.6 ADD SCENARIO FORM

Acceptance criteria can be added as soon as a user story was submitted. Since we want to promote the scenario model from BDD (North, 2006), a template with the appropriate syntax is added as a placeholder in the text area. An example of the scenario template is shown in Figure 4-10.

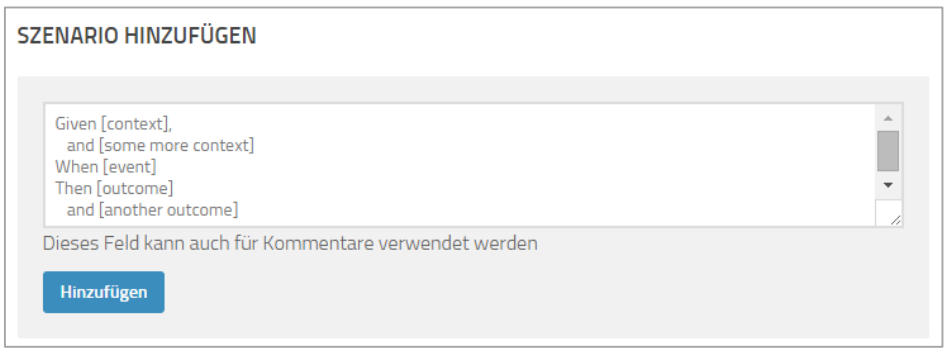

FIGURE 4-10. ADD SCENARIO SCREEN

4.3.7 MOBILE USER INTERFACE

In recent years we have seen a steady increase in the usage of mobile devices (Rivera & der Meulen, 2013). Therefore, it was an important requirement for us to have an interface that is user-friendly and applicable to smartphones and tablets. The *Hueman* theme (Agnarson, 2015) and the Captain Up plugin ("Captain Up," 2015) are both designed for mobile devices. This makes it possible for stakeholders to interact with the platform on their personal devices. Figure 4-11 shows a screenshot of the site taken from a mobile phone. In the lower left corner of the image we can see how the user is being rewarded with a badge and 50 points for his last activity.

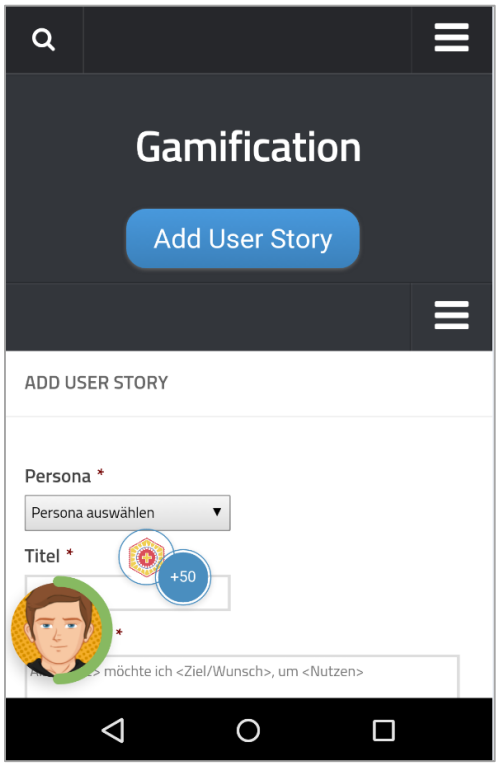

FIGURE 4-11. MOBILE USER INTERFACE

4.4 GAME ELEMENTS PORTFOLIO

The following table summarizes the diversified portfolio of game elements used for the artifact design. For each game element a column describes which party provides the feature. Furthermore, the third column contains various human desires that are satisfied by these game elements, which are based on Table 9-3 in "Appendix A3: Mapping Game Elements to Motivation".

Game Element	Provided by	Affected Motivation
Points	Captain Up	Order, Status, Saving
Badges	Captain Up	Power, Order, Saving
Leaderboard	Captain Up	Power, Order, Status
Levels	Captain Up	Order, Independence, Status
Challenges	Captain Up	Curiosity, Independence, Power
Activity feed	Captain Up	Power, Order, Status
Avatar	Wordpress plugin	Power, Independence, Status

Onboarding	Sequence of Wordpress pages	Curiosity, Independence, Tranquility
Game Master	Wordpress Plugin	Curiosity, Social Contact, Status
Storytelling	Animated character called Tom PowToon animation software	Curiosity, Independence, Tranquility
Video	YouTube Video for user story PowToon Animation for business case	Curiosity, Order, Tranquility
Facial Animation	Voki speaking avatar software	Curiosity, Order, Tranquility
Progress Bar	Wordpress plugin	Order, Tranquility
Quiz	Wordpress plugin	Curiosity, Independence, Order
Timer	Wordpress plugin	Order, Tranquility
Like	Wordpress plugin	Power, Status, Vengeance
Prize	Physical reward	Power, Independence, Status

TABLE 4-3. GAME ELEMENTS PORTFOLIO

The intention behind a diversified gamification RE platform is to satisfy as many intrinsic human desires as possible. Once some of these desires are affected, it is assumed that stakeholders will be more active in requirements engineering workshops and consequently deliver better results.

The game elements were selected whenever an opportunity was found to increase motivation or eliminate cumbersome tasks. For instance, reading a business problem is less interesting than watching a video (Alterio & McDrury, 2003). Furthermore, studies have shown that videos can have a higher impact on the learning process (Karppinen, 2005).

The design was carefully developed and maintained, ensuring that the tool was not overloaded with too many elements. After all, the users shouldn't get distracted by gamification, but rather stay focused on their main task. In the next part, the artifact design evaluation in the context of a controlled experiment is described.

5 EXPERIMENT

An empirical study is conducted to evaluate the effect of gamification within the setting of a controlled experiment. As explained below, an experimental research is ideal to refute or establish the validity of our hypothesis. Furthermore, the intention behind an experiment is to provide insight into cause-and-effect by demonstrating what outcome occurs when gamification is deployed to requirements engineering.

There is an increasing understanding in the software engineering community that empirical studies are required to improve processes, methods, and tools for software development and maintenance. Tichy (1998), Jedlitschka, Ciolkowski, & Pfahl (2008) and Runeson & Höst (2009) proposed some arguments regarding why computer scientists should experiment more. For instance, an empirical research can build a reliable knowledge base, lead to new and unexpected insights and help push into unknown areas. Nonetheless, the downside of an experiment is that it can become costly and there are many variables to control. Therefore, a control group resembling the treatment group in many demographic variables serves as comparison when gamification is evaluated.

This chapter contains the planning of the experiment. Section 5.1 defines the scope of the experiment. Next, Section 5.2 covers the design and Section 5.3 the process. In Section 5.4 we describe how the experiment was measured and which instruments were used. Finally, Section 5.5 is devoted to internal and external validity.

5.1 SCOPE

The first activity is scoping the experiment by defining its goal for the problem to be solved. In order to capture the scope, the Goal-Question-Metric (GQM) is used as originally suggested by Basili & Rombach (1988). The framework consists of five elements, which are object of study, purpose, quality focus, perspective and context. From this template we propose the following scope for our experimental research:

> Analyze the theory of gamification in requirements engineering
> for the purpose of evaluation
> with respect to its effectiveness and influence
> from the point of view of the researcher
> in the context of a stakeholder sample expressing their requirements.

5.2 DESIGN

The gamification theory is evaluated in an offline laboratory setting where a limited part of the requirements engineering process is executed in a controlled environment. The experiment is intended to compare the cause-and-effect relationship between the results obtained from a treatment sample against a control sample (Wohlin et al., 2012). In this laboratory experiment the treatment group receives an intervention in form of gamification to incentivize intrinsic and extrinsic motivational factors (see Figure 5-1). The goal of the experiment is to measure the response of the intervention by means of an ex-post measurement of the expected effect.

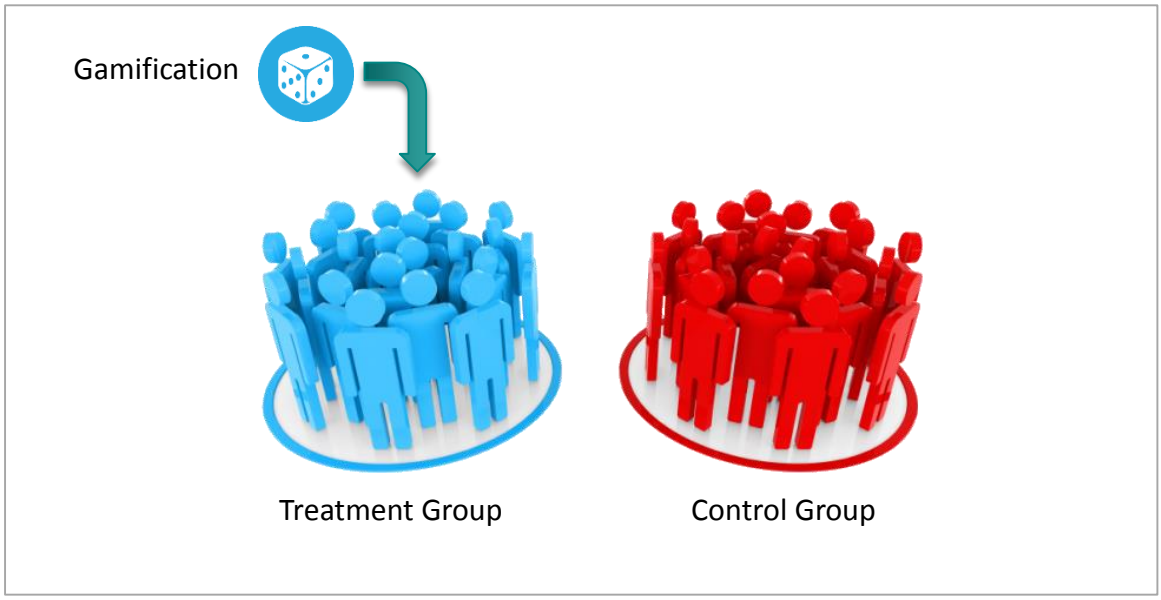

FIGURE 5-1. LABORATORY EXPERIMENT

5.2.1 SELECTION OF SUBJECTS

To generalize the results, Wohlin et al. (2012) suggested to execute the experiment in a real software project, with professional staff members. The selection of subjects must be representative for the population. Therefore, employees from MaibornWolff are selected and provided with a real business case. The company was founded in 1989 and occupies more than 140 employees with an average age of 33.4 years. Furthermore, 30% of the workforce consists of women (“Facts and Figures,” 2015).

Since it is practically impossible to select subjects from a list of the population at random, a non-probability sampling was conducted. For this the nearest and most convenient persons are selected, which resolves in a quasi-experiment (Wohlin et al., 2012). To find enough volunteering and motivated participants, an invitation was announced on the corporate intranet. Furthermore, the experiment was mentioned in the weekly staff meeting where people were convinced to participate in the experiment.

Although it is assumed that a good sample size should consist of minimum 30 people to obtain statistical significant results (Gravetter & Wallnau, 2013), we had to rely on existing resources provided by MaibornWolff. Furthermore, based on the feedback from the pilot phase we figured out that too many people can cause confusion on the platform. Both available employees and lessons learnt from the pilot study led to an allocation of 6 people per group.

The experiment uses a balanced design, which means that participants are divided into two groups with equal number of subjects. It was also ensured that both groups contain similar domestic characteristics and professional work experience. Balancing is desirable because it both simplifies and strengthens the statistical analysis of the data (Wohlin et al., 2012). Once equivalent groups were formed, both were treated identically, except for the gamification variable that was removed from the control group.

5.2.2 DESIGN TYPE

The experiment is designed to assess effect of a change in the method of requirements engineering by forming a treatment group and a control group (Bhattacherjee, 2012). The treatment effect is

measured by the difference in the posttest scores. Most common is to compare the means of the dependent variable for each treatment. Therefore, a one factor with two treatments experimental design type is applied (Wohlin et al., 2012).

The measurement of the dependent variables can be influenced by extraneous variables, called covariates. These variables need to be controlled to eliminate their effect on the dependent variable and therefore, allow for more accurate cause-and-effect measurements. A covariance design is a special type of pretest posttest control group design where the pretest is used to measure covariates of interest (Bhattacherjee, 2012). The design notation is shown in Figure 5-2. “R” represents random assignment of subjects to groups, “X“ represents the treatment administered to the treatment group, “C“ represents the covariates and “O“ the observation.

R	C	X	O_1	(Treatment Group)
R	C		O_2	(Control Group)

FIGURE 5-2. COVARIANCE DESIGN

In the pretest participants were asked for their work experience. Furthermore, their Reiss profiles (Reiss & Havercamp, 1998) were collected to study particular intrinsic motivational tendencies. These information were used as input to divide the subjects into two balanced groups. On the other hand, the posttest contained a survey to test our hypotheses.

5.3 PROCESS

The planning process for the experiment is used as a checklist and guideline of what tasks have to be done. It is a crucial step in an experiment to ensure that the results from the experiment become useful. Poor planning may ruin any well-intended study (Wohlin et al., 2012).

The experiment is divided into six phases. The process is partly iterative designed and it is possible to go back and refine a previous activity before continuing with the experiment. The main exception is when the execution of the experiment has started. Then it would not be possible to return to the scoping and planning of the experiment, since the subjects would already be influenced by the experiment. The following image illustrates the different phases of the experiment.

FIGURE 5-3. EXPERIMENT PROCESS

5.3.1 PILOT STUDY

Before conducting the real experiment, the artifact design was tested in a preliminary study. This ensured to evaluate feasibility, necessary time, design problems, select an appropriate sample size and improve upon the study design. Hereby, two internal employees from MaibornWolff received an hour to interact with the prototype and fill out all surveys. Afterwards, the pilot testers were interviewed with a structured questionnaire to obtain relevant feedback for improvements (see Appendix B: Pilot Study). The following list contains some change requests that were adjusted in the artifact design:

- ***Reduce the number of badges***: There were too many badges involved that were causing confusion.
- ***No points for page visits***: The purpose of page visit points was not comprehensible and caused distraction.
- ***Reduce the size of the onboarding program***: The onboarding program was initially too long and the pilot participants started to skip pages.
- ***Announce quiz at the beginning of the onboarding program***: One participant was negatively surprised when the quiz appeared unexpectedly.
- ***Add more content information to personas***: The personas were to week formulated.
- ***Change the chat feature to enhance collaboration***: The chat did pop-up when somebody sent a message. Furthermore, it was not possible to send private messages.

For the pilot phase only the gamified platform was tested. The reason is because the gamified platform requires at least two players to experience the competitive game elements, such as leaderboard or activity feed. Furthermore, the gamified version builds upon the non-gamified prototype. Therefore, suggested improvements that were not related to the game features also affected the non-gamified system.

5.3.2 PRETEST

After completing the pilot phase and incorporating relevant changes, the participants for the experiment were selected. All participating employees received a survey beforehand to answer demographic and work related questions. This type of information allowed us to control extraneous variables and cluster the participants into two equal balanced groups. The survey results and the grouping of the participants can be found in "Appendix D1: Pretest" and "Appendix D2: Experimental Groups". The following table summarizes and compares important indicators for the segmentation of the selected subjects:

	Experimental Group	Control Group
IT Consultants	4	4
Software Engineers	1	1
Enabler (employee for internal operations, e.g. Marketing, Accounting)	1	1
Males	3	3
Females	3	3
Years of IT experience (AVG)	6,7	7
RE Experience (AVG)	4	4,8
User Story Experience (AVG)	5,2	4,2

TABLE 5-1. SUBJECTS SEGMENTATION

Information related to personal motivation was provided by the management in form of the Reiss profile test (Reiss & Havercamp, 1998). This data can be found in "Appendix D3: Reiss Profile". All collected information for this experiment were kept confidential and anonym. This was also assured to the participants.

5.3.3 OPERATION

Next, the experimental research phase was executed. Both groups were briefed simultaneously at the same time and in the same room. Relevant information related to the experiment and the RE platform were provided in this session. Furthermore, user stories and scenarios were briefly introduced. Explaining the guidelines for both groups at the same time ensured that all participants received exactly the same information.

To ensure that the two groups would not interfere with each other, they were kindly asked to not communicate within or between the two groups. For communication purpose they may only use the integrated chat. To emphasize this statement, the groups were led to believe that the purpose of the experiment is to study online collaboration in requirements engineering. To avoid the possible threat of a competition between the two groups, they were also told that they are working on two independent business cases.

After the introduction both groups had 2 hours to interact with the digital RE platform at their personal desktops. The start and end time of the game was for both groups the same.

5.3.4 POSTTEST

After time has expired, the prototype was taken offline and all participants received an e-mail with an invitation to a post-survey. The survey results were intended to measure emotions and cognition during the interaction with the prototype. Questions were asked to rate the difficulty of writing user stories and scenarios. The questionnaire also contained two open questions where participants could share their positive and negative experience with the prototype. The gamification group had to answer additional questions related to the fun factor of the implemented game elements.

5.3.5 FOCUS GROUP

Subsequently, the two produced sets of requirements were compared within and between the groups. Duplicates and nonsense requirements were separated out. The remaining requirements were then prepared for a group of experts who judged their quality. The quality assurance team examined the results independently and did not take part in the operational phase of the experiment. This avoided the threat of having stakeholders rate their own requirements and therefore, influence the outcome. These experts mainly consisted of experienced Scrum experts and coaches.

With the aid of the INVEST model (Wake, 2003) the experts were asked to rate the characteristics and value of the requirements. The level of creativity was also measured in a separate question. Because of the large number of produced requirements we had to sample them to a more manageable size.

A great number of MaibornWolff employees were also asked to rate the user stories with the aid of the Kano model (Kano et al., 1984). This step was responsible to collect data to measure the external value of the requirements from a user perspective.

5.3.6 ANALYSIS

Finally, the produced data from the two groups were compared with appropriate statistical methods. Based on the data we either accepted or rejected our null hypotheses and then tried to generalize the results to the environment that is similar to the experimental setting.

Due to the small sample size, we used a significance level of $\alpha = 0.05$, two-sided. If the observed sample result was below the significance level, then we could conclude that the observed effect actually reflects the characteristics of the population rather than just sampling error (Fisher, 1936).

5.4 INSTRUMENTATION

5.4.1 BUSINESS CASE DESCRIPTION

To obtain valuable results from the experiment, the gamified RE artifact was tested in the context of a real existing business problem at MaibornWolff. This not only enabled the company to get an initial set of useful requirements that can be implemented in near future, but also reinforced the subjects' willingness and motivation to participate in the experiment.

Every week MaibornWolff conducts a company-wide meeting where employees can share news and insights from their projects. This meeting occurs every Friday at noon at the headquarters in Munich. However, in the case of MaibornWolff not all employees can attend on-site. Some are working in Berlin, others are located in Frankfurt and few employees also work from home. This makes it inherently difficult to keep all staff members updated in real-time. Managing the dissemination of information is very labor-intensive for the meeting organizer and should be reduced to a minimum.

VIDEO CONFERENCING SOFTWARE

A possible solution to encounter this problem is to make-or-buy an effective video conferencing system that can broadcast the meeting to all locations. Furthermore, it should also be possible for people to actively participate in the meeting independently of their location. For this very reason the company is currently planning to evaluate and compare different software solutions. To the present day the requirements for such a tool are still unclear and first need to be explicitly defined by the stakeholders. This is where our gamified RE tool came into play. With the aid and knowledge of MaibornWolff employees we tried to gather as many useful requirements to address such a video conferencing tool as possible.

PERSONAS

For this particular business case three personas were defined from which perspective users wrote their user stories. These three personas represented possible user groups for the weekly-meeting and were defined with human judgement (Brickey, Walczak, & Burgess, 2012).

The following table lists the three fictitious characters together with a brief profile description.

	Felix	**Michael**	**Anna**
			
Age:	▪ 25 years old	▪ 40 years old	▪ 32 years old
Traits:	▪ Technophiliac ▪ Methodological ▪ Extroverted	▪ Structured ▪ Communicative ▪ Perfectionist	▪ Goal-oriented ▪ Dynamic ▪ Creative

Interests:	▪ Role playing games ▪ Game of Thrones ▪ Comics	▪ Soccer ▪ Reading ▪ Art	▪ Travelling ▪ Yoga ▪ Painting
Job:	▪ Software Engineer	▪ Line Manager	▪ Test Manager
Roles:	▪ Meeting Presenter ▪ Question Responder	▪ Meeting Attendee ▪ Questioner	▪ Meeting Organizer ▪ Meeting Moderator
Objectives:	▪ Present interesting trends to the entire organization ▪ Present topics when not at the headquarters in Munich ▪ Answer the audience's questions	▪ Listen to other speakers ▪ Follow the meeting from a personal device ▪ Ask questions when not physically attending the meeting	▪ Rapid preparation and mobile setup of the meeting ▪ Flexible switching between moderator, speakers and audience ▪ Minimal effort for follow-up tasks

TABLE 5-2. PERSONAS USED FOR BUSINESS CASE

Felix represents employees who want to present new insights from their projects. He also likes to answer questions to other audience. Michael is a person sitting in the audience and listening to the presentations. Because Michael is sometimes working from home, he cannot always participate in the weekly meeting. Anna is the third persona and is responsible to organize and moderate the meeting. She ensures that all MaibornWolff employees receive a meeting invitation beforehand and the meeting minutes afterwards. Anna also takes the role of a moderator during the meeting. Their background stories were told by aid of Voki (Otway, 2012) for the gamified version, respectively with written texts for the non-gamified version.

5.4.2 GAMIFICATION API CONFIGURATION

After developing the business case together with the related personas, the Captain Up plugin ("Captain Up," 2015) was configured for the experiment. Actions, badges and levels had to be defined in the admin panel, which are responsible to keep players engaged and get them work towards a clear defined goal (Werbach & Hunter, 2012).

ACTIONS

Points are rewarded for several actions within the gamified platform. They give players continuous feedback for their actions and kept them continuously engaged. Although it is possible to distribute points for barely any action with the Captain Up plugin ("Captain Up," 2015), we decided to rather keep it on a minimum level and not overextend the player. Therefore, we preferred to only reward players for writing user stories and submitting scenarios. The following table and image consists an overview of obtainable points.

#	Action Name	Description	Time	Points
1	**Write user story**	Add a user story to the system	3'	**30**
2	**Submit scenario**	Add a scenario (acceptance criteria) to an existing user story	1'	**10**

TABLE 5-3. ACTIONS

FIGURE 5-4. EXAMPLE OF ACTIONS

Time is used as a reference point to calculate the value of a certain action. The efforts to accomplish a certain action was provided as feedback from the pilot phase. Finding new user stories is more sophisticated because they require more mental labor and creativity compared to scenarios. Furthermore, according to Pichler (2010) (an Agile Product Management and Scrum consultant) a user story should preferably consist between 3 and 5 acceptance criteria. Therefore, we decided that the value of writing user stories are worth three times as much. To increase the perceived value of these actions, the time value was multiplied with a factor of 10.

BADGES

Badges are rewarded when certain actions or a combination of actions (challenge) are achieved. Finishing a challenge also is rewarded with bonus points. The following table shows an overview of the implemented badges and how they could be achieved. Figure 5-5 shows a screenshot of four example badges.

#	Badge Name	Description	Time	Bonus Points
1	**Hello World**	Visit the homepage for the first time	1’	**10**
2	**Certified**	Read how to write user stories and scenarios	2’	**20**
3	**Gamer**	Read the rules of the game	2’	**20**
4	**Identified**	Change the avatar for the first time	2’	**20**
5	**Onboarded**	Complete the onboarding program	10’	**100**
6	**Scenario Writer**	Submit 5 scenarios	5’	**50**
7	**Scenario Expert**	Submit 15 scenarios	15’	**150**
8	**Scenario Master**	Submit 30 scenarios	30’	**300**
9	**User Story Writer**	Write 3 user stories	9’	**90**
10	**User Story Expert**	Write 10 user stories	30’	**300**
11	**User Story Master**	Write 20 user stories	60’	**600**
12	**User Story King**	Write 25 user stories and 35 scenarios	110’	**1100**
Total Points:				**2760**

TABLE 5-4. BADGES

FIGURE 5-5. EXAMPLE OF BADGES

The list of badges was developed to ensure that players experienced an optimal Flow (Csikszentmihalyi, 1991) within a 2 hour timeframe. Badges 1-5 are very simple to achieve and are rewarded during the onboarding program. This gives the players some positive feedback on a very early stage and makes them familiar with the badge system. The time values for these badges were defined based on the pilot study results and were multiplied by 10 as well. For instance, the onboarding program requires approximately 10 minutes to be completed. Therefore, 100 points are rewarded for reading the tutorial.

After completing the onboarding program, users have the opportunity to solve several challenges and receive badges 6-12. Bonus points are calculated on the basis of the estimated achievement time. For instance, to obtain badge 12 a user must write 25 user stories and 35 scenarios. Therefore, he needs to spend roughly 75 minutes writing user stories and 35 minutes writing scenarios, resulting in a total of 110 minutes. With the following formula it is possible to achieve all badges within 2 hours.

Onboarding (10') + 25 User Stories (75') + 35 Scenarios (35') = 120 minutes

If a player completes all quests, he would possess in total 3860 points, which is calculated with the following formula:

Bonus Points (=2760 pts) + 25 User Stories (=750 pts) + 35 Scenarios (=350 pts) = 3860 points

LEVELS

Throughout the game, players can level up when enough points are collected. This allows us to escalate the difficulty of the overall progression (Werbach & Hunter, 2012). For the user story game we defined seven levels. Each level has a name, a description and number of required points as shown in Table 5-5. An example of a level 2 is shown in Figure 5-6.

Level	Name	Description	Points Required	Points Difference
1	**Rookie**	You have joined our ranks. Ahead lies adventure, wisdom and an abundance of points. Work hard, and you will level up to achieve Superstardom.	**0**	0
2	**Apprentice**	The world is no longer new and the training wheels have been removed. You've proven your dedication to the cause, yet have much to learn. Press on, faithful apprentice!	**50**	50
3	**Journeyman**	The road to true knowledge begins with the understanding of how much there is to learn. Though you have come far, the journey has just begun.	**270**	220
4	**Specialist**	Few have made it as far as you've come, and fewer yet will make it further. But the world is in need of people like you. Will you rise to the call?	**730**	460
5	**Champion**	It is time for us all to stand and cheer for the doer, the achiever – the one who recognizes the challenges and does something about it.	**1480**	750
6	**Veteran**	Followers journey from the ends of the earth to behold your greatness and study under your tutelage. I hear they even turned you into an action figure.	**2560**	1080

7	**Legend**	You walk into a party and the room goes quiet. You're a living legend that we feel privileged to know. What a life you have led, but the show isn't over yet.	**4000**	1440

TABLE 5-5. LEVELS

FIGURE 5-6. EXAMPLE LEVEL

According to Werbach & Hunter (2012) the level of difficulty should not be completely linear, but rather steadily increase throughout the game. Chris Bateman, a game designer and philosopher posted several basic mathematical principles for regression ratios on his blog (Bateman, 2006). Chris proposed creating smooth curves with low progression ratios to create gentle pressure on the player. For this very reason a polynomial progression was selected, which is basically an improvement over exponential progression. The following formula was used to calculate the points required for each level:

Points required for level (n+1) = basic points * level (n) ^{y}

Basic points is a constant that decides how much points is required to go up a level. To allow players level up at the beginning of the game, the first level was set to 50 points (*constant*). Furthermore, since users can receive nearly 4000 points, we decided to set this as the limit for the last level. Based on the constant (50) and the points required for the final level (4000) we received a polynomial function with *y = 2.4456* as shown in Figure 5-7.

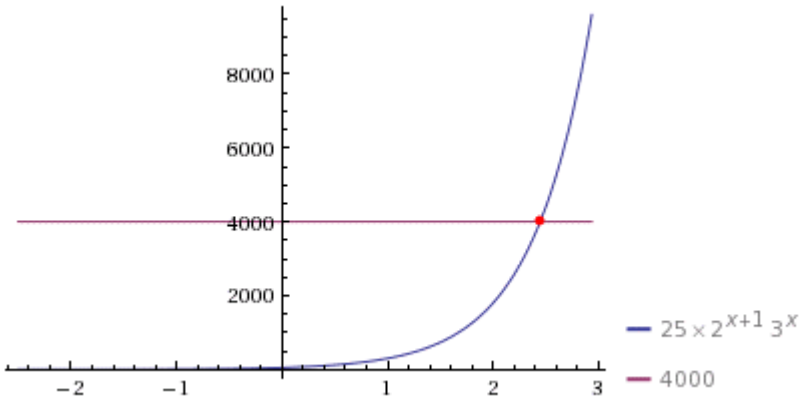

FIGURE 5-7. POLYNOMIAL FUNCTION

This provided us the following formula to calculate the points required for every level:

Points required for level (n+1) = 50 * level (n) $^{2.4456}$

With this formula the levels were reversely determined and rounded to full numbers. The following graph shows the total points required for each level together with the point gap per levels.

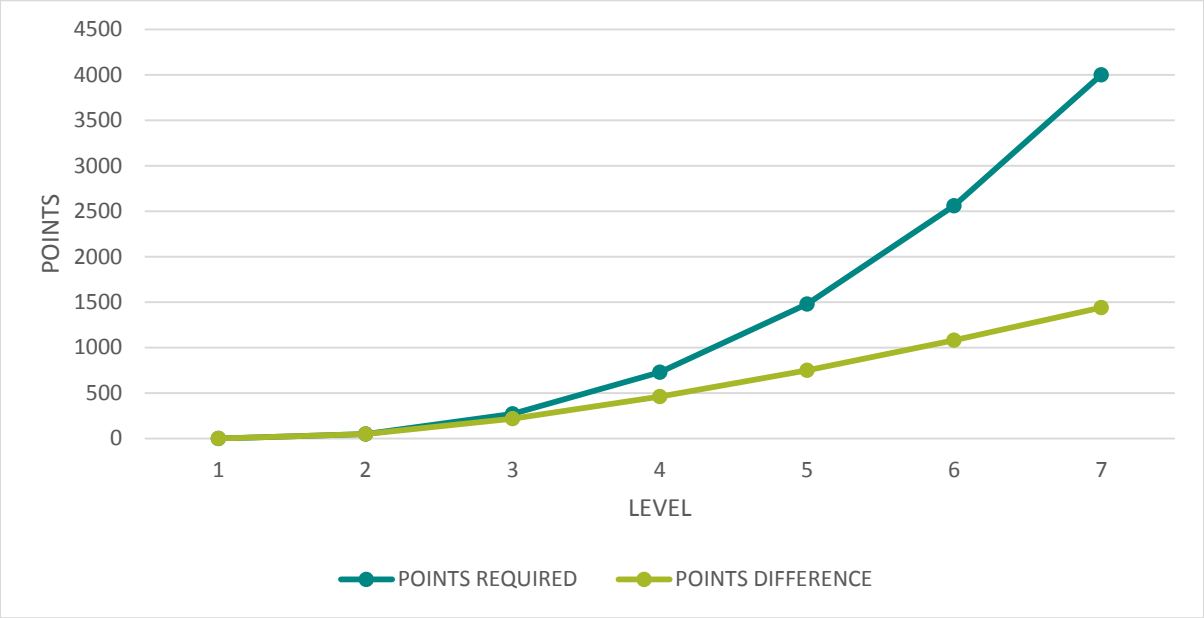

FIGURE 5-8. POINTS REQUIRED PER LEVEL

PRIZE

For the gamification team there was also a prize awarded for the winner of the game. This information was announced in the onboarding program of the gamification group. The winner received an Amazon gift card with a value of €25. The reward was not necessarily given to the first player on the leaderboard. This would have risked people writing pointless and trivial user stories just to win the prize. The number of likes a player received for their user stories were also considered to avoid a quality decrease. Every received like added 100 points to the total sum. Together with the achieved points and number of likes a mixed calculation was determine the best player.

5.4.3 MOTIVATION

The Reiss profile test results (Reiss & Havercamp, 1998) were used to gather motivational tendencies. These data already existed, because every employee at MaibornWolff had to fill out the test when joining the company. The result is a list of 16 basic human desires that can range between -2 (low desire) and +2 (strong desire). The population's normal values range between -0.79 and +0.79. People with values above and below the norm have substandard respectively outstanding manifestations. Values close to 0 signal a neutral behavior and have to be further analyzed in a face-to-face discussion.

5.4.4 GAME ELEMENTS

For the gamification group a set of questions is asked to test how satisfying the particular game elements are. This is measured on a 6-point Likert scale as described in Table 5-6. For each game element an associated image is presented to help users refresh their memory. Game elements that were not noticed by the players are ignored for the statistical validation.

- Did not notice	1. Not at All	2. A Little	3. Moderately	4. Quite a Bit	5. Extremely

1	Points
2	Badges

3	Leaderboard
4	Levels
5	Challenges
6	Activity Feed
7	Avatar
8	Onboarding program
9	Game Master
10	Storytelling
11	Videos
12	Facial animation
13	Progress bar
14	Quiz
15	Timer
16	Like
17	Prize

TABLE 5-6. SATISFACTION OF GAME ELEMENTS

5.4.5 USER ENGAGEMENT

EMOTIONS

To measure people's emotions the Positive and Negative Affect Schedule (PANAS) questionnaire (Watson et al., 1988) is used. PANAS uses self-reported responses to verbal questions that comprises two psychometric scales to measure Positive Affect (PA) and Negative Affect (NA). PANAS was developed based on the fundamental theory that the two dimensions are "independent, uncorrelated dimensions" (Watson & Tellegen, 1985). The researchers performed several specific experiments with a Varimax rotation of the two factors to ensure that the PA and NA scales would be uncorrelated. Cronbach alpha coefficient for the Positive Affect Scale ranged between 0.86 and 0.90; for the Negative Affect Scale between 0.84 and 0.87.

PANAS requires the respondent to answer a 20-item test using 5-point scale that ranges from "not at all" (1) to "extremely" (5). The measurement instrument is shown in Table 5-7.

1. Not at All	2. A Little	3. Moderately	4. Quite a Bit	5. Extremely

1	Interested	11	Irritable
2	Distressed	12	Alert
3	Excited	13	Ashamed
4	Upset	14	Inspired
5	Strong	15	Nervous
6	Guilty	16	Determined
7	Scared	17	Attentive

8	Hostile	18	Jittery
9	Enthusiastic	19	Active
10	Proud	20	Afraid

TABLE 5-7. POSITIVE AND NEGATIVE AFFECT SCHEDULE (PANAS) (WATSON ET AL., 1988)

In this experiment we only measure Positive Affect. We expect gamification to evoke fun emotions and pleasure, rather negative ones such as being scared or afraid (Attfield et al., 2011; Burke, 2014). To only measure PA, the scores on items 1, 3, 5, 9, 10, 12, 14, 16, 17, and 19 have to be added together. Higher scores represent higher levels of positive emotions.

COGNITION

To measure cognition we rely on the Flow theory from Csikszentmihalyi (1991). One out of many approaches to measure Flow is with the aid of the Flow Short Scale (FSS) (Rheinberg et al., 2003). The FSS is an instrument to measure flow in any domain and is a further development of Rheinberg's flow scale (Rheinberg, 1987). Flow components are qualitatively derived via 10 items that are measured on a 7-point Likert scale (see Table 5-8). These scales range from 1 (not at all) to 7 (very much).

The FSS has been validated and successfully applied in different domains with high reliability results (Cronbach $\alpha > 0.90$) (Engeser & Rheinberg, 2008). For this very reasons, cognition is measured in this study with the FSS as described in the table below.

1. Not at all	2.	3.	4. Partly	5.	6.	7. Very much

1	I feel just the right amount of challenge
2	My thoughts/activities run fluidly and smoothly
3	I do not notice time passing
4	I have no difficulty concentrating
5	My mind is completely clear
6	I am totally absorbed in what I am doing
7	The right thoughts/movements occur of their own accord
8	I know what I have to do each step of the way
9	I feel that I have everything under control
10	I am completely lost in thought

TABLE 5-8. SHORT FLOW SCALE (RHEINBERG ET AL., 2003)

The flow items can be interpreted via two factors: *fluency* of performance (items 2, 4, 5, 7, 8, 9) and *absorption* by activity (items 1, 3, 6, 10) (Rheinberg et al., 2003).

The FSS also contains three 9-point Likert scale questions to measure the experienced difficulty of the task, perceive skill and perceived balance. For this research we only adapted the first question to measure how difficult it is to use personas, write user stories and submit scenarios. These three questions are formulated in Table 5-9.

1. easy	2.	3.	4.	5.	6.	7.	8.	9. difficult

1	Writing user stories was …
2	Writing scenarios was …
3	Using personas for me was …

TABLE 5-9. EXPERIENCED DIFFICULTY

For the research we also added a general question on a 9-point Likert scale to measure the overall satisfaction of the requirements elicitation platform as shown in the table below. This information would reveal if the gamification platform is more appreciated.

1. Not at All	2. A Little	3. Moderately	4. Quite a Bit	5. Extremely

1	How did you like the software tool overall?

TABLE 5-10. GENERAL SATISFACTION

BEHAVIOR

To study the concrete details of how many users visited and interacted with our online RE platform, we integrated several web analytics software in the backend. With these web analytics we are able to collect various amount of data and analyze which of the two platform received more traffic.

We installed two Wordpress plugins for data collection. The first is called *Slimstat* ("Meet the leading web analytics plugin for WordPress," 2015). The advantage of Slimstat is that it uniquely assigns all activities to a user, which can then be used for our t-tests and correlation analyses. With Slimstat we can record the following type of information per user:

- Total number of page views (per user)
- Average time on site (per user)
- Average number of pages per session (per user)

Next to Slimstat, we use a second plugin to track the number of mouse clicks and touchscreen taps per visitor ("Hotspots Analytics - Heatmaps, User Activity & Custom Event Tracking," 2015). This provides us with click-through information, from which a heatmap (Nielsen, 2006) can be generated to better analyze the behavior. An example of such a heatmap is shown in "Appendix C: Heatmap".

5.4.6 PERFORMANCE

Performance is measured on a quantitative and qualitative basis to test our second hypothesis. The quality and creativity dimension requires an additional step to make the requirements comparable. The requirements were analyzed and prepared for a group of Scrum experts who subjectively rated the value and level of innovation.

PRODUCTIVITY

Productivity is concerned with the throughput and content creation. This dimension is measured with the following metrics:

- Total number of user stories
- Total number of scenarios
- Average number of scenarios per user story

QUALITY

The quality of user stories is rated with the expertise of Scrum professionals. For this step the requirements are measured with the INVEST (Wake, 2003) and the Kano model (Kano et al., 1984). Because of a large pool of gathered requirements the number of user stories had to be reduced to a manageable set. Every characteristic (INVEST) of the user stories was rated with a 5-point Likert scale (Naveed, 2014). The following scale was used to measure each characteristic in INVEST:

- ***1 - Definitely not***
- ***2 - Not sure***
- ***3 - Maybe***
- ***4 - Looks like***
- ***5- Definitely***

Due to the large sample size of requirements, the focus remained on evaluating the collected user stories. Scenarios only added to the evaluation sheet to provide the reviewer with more content. Their intrinsic quality was omitted for this research.

The importance of the user stories are classified by asking a pairwise set of functional and dysfunctional questions (Kano et al., 1984). Hereby, we asked several MaibornWolff employees and future potential system user to evaluate each user story with the following two questions:

- ***How would you feel if this user story was present?***
  - *I like it*
  - *I expect it*
  - *I am neutral*
  - *I can tolerate it*
  - *I dislike it*
- ***How would you feel if this user story was absent?***
  - *I like it*
  - *I expect it*
  - *I am neutral*
  - *I can tolerate it*
  - *I dislike it*

The gathered user stories were randomly shuffled so that it was not possible to identify from which group the user stories came from.

CREATIVITY

To measure creativity a 5-point Likert scale was added to the same evaluation sheet. The Scrum experts could assess the novelty of the requirements by selecting from a list of 5 innovation levels which they think was appropriate. The same scale to measure INVEST was applied with the question “how novel is the user story?”.

- ***1 - Definitely not***
- ***2 - Not sure***

- ***3 - Maybe***
- ***4 - Looks like***
- ***5- Definitely***

5.4.7 STAKEHOLDER EXPERTISE

Stakeholder expertise served as a control variable to test if performance is influenced by people's previous work experience. The data provided was used to mix participants into two equal balanced groups. An internal employee at MaibornWolff additionally cross-checked the groups to ensure that the participants were equally distributed. The data was collected in the pretest and includes the following 7 questions:

#	Question	Type of Answer
1	Name, Surname	Text Field
2	Age	Text Field
3	Gender	Radio Buttons (M/F)
4	Job grading	Dropdown List
5	professional experience within IT (in years)	Text Field
6	How familiar are you with requirements engineering?	9-Point Likert Scale
7	How familiar are you with user stories?	9-Point Likert Scale

TABLE 5-11. STAKEHOLDER EXPERTISE SURVEY

5.5 VALIDITY EVALUATION

A fundamental question concerning the outcome from an experiment is how valid the results are and therefore, validity should already be considered in the planning phase (Wohlin et al., 2012). In order to test the validity of this study, the factors that may affect our hypotheses had to be thoroughly examined. First, we look at internal validity, which examines whether the methods used and reasoning based upon the discovered information is sound. Next, we explore the external validity, which concerns whether the conclusions drawn can be generalized to and across individuals, settings, and times.

5.5.1 INTERNAL VALIDITY

Internal validity refers to the causal conclusion between two variables (Bhattacherjee, 2012). In this case the variable of gamification and its influence on the motivation of stakeholders in requirements engineering is measured. There is always the chance that another unknown factor contributes to the results and findings. Many potential causal relationships are already eliminated by using a control group, ensuring that our results stand up to rigorous questioning (Wohlin et al., 2012). Since the research object is of artificial nature, it might be imperfect or inaccurate. Mistakes and misleading conclusions can be drawn by using the artifact ineffectively. To avoid some errors, a pilot study was conducted to improve the artifact design and make modification to the experimental structure.

In this research three main threats to internal validity are identified. These are selection of game elements, instrumentation and selection of subjects.

First, the *game mechanics and elements* were retrieved with an explorative approach. The selection and implementation of these elements with respect to intrinsic motivation were realized for a

situation that has not been clearly defined by literature yet. To mitigate this threat, a broad range of game elements, designed to stimulate different intrinsic motivation, was selected. To find out what impact these elements have on an individual's behavior, a set of questions regarding the enjoyment factor was included in the posttest.

Next, *poor question wording and bad instrumentation* can have had a negative effect on the reliability of the measures. To avoid this risk we decided to use well standardized questionnaires with high validity and reliability, such as the Short Scale Flow (Rheinberg et al., 2003) or the Positive and Negative Affect Schedule (Watson et al., 1988).

A further important threat to internal validity is the *selection of subjects*. For the experiment we could not randomly select the subjects, 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. Furthermore, since we are conducting the experiment within an IT environment, there is less heterogeneity in our study group. Nonetheless, there were still some people with less requirements engineering knowledge enrolled in the experiment.

Fernandes et al. (2012) concluded in his research study that the graphical user interface of his software had an impact on user satisfaction. To avoid the same mistake an aesthetic theme was used for the layout of the prototype (Agnarson, 2015). Furthermore, with an experimental research we had the ability to link cause and effect through treatment manipulation, while controlling for other disruptive effect of extraneous variable such as the graphical interface.

5.5.2 EXTERNAL VALIDITY

With external validity we want to state that the results from our experiment are valid outside the actual context in which the experiment was run. There are some potential problems with regard to the external validity of the results. During an experiment people find themselves in an *unfamiliar situation*, making it difficult to guarantee that the outcome is caused by the intervention (Bhattacherjee, 2012). This is why we let people interact with the artifact design at their personal desktops. Furthermore, there are some limitations to our experimental condition. First of all, our *sample size* is relatively small to make significant conclusions (Fisher, 1936). The reason is because management could only free up a limited amount of employees for the experiment. Moreover, it could get very confusing to manage user stories on an online platform when too many stakeholders are interacting at the same time. To avoid threats to external validity we tried to make the experimental environment as realistic as possible by providing them with a real company internal business case.

A further threat to external validity is the possible *interference between the two experimental groups*. If the groups would share their experiences during the experimental phase, they could recognize the intension behind the experiment and start manipulating the outcome. To avoid this problem both groups were told that they are working on two separate and independent cases. We also kindly asked all participants in the introduction to not communicate within and between the two groups by giving them the impression that the goal behind the experiment is to test requirements engineering in an online environment.

6 RESULTS

Before the operation of the experiment was performed, all participants were simultaneously briefed together in the same room and handed the exact same instruction guidelines. By that time the participants did not know their fellow group members and would not find out before logging into the system for the first time. All participants received the same target goal for the experiment, namely to produce an initial backlog of user stories as a team that can be used for further RE analysis.

The operation of the experiment went quite smooth, however, with some issues facing the treatment group. First, there was a participant dropout after 10 minutes because of an IT emergency case. This left the gamification group with only five people. The data from this particular participant was omitted for the data analysis. Next, the treatment group experienced several performance issues, presumably caused by the gamification API. The average page load time was 2150 milliseconds, whereas the control group only experienced an average lag of 460 milliseconds. A tolerable waiting time on the web is approximately 2 seconds for a page to be loaded (Nah, 2004).

Furthermore, there were some issues with the Safari browser. Unfortunately, the gamification API did work as intended for this browser. Two players who used Safari were notified by the game master via chat after 10 minutes. Other than that, barely any interaction was required by the game master. The gamification group was continuously engaged in the requirements elicitation process. Only in case of performance troubles the game master had to intervene.

After the experiment ended and all participants had filled out the post-survey a review session was organized. In this meeting the intension behind the experiment was disclosed and an open discussion was started. Despite the same objective that was given to both groups, the treatment group was primarily working independently for themselves. There was barely any interaction between the users of this group. On the other side, the control group was intensively collaborating and forming a strategy to achieve the predefined target. On the downside, this group produced much less requirements and scenarios than the experimental group.

The following parts of this chapter represent the aggregated data from the two prototypes, which were statistically analyzed and visualized with appropriate diagrams. Data to assess the quality and creativity of the user stories were further evaluated by several Scrum experts and internal employees.

6.1 GAMIFICATION

Figure 6-1 represents a bar chart with how fun the experimented game elements were in ascending order. Storytelling was perceived as the most joyful game element (M = 4.6, SD = 0.548), followed by the onboarding program (M = 4.4, SD = 1.342) and videos (M = 4.4, SD = 0.548).

The leaderboard (M = 3.75, SD = 0.5), the badges (M = 3.4, SD = 0.548), the points (M = 3.2, SD = 1.304) and the levels (M = 2.8, SD = 0.447) were rated relatively low, although often illustrated as good examples in literature (Werbach & Hunter, 2012).

The least pleasurable game element was the facial animation (M = 2.5, SD = 0.577) that was used to introduce the different personas. Feedback from the open discussion stated that the animations provided by Voki (Otway, 2012) were too childish. The text-to-speech was also very disturbing and not

well appreciated by the players. The participants proposed that a video with a real person and voice could have improve the value of this element.

An insightful difference was found between our own customized game elements (M = 4.021, SD = 1.012) and the standardized elements that many gamification platforms offer out of the box (M = 3.438, SD = 0.801), $t(78) = 2.682$, $p < 0.05$. Detailed statistics about the mean and standard deviation of all game elements can be found in "Appendix E1: Gamification".

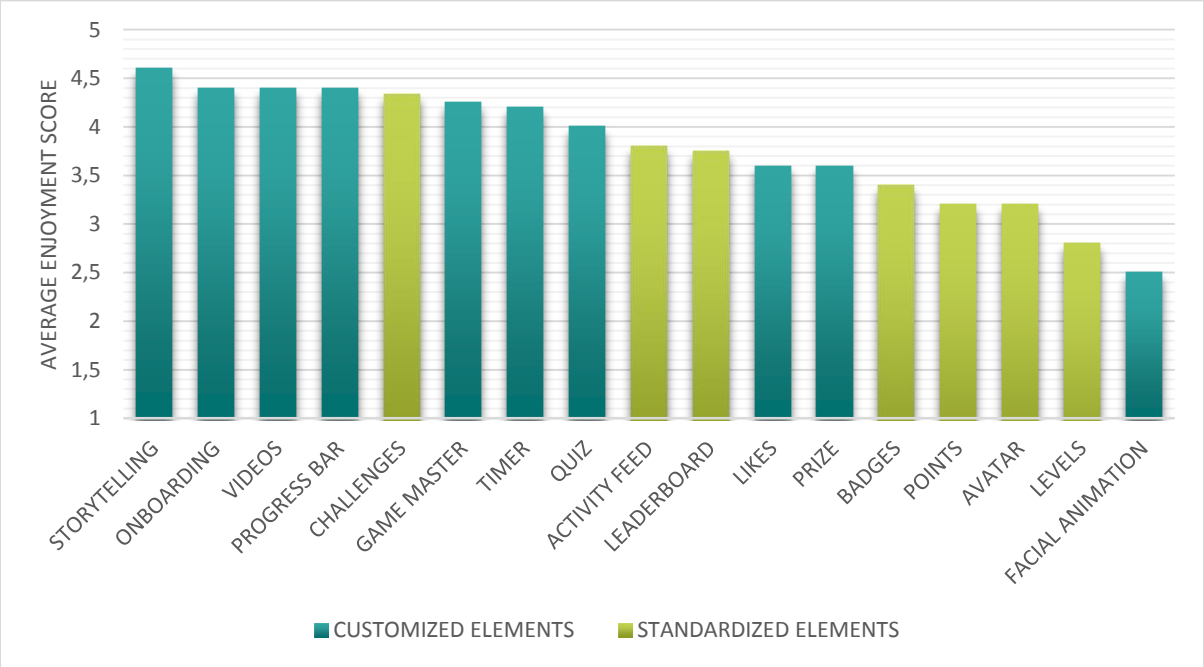

FIGURE 6-1. GAME ELEMENTS

6.1.1 EMOTIONS

Users interacting with the gamified prototype did not report a higher positive affect (PA) (M = 36.8, SD = 4.025) compared to the results reported by the control group (M = 37.0, SD = 4.0), $t(9) = -0.082$, $p > 0.05$. The group means are visualized in the bar chart below and statistically compared in Table 9-12 in "Appendix E3: Emotions".

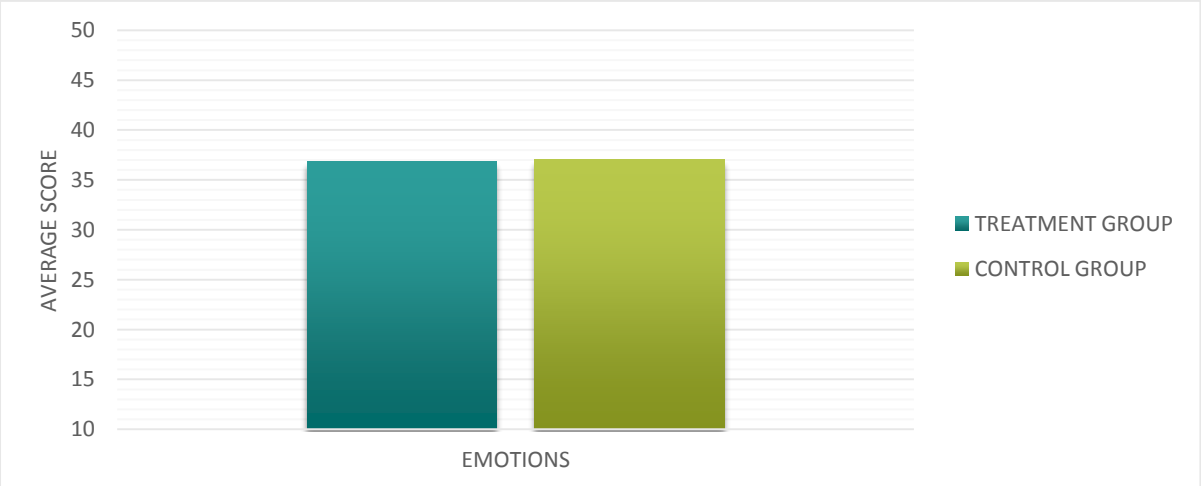

FIGURE 6-2. POSITIVE AFFECT (PA)

6.1.2 COGNITION

Figure 6-3 shows that the experimental group experienced more flow (M = 50.4, SD = 7.635) than did the control group (M = 43.333, SD = 5.645). However, this difference was not statistical significant, $t(9) = 1.767$, $p > 0.05$. The data are captured in Table 9-13 in "Appendix E4: Cognition".

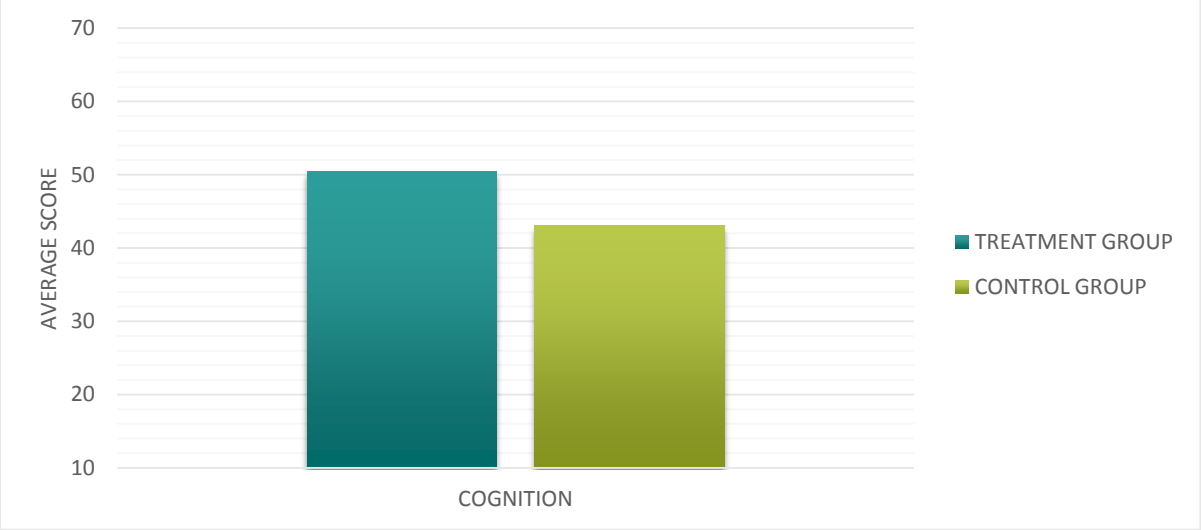

FIGURE 6-3. COGNITION

6.1.3 BEHAVIOR

PAGE VISITS

Over a two-hour time period, participants interacting with the gamification prototype significantly caused more page visits (M = 161.0, SD = 40.367) than did the control group (M = 88.833, SD = 38.338), $t(9) = 3.036$, $p < 0.05$. Figure 6-4 shows a diagram with a time plot of page visits between 09:30 and 11:30. The data for the statistics can be found in "Appendix E8: Behavior".

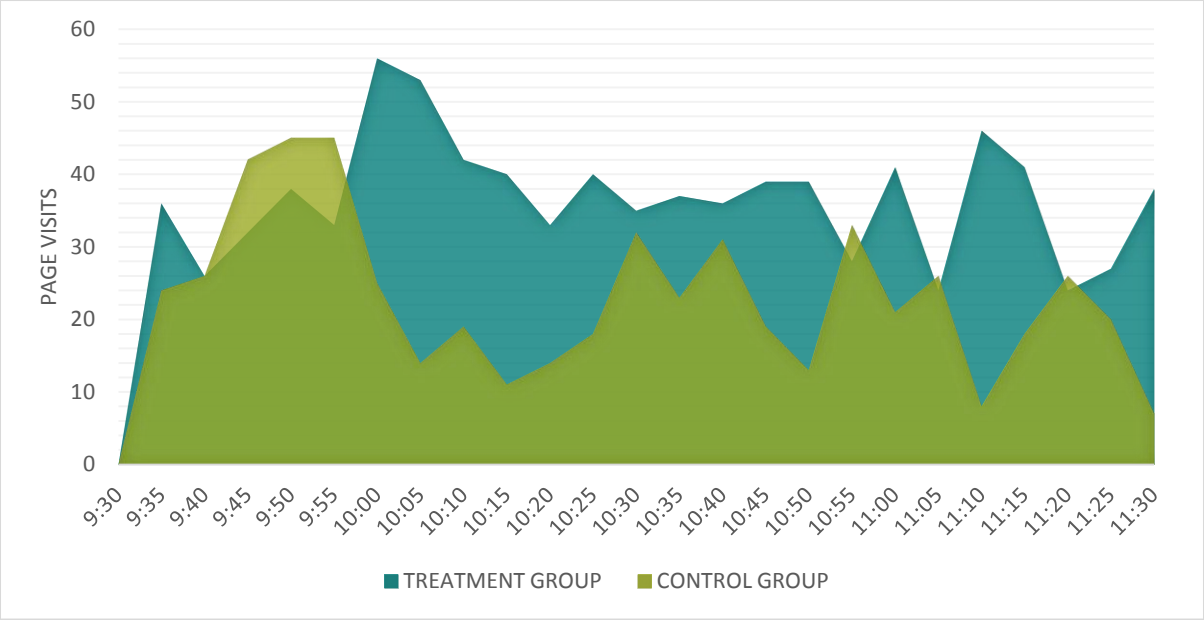

FIGURE 6-4. PAGE VISITS

COLLABORATION

During the two-hour experiment the control group wrote more text messages (M = 24.167, SD = 1.732) than the treatment group (M = 1.0, SD = 1.732), $t(9) = -2.65$, $p < 0.05$). The bar chart below visualized the total amount of messages sent between the two groups. The statistical data from SPSS are presented in "Appendix E9: Chat Messages".

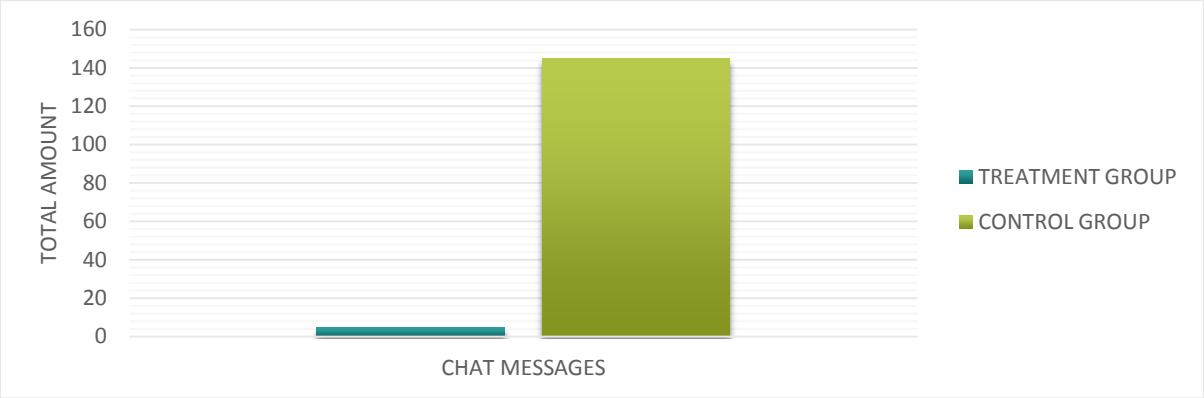

FIGURE 6-5. CHAT MESSAGES

CONTROLLING SOCIALIABILITY

Although both groups had a participant who initiated a discussion, the interaction was rather low in the treatment group. To ensure that collaboration in the control group was not naturally caused, we examined two basic human desires that were provided by the Reiss Profile (Reiss & Havercamp, 1998). These are *social contact* and *independence*. Values close to +2 for "independence" state that a person is more team oriented. Values close to -2 mean that a person prefers to be more independent. For "social contact" values close to +2 signifies that a person appreciates sociability, whereas values close to -2 implies that somebody prefers to be more reclusively.

No significant difference could be find in independence between the treatment group (M = 0.43, SD = 0.595) and the experimental group (M = 0.157, SD = 0.609), $t(9) = 0.747$, $p > 0.05$. Neither did the treatment group show any difference in social contact (M = -0.166, SD = 0.869) than did the experimental group (M = -0.258, SD = 1.006), $t(9) = 0.161$, $p > 0.05$.

Figure 6-6 shows a horizontal bar chart of the average social interest between the two groups. The statistical data from SPSS are presented in "Appendix E11: Social Interest".

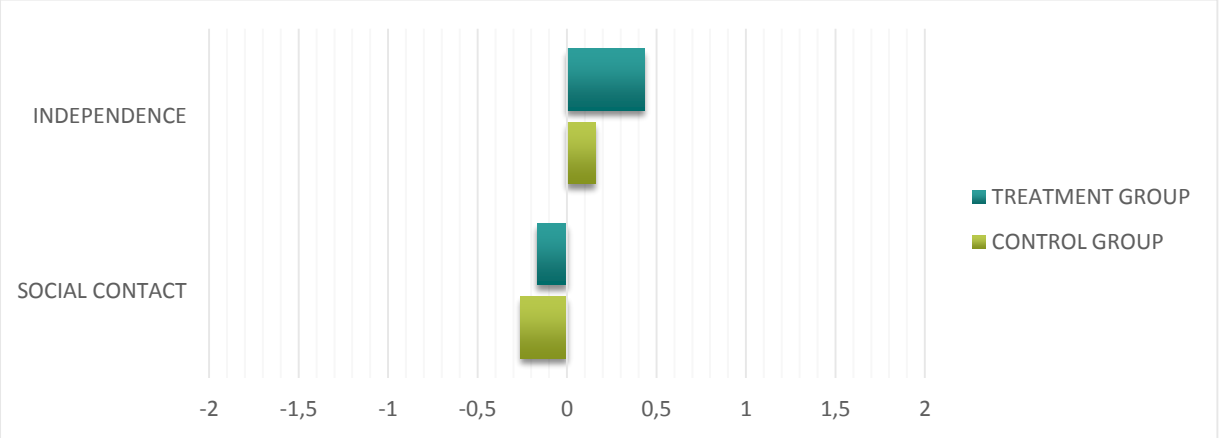

FIGURE 6-6. SOCIALBILITY

HEATMAPS

The behavioral difference between the two groups could be made visible by showing a heatmap of the two graphical user interfaces. When looking at the former image it is obvious that most of the mouse clicks from the treatment group occurred within the add user story form (see Figure 6-7). On the other side, the mouse clicks from the control groups were mainly performed at the bottom right corner of the image where the chat feature was located (see Figure 6-8).

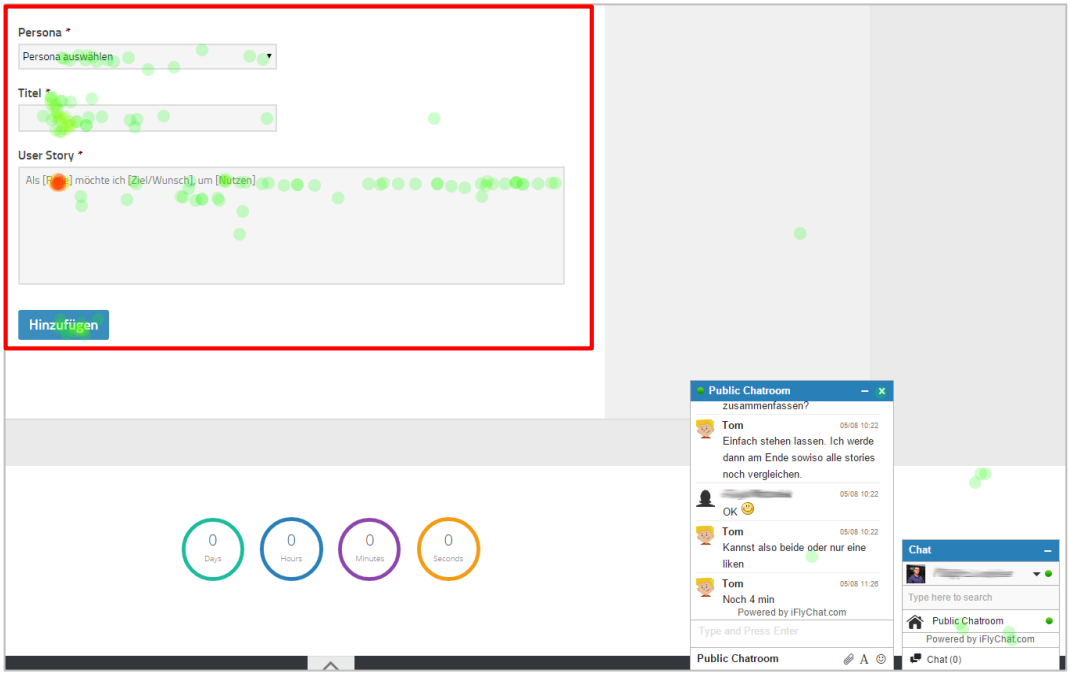

FIGURE 6-7. HEATMAP TREATMENT GROUP

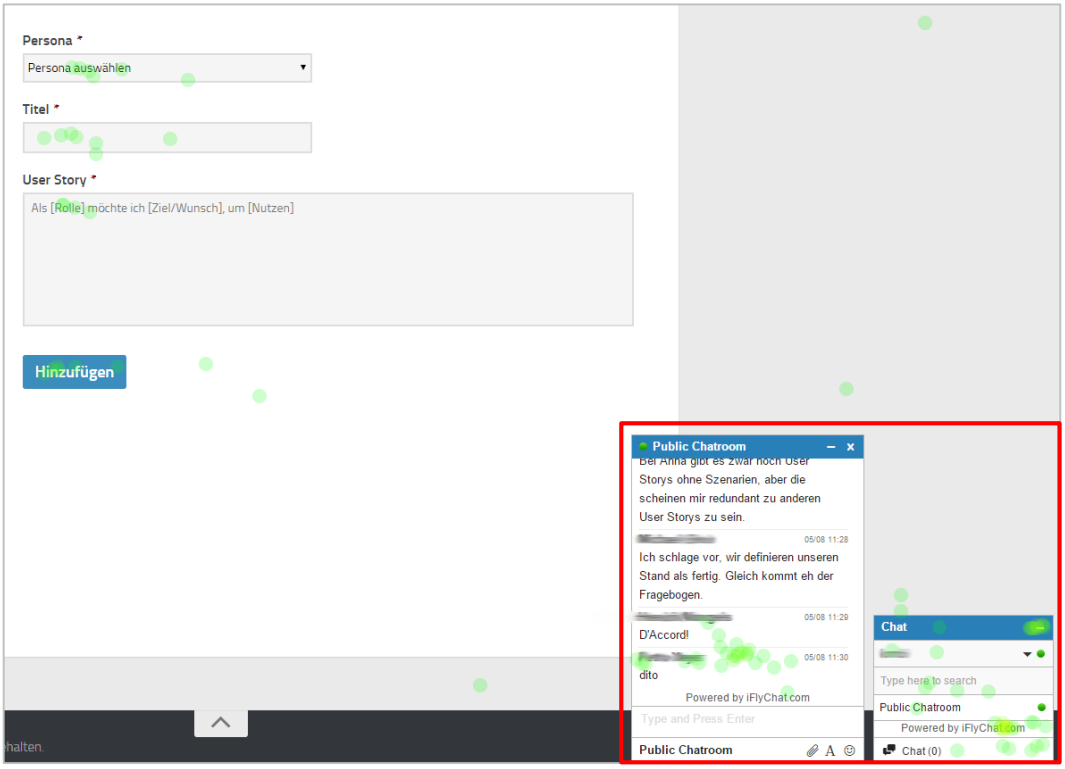

FIGURE 6-8. HEATMAP CONTROL GROUP

6.2 PERFORMANCE

6.2.1 PRODUCTIVITY

The treatment group had written in total 51 user stories, whereas the control group had written less than halve as much requirements (21). The average number of written user stories per participant for the treatment group (M = 10.0, SD = 2.345) was significantly higher than those of the control group (M = 3.5, SD = 2.258), $t(9) = 4.673$, $p < 0.05$.

The total number of written scenarios was nearly 4 times higher for the treatment group (68) than the control group (18). On average, the number of added scenarios per participant was significantly higher for the treatment group (M = 13.4, SD = 5.727) than those of the control group (M = 3.0, SD = 3.847), $t(9) = 3.597$, $p < 0.05$.

Figure 6-9 represents a bar chart containing the total number of produced user stories and scenarios for both treatment and control group. The statistical data from SPSS can be found in "Appendix E10: Productivity".

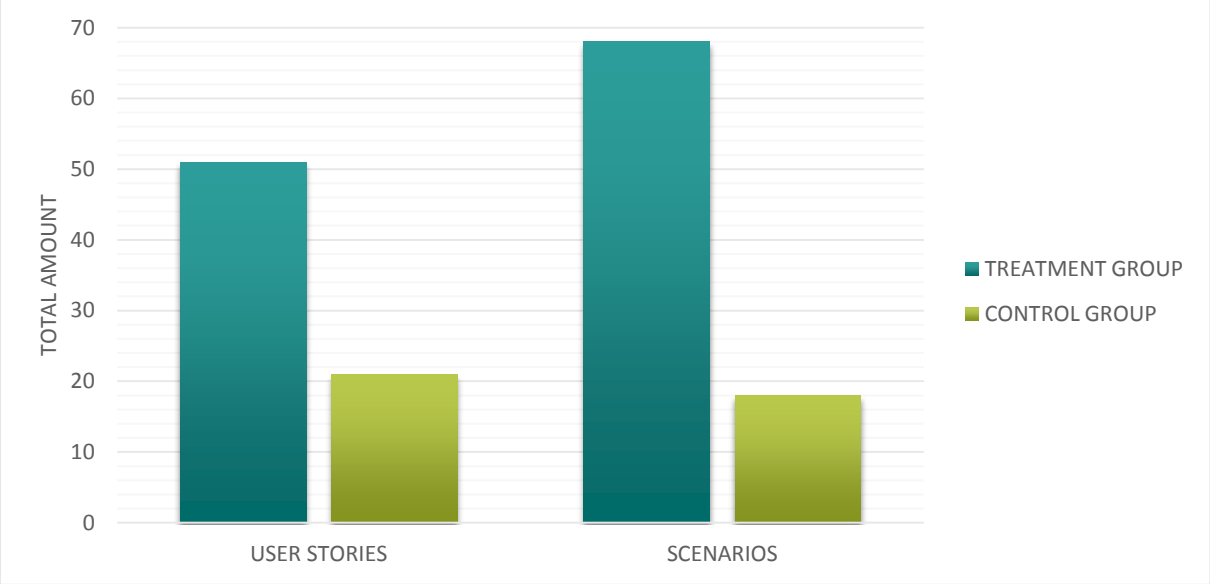

FIGURE 6-9. REQUIREMENTS PRODUCTION

Figure 6-10 and Figure 6-12 show two additional images with an accumulation of the user stories and scenarios over the two-hour time period. From the images it is obvious that user stories and scenarios were constantly produced until the end of the game. For the treatment group we identified that the last user story was written before 11 o'clock.

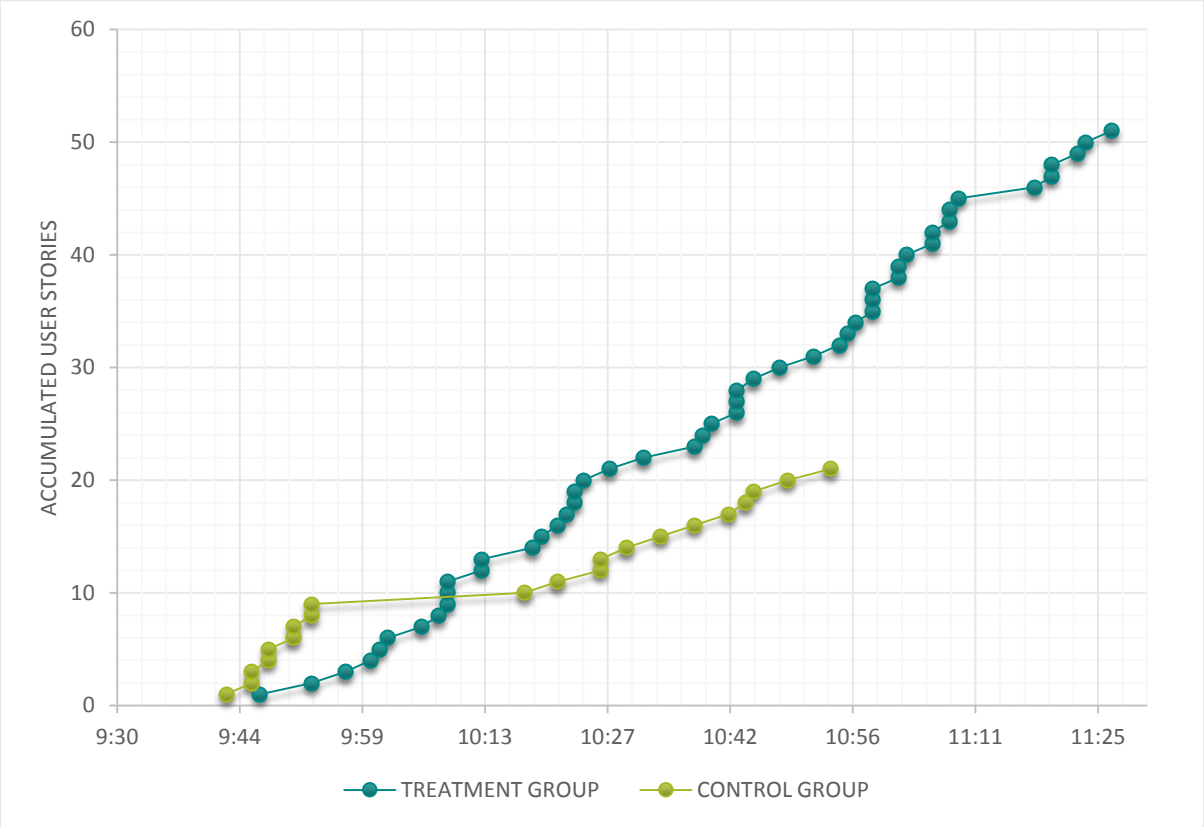

FIGURE 6-10. ACCUMULATED USER STORIES OVER TIME

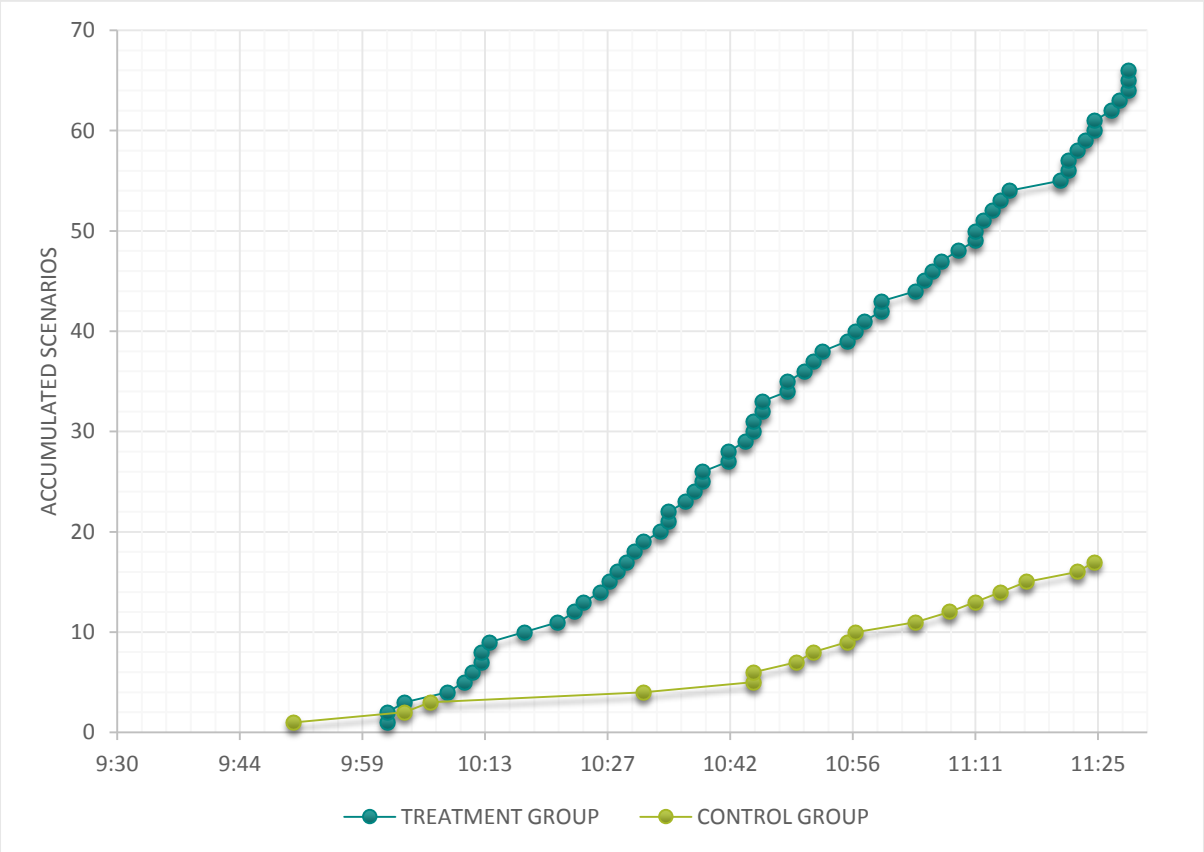

FIGURE 6-11. ACCUMULATED SCENARIOS OVER TIME

6.2.2 QUALITY

USER STORY CATEGORIES

It is striking to see that basically all stories written by the control group were also represented with similar context in the backlog of the treatment group. To better manage and analyze these user stories, we first organized them into categories. Figure 6-12 represents a structure of the different categories. The gray boxes only represent categories for the user stories in the control group backlog. The gray boxes together with green boxes represent categories existing in the treatment group backlog. Overall, 10 sub-categories were identified by the treatment group and 5 sub-categories by the control group.

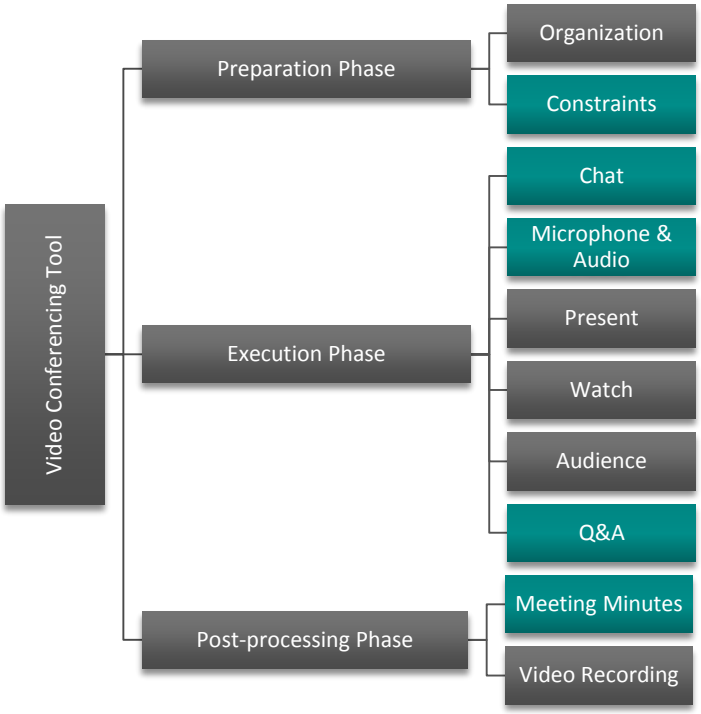

FIGURE 6-12. USER STORY CATEGORIES

After thoroughly reading and assessing all user stories, we identified that both groups used three wrong personas for their stories. Furthermore, both groups produced 10 duplicates. Comparing this number to the total amount of requirements, we receive a 20% duplicate rate for the treatment group and a nearly 50% duplicate rate for the control group. The user stories were not formulated exactly the same, but nearly identical in their content.

The stories from the treatment group wore more specific formulated. The user stories gathered by the control group were written on a more generic level of detail. Nonsense requirements could not be found in any of the groups. Table 6-1 summarizes and compares the findings between the two experimental groups. The most important differences were highlighted in bold font.

Treatment Group	Control Group
10 Categories	5 Categories
3 wrong applied personas (caused by the same person)	3 wrong applied personas (caused by 3 different people)

10 Duplicates	10 Duplicates
19,6% Duplicate Rate	**47,6% Duplicate Rate**
User Stories were **more specific**	User Stories were **more abstract**

TABLE 6-1. GROUP COMPARISON

ABSTRACTION LEVEL

If we take a closer look at the content of the requirements, we identified that most of the user stories written by the control group were very abstract. Some of them even appear to resemble epics (Cohn, 2004) rather than user stories. An example of the different level of detail is given in Table 6-2. The right column contains a user story written by the control group on a very high level, whereas the left side contains three detailed user stories provided by the treatment group.

Treatment Group	Control Group
As a meeting attendee I want to have a chat so that I can ask my questions.	***As a meeting attendee I want to ask questions remote so that I can participate in the meeting when not on site.***
As a meeting attendee I want to be hooked-up live via camera when asking questions so that my fellow colleagues can see me.	
As a meeting attendee from remote I want to understand questions in the audience so that I can follow the discussion and ask my own questions.	

TABLE 6-2. USER STORY EXAMPLES

After removing the duplicates within the groups the number of user stories per category were counted. The following star plot shows the number of user stories per category between the two groups. From the image it is obvious that the treatment group covers more requirements in all categories.

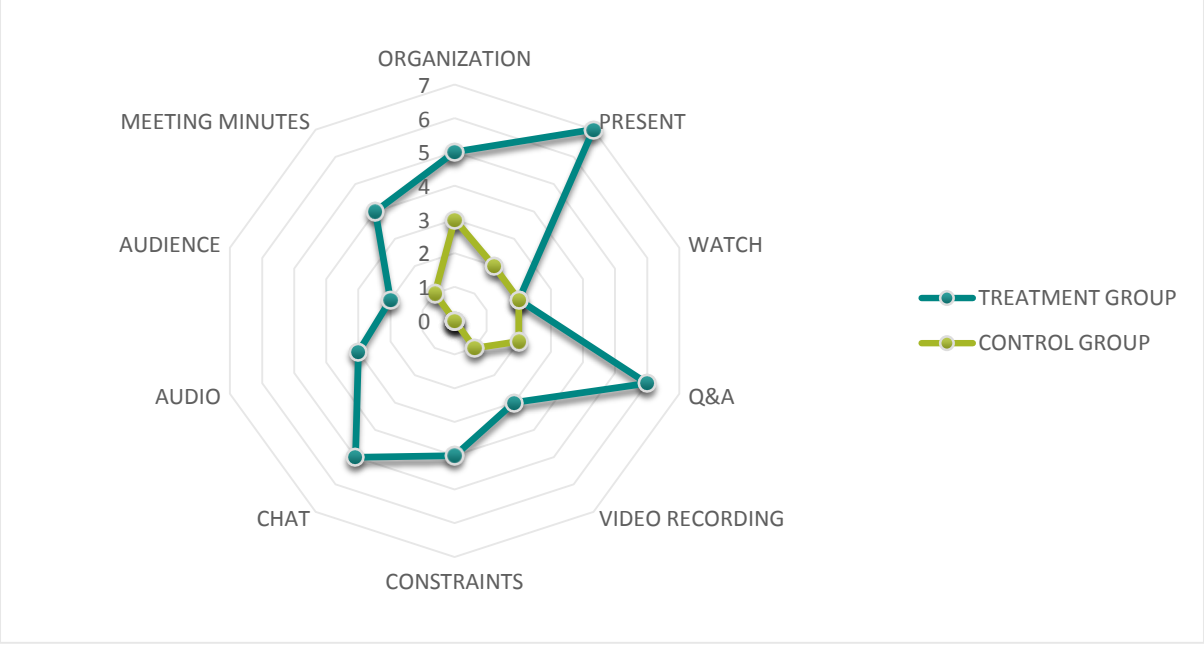

FIGURE 6-13. NUMBER OF USER STORIES PER CATEGORY

USER STORY SAMPLING

The intrinsic and extrinsic quality of the user stories were measured with the INVEST approach developed by Wake (2003) and the model from Kano et al. (1984). After the removal of all duplicates the requirements list still contained 52 user stories. For the evaluation we sampled the user story to a better manageable size. All 11 user stories from the control group were selected. The stories gathered by the treatment group were reduced to 25. We ensured that the list contained user stories from all categories. A further important factor was that each stakeholder was represented with at least one user story in requirements list. Finally, the sampling also ensured that the selected user stories were well distributed over the timeline.

Two user stories had an identical content in both groups. These two requirements were only listed once on the evaluation form. However, we had to ensure that the data influenced the statistics for both groups. The requirements list consisted of 36 user stories. Considering the two equal user stories, the list was made up of total 38 user stories.

INVEST MODEL

The intrinsic quality of the user stories was measured with the INVEST model (Wake, 2003). Five Scrum experts were asked to rate the intrinsic quality based on 5 requirements characteristics. The following table shows the list of experts that subjectively assessed the user stories with their professional expertise.

Expert	Professional Role	Agile Experience
Expert 1	IT Consultant	2 years as a Scrum Developer
Expert 2	Business Unit Manager	7 years as an Agile Consultant (Certified Scrum Master)
Expert 3	Software Engineer and IT Architect	9 years in different Scrum roles (Certified Scrum Professional und Scrum Master)
Expert 4	Agile Consultant	3 years as a Product Owner (Certified Scrum Product Owner)
Expert 5	IT Consultant	1 year as a Product Owner (Certified Scrum Master)

TABLE 6-3. SCRUM EXPERTS

Almost all user story characteristics, except for two, were statistically higher in the treatment group. The results of the INVEST approach is visualized in the bar chart in Figure 6-14. The statistical calculations are listed in Table 9-23 in “Appendix E12: Quality”.

- The user stories of the treatment group were more independent (I) (M = 4,022, SD = 0.95) than those of the control group (M = 3.436, SD = 1.302), $t(78.481) = 3.025$, $p < 0.05$.
- The user stories of the treatment group were slightly better negotiable (N) (M = 3.985, SD = 1.099) compared to the control group (M = 3.891, SD = 1.048). However, this difference was not significant $t(188) = 0.543$, $p > 0.05$.

- The user stories of the treatment group were slightly less valuable (V) (M = 3.933, SD = 1.052) than those of the control group (M = 4.055, SD = 1.061). However, no significant difference could be identified, $t(188) = -0.718$, $p > 0.05$.
- The user stories of the treatment group enabled for better estimations (E) (M = 3.504, SD = 1.177) than those of the control group (M = 2.418, SD = 1.213), $t(188) = 5.714$, $p < 0.001$.
- The user stories of the treatment group were smaller (S) (M = 3.244, SD = 1.187) compared to the control group (M = 2.364, SD = 1.007), $t(188) = 4.837$, $p < 0.001$.
- The user stories gathered by the treatment group were better testable (T) (M = 4.193, SD = 1.04) than those gathered by the control group (M = 3.418, SD = 1.37), $t(80.546) = 3.772$, $p < 0.001$.

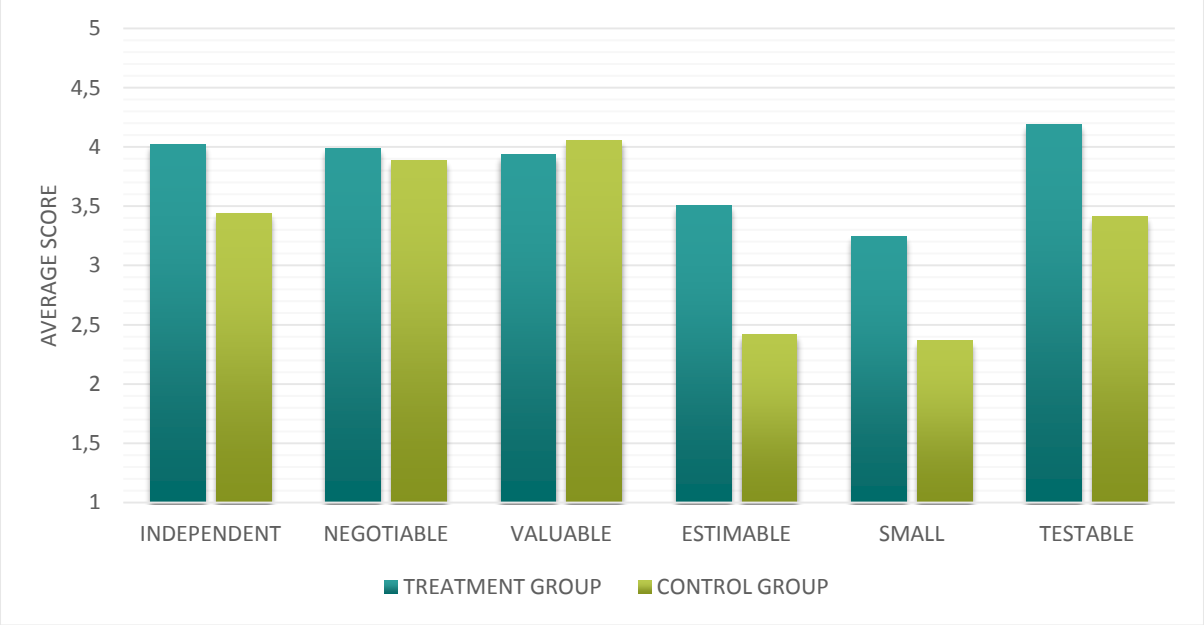

FIGURE 6-14. INVEST EVALUATION

The largest gap between the two groups was found in the characteristics “estimable”, “small” and “testable”. This is probably due to the fact that most of the user stories from the control group were written in form of epics (Cohn, 2004) rather than real user stories.

CORRELATION ANALYSIS

- There was a strong positive correlation between “estimable” and “small”, $r(180) = 0.706$, $p < 0.001$.
- The characteristics “estimable” and “testable” also correlated very strong, $r(180) = 0.708$, $p < 0.001$.
- A moderate positive correlation was found for “small” and “testable”, $r(180) = 0.556$, $p < 0.001$.

These correlation finding makes sense, since the development efforts for smaller user stories are easier to estimate and derive test cases from. The two identical user stories existing in both groups were only considered once for the Pearson’s correlation analysis, which is why N equals 180. The statistics can be found in Table 9-24 in “Appendix E12: Quality”.

From these results we can suggest that the epics written by the control group should be chunked down into further pieces to improve the value of estimable and small.

KANO MODEL

Due to the fact that most MaibornWolff employees are future potential users of the video conference software, it was a logical consequence to have them rate the external quality of the gathered user stories. The requirements list was sent together with the Kano questionnaire to several MaibornWolff employees. In total 13 internal employees responded to the inquiry, which were then used to classify the priority of the requirements.

Figure 6-10 shows a bar chart with the total number of user stories per Kano category. The results clearly indicate that the treatment group identified more attractive and indifferent requirements. The number of must-be requirements were similar in both groups. The raw data of the Kano results are captured in Table 9-25 in “Appendix E12: Quality”.

It is important to keep in mind that it was necessary to sample the user stories from the treatment group. The total size of the bars for the treatment group in this case would be even higher.

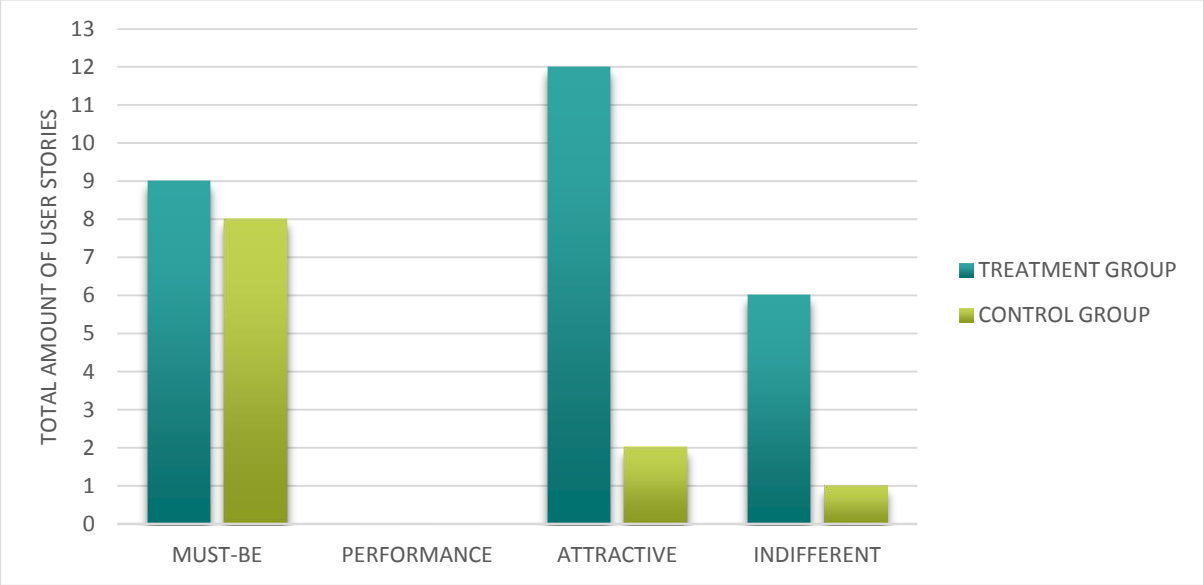

FIGURE 6-15. KANO EVALUATION

The following two pie charts show the proportion of the categories within the two experimental groups. In Figure 6-16 we can see that nearly half of the requirements were categorized as attractive requirements within the treatment group. Must-be requirements account for one third and indifferent requirements for approximately a quarter of all user stories.

On the other hand, most of the requirements in the control group were prioritized as must-be requirements, followed by attractive and indifferent requirements. Performance requirements were not represented in any of the groups. The relational allocation of these prioritizations are displayed in Figure 6-17.

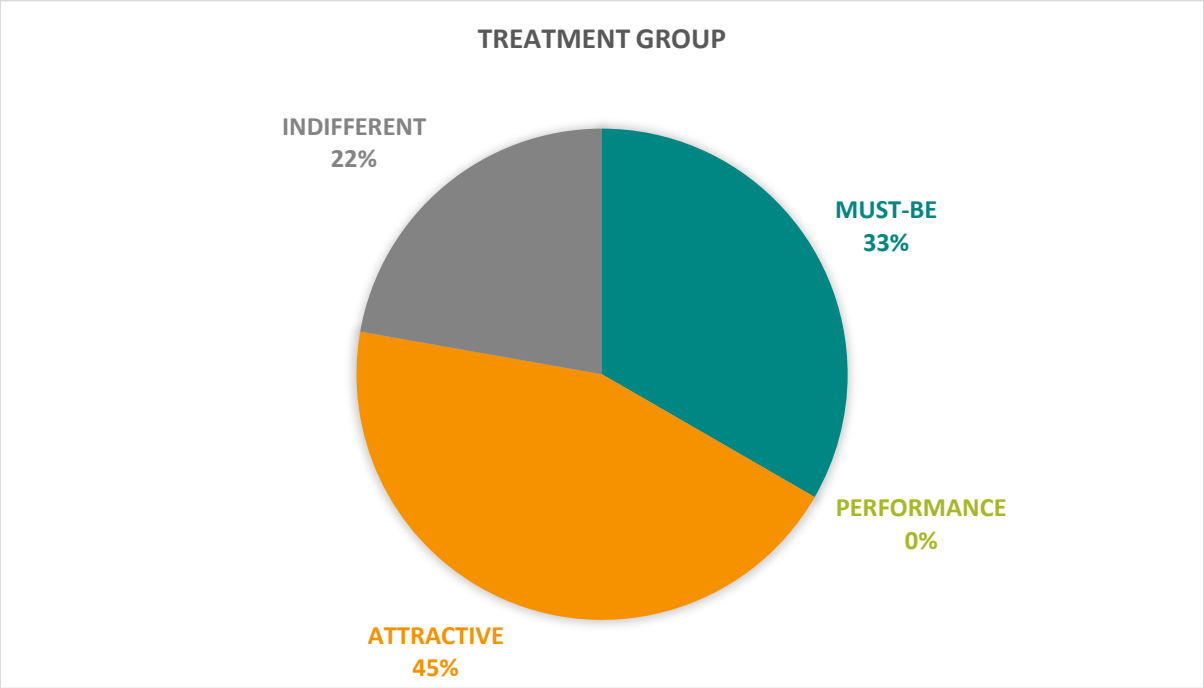

FIGURE 6-16. KANO EVALUATION WITHIN TREATMENT GROUP

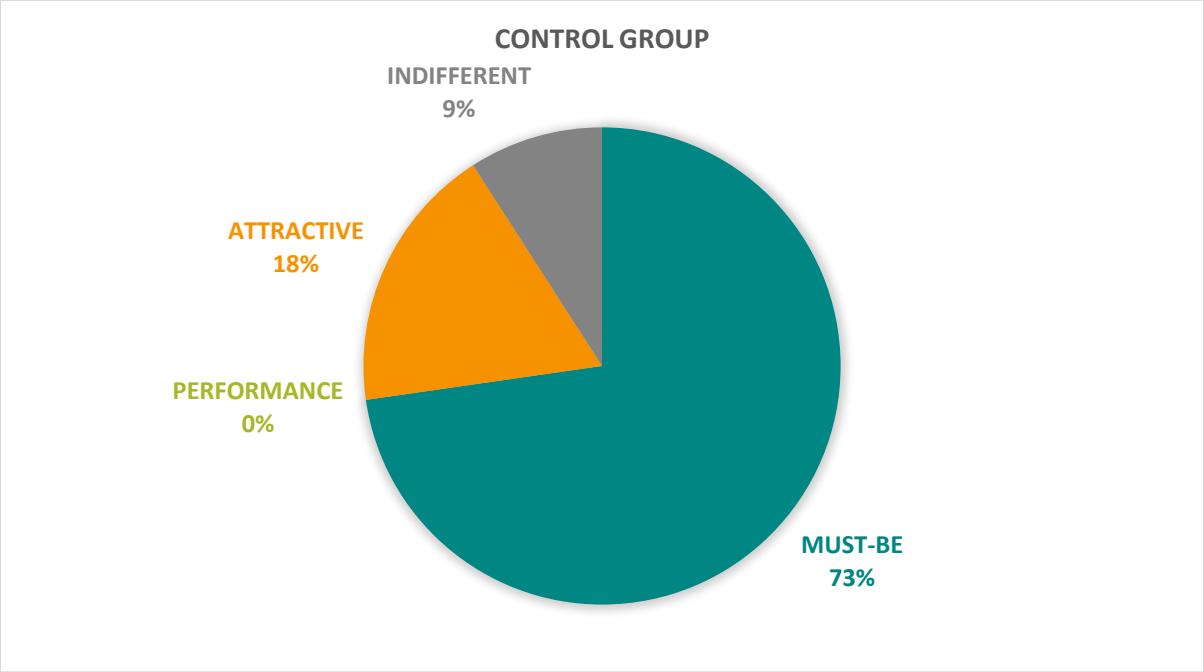

FIGURE 6-17. KANO EVALUATION WITHIN CONTROL GROUP

CUSTOMER SATISFACTION COEFFICIENT

With aid of the Kano model it was possible to calculate the customer satisfaction and dissatisfaction coefficient per user story (Berger et al., 1993). The following image shows a quadrant with a satisfaction value on the vertical axis and a dissatisfaction value on the horizontal axis. The user stories were then plotted with their CS+ and CS- value for each group.

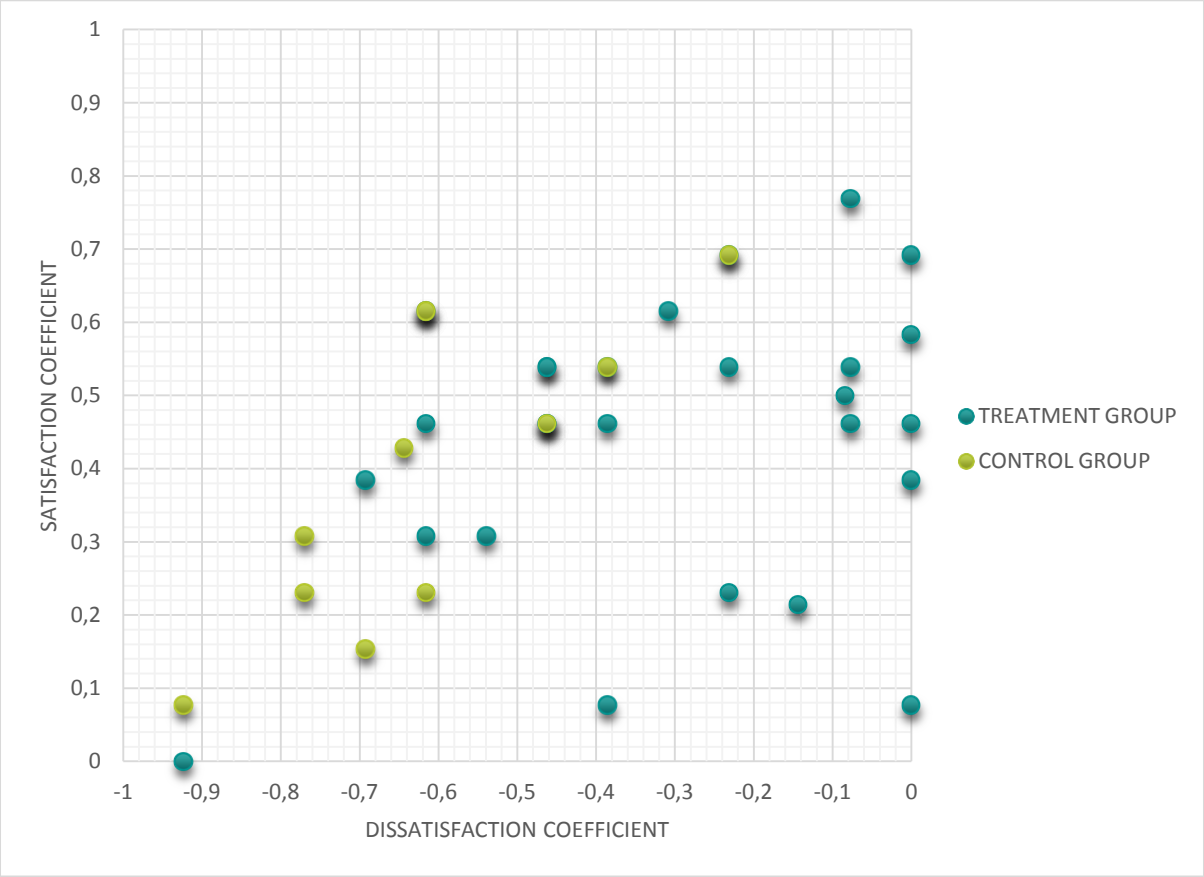

FIGURE 6-18. CUSTOMER SATISFACTION COEFFICIENT

Based on the statistical tests from Table 9-26 in "Appendix E12: Quality" the customer satisfaction coefficient (CS+) for the treatment group was slightly higher (M = 0.441, SD = 0.193) than that of the control group (M = 0.396, SD = 0.209). However, the difference was not significant, $t(36) = 0.634$, $p > 0.05$.

On the other side, the customer dissatisfaction coefficient (CS-) was significantly lower for the treatment group (M = -0.331, SD = 0.256) compared to the control group (M = -0.611, SD = 0.194), $t(36) = 3.468$, $p = 0.001$. The following image shows the mean average CS+ and CS- values per group.

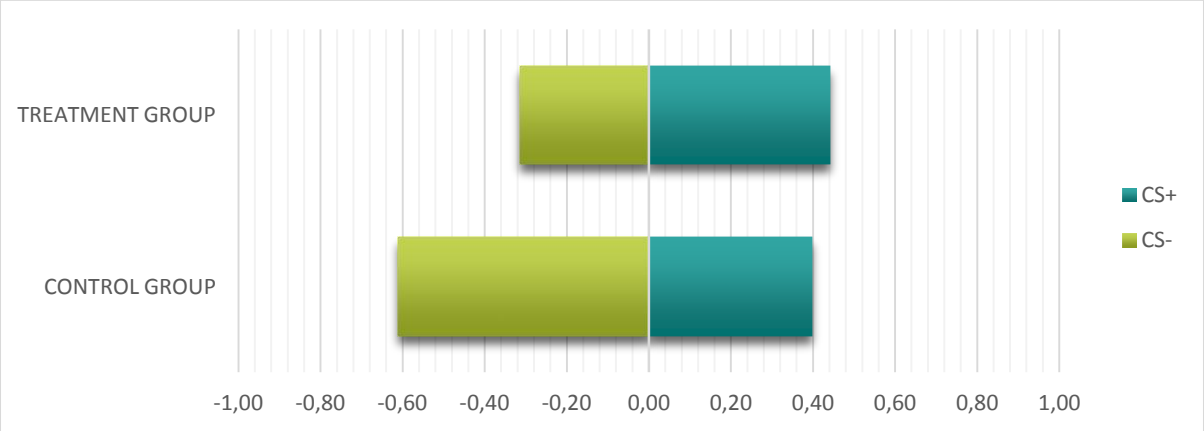

FIGURE 6-19. MEAN CUSTOMER SATISFACTION VALUE

6.2.3 CREATIVITY

Creativity was significantly higher for the treatment group (M = 3.044, SD = 1.085) compared to the control group (M = 2.236, SD = 0.922), $t(188) = 4.853$, $p < 0.001$. The following bar chart visualizes the difference in creativity between the two experimental groups. The statistical data for creativity can be found in Table 9-27 in “Appendix E12: Quality”.

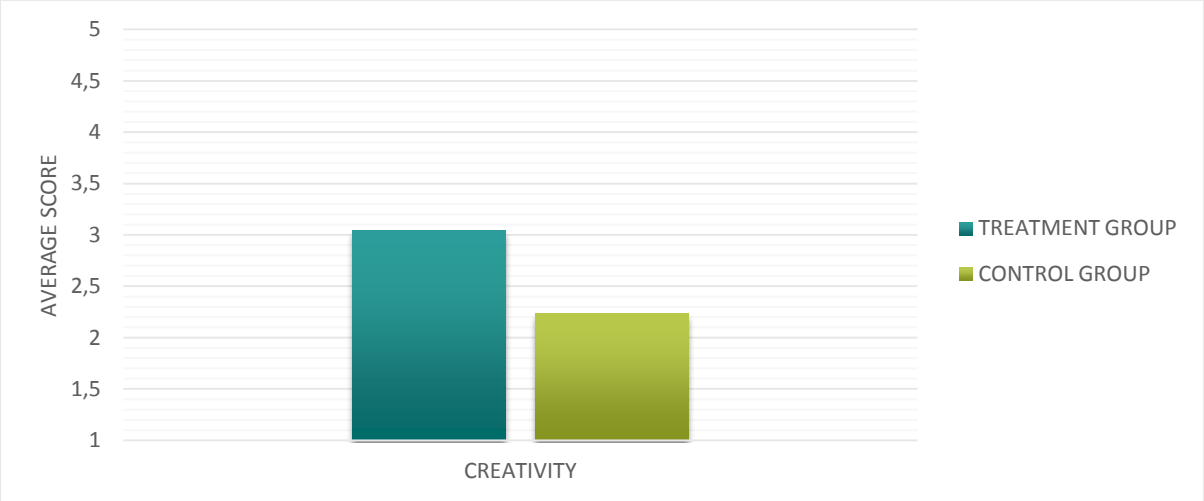

FIGURE 6-20. CREATIVITY

When comparing the mean of the individual user stories in ascending order, we can see that barely all stories written by the control group are in the lower half of the following diagram. The means and the standard deviations for the user stories can be found in Table 9-28 in “Appendix E12: Quality”.

For the graph is important to know that the content of user story A05 and B04 as well as A15 and B16 are identical, but only evaluated once in the list. The data were considered in both groups.

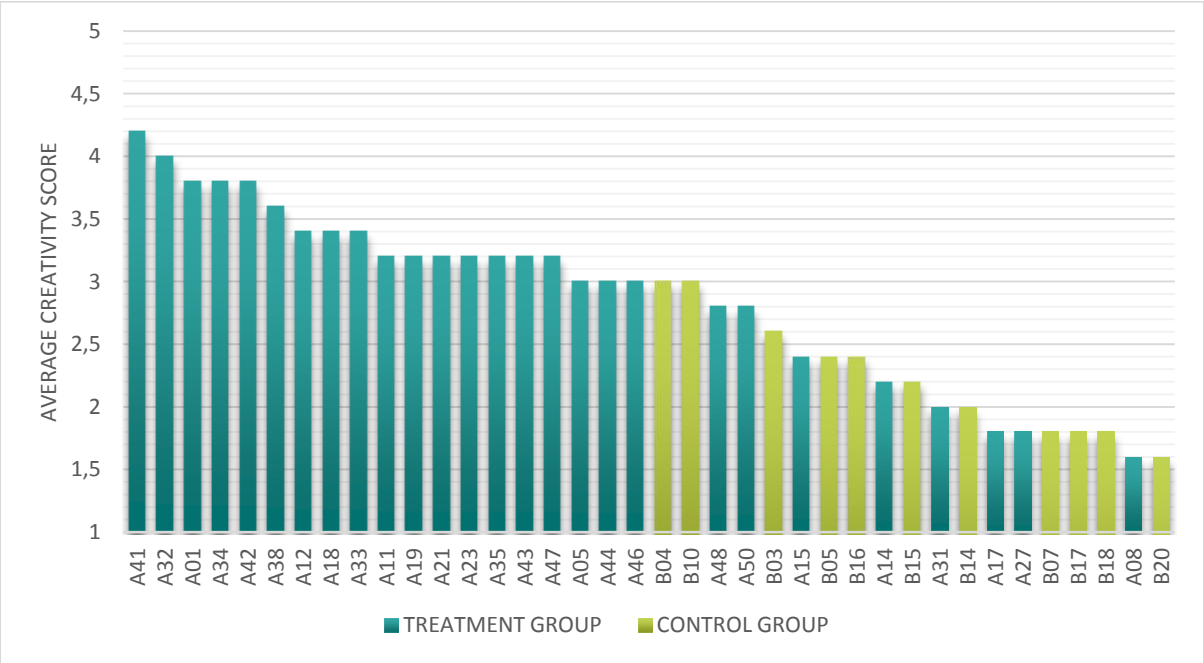

FIGURE 6-21. MEAN CREATIVITY OF USER STORIES

A further attempt to illustrate the importance of creativity was with the aid of a time plot. The timeline in Figure 6-22 shows the level of creativity per user story from both experimental groups between 09:30 and 11:30. We can see that there was a slight increase in creativity for the control group. The trend line for the treatment group was rather decreasing over time.

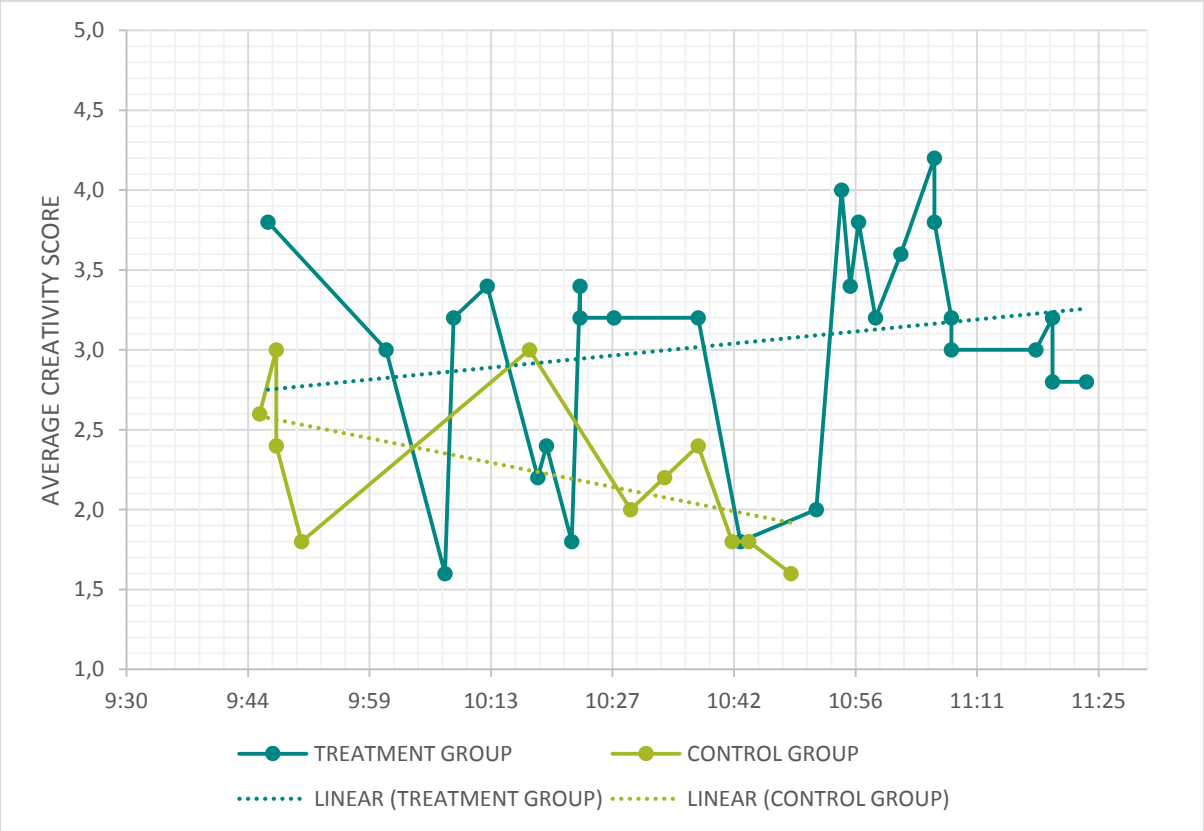

FIGURE 6-22. CREATIVITY OVER TIME

RELATIONSHIP BETWEEN CREATIVITY & KANO CATEGORIES

We can see from Table 9-29 in "Appendix E14: Relationship Between Creativity & Kano Categories" that there is a significant difference between creativity scores within the Kano categories as determined by one-way ANOVA ($F(2, 35) = 18.934$, $p < 0.001$).

A Tukey post-hoc test (Turkey HSD) revealed that creativity was higher for attractive ($M = 3.267$, $SD = 0.569$) and indifferent requirements ($M = 3.229$, $SD = 0.423$) compared to must-be requirements ($M = 2.2$, $SD = 0.511$). No statistically differences were found between attractive and indifferent requirements, $p > 0.05$. The creativity scores per category are visualized in Figure 6-2.

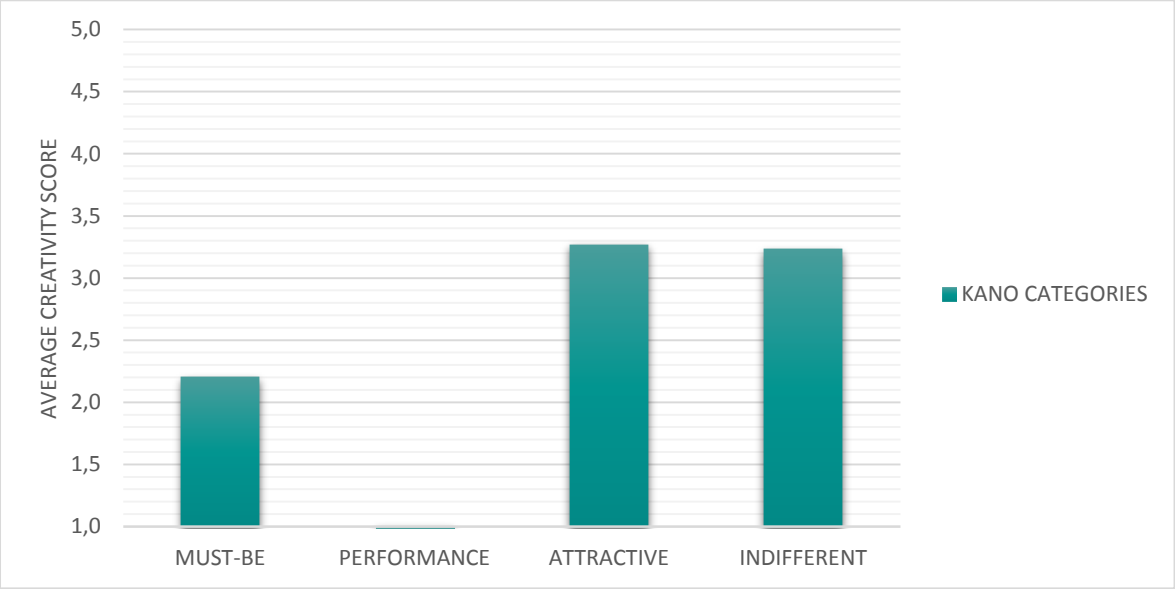

FIGURE 6-23. CREATIVITY SCORES WITHIN KANO CATEGORIES

Finally, creativity strongly correlated with the Kano categories, $r(36) = 0.632$, $p < 0.001$ as shown in Table 9-30 in "Appendix E14: Relationship Between Creativity & Kano Categories". Therefore, requirements with high creativity were more likely to be categorized as attractive or indifferent requirements. On the other side, requirements with low creativity scores were classified as must-be. The two duplicates between the groups were removed for this analysis in order to increase validity.

We decided to apply the Spearman's rho correlation coefficient, because – compared to Pearson coefficient – no linear relationship was postulated.

7 DISCUSSION

The experimental results show that the two groups exhibited different behaviors while interacting with the artifact design. Despite the same instructions given at the beginning to all subjects, this behavioral difference had a substantial impact on the outcome and quality of the trial. The group who was exposed to various game elements not only gathered significantly more user stories and scenarios, but also the quality of their proposed requirements was much higher. Intrinsic quality was measured with INVEST (Wake, 2003) and customer satisfaction with the Kano model (Kano et al., 1984). Most user stories of this group were categorized as attractive requirements, which have the greatest impact on customer satisfaction. Moreover, this group undisputedly outperformed the control group in the creativity dimension as well. Since this group identified many attractive requirements, it was a logical sense that these contained more creative ideas. The behavioral difference becomes even clearer when looking at the last ten minutes of the experiment. In Figure 6-1 from previous chapter we can see that the page visits of the treatment group sharply increased towards the end, while the number of views for the comparing group dropped down to zero.

On the other side, the control group was very communicative during the execution phase of the experiment. The recorded data indicated 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 we were able to deduce that this group was causing mutual obstruction. As a result, not only was creativity of their user stories lower, but also the intrinsic and extrinsic quality. Great uncertainty was inhibiting them from being creative and gather more than just must-be requirements. The last three characteristics of the INVEST model (estimable, small and testable) (Wake, 2003) were rated very poorly by the five Scrum experts, since most user stories were resembling epics rather than user stories.

Although collaboration is considered to be one of many success factors for requirements engineering (Kappelman et al., 2006), in this particular research we provided some evidence that it can also have negative consequences. While communication was absent within the treatment group, they still were able to outperform the comparison group in several dimensions. Due to the fact that their type of thinking was more diverging and the trail of thought much deeper, they were able to gather a great number of creative requirements in higher quality (Stroebe & Nijstad, 2004).

This chapter first discusses the measured concepts in Section 7.1, 7.2, and 7.3. Next, the hypothetical model is evaluated in Section 7.4, followed by a conclusion part in Section 7.5. Furthermore, we answer our research questions in Section 7.6 and describe the limitations in Section 7.7. Finally, some advice for future research is given in Section 7.8.

7.1 GAMIFICATION

7.1.1 COMPETITION VS. COOPERATION

Both groups were briefed simultaneously at the beginning of the experiment with the objective to create an initial backlog of user stories as a team. While the control group followed these guidelines and worked towards this goal as a team, the subjects who received the gamification treatment created user stories independently for themselves, and not as a team. The reason for this effect can be

explained by the virtual competition between the players that these game elements simulated in the playful prototype. For instance, the leaderboard served as a high-score list where leaders could gain recognition, which then should have motivated other players to take action (Werbach & Hunter, 2012). However, these competitive game elements that very often satisfy people's desire for status also bear some risks. The experiment in this research has confirmed current knowledge that a leaderboard can create unhealthy rivalries between employees in an organization. When people are competing against each other, then valuable knowledge exchange might be at serious risk. In worst case, an artificial competition might even arouse public humiliation and cause ethical damages to the company (Turco, 2013). Mario Herger (2014) mentioned in his book that "adding stress element through competition on employees can never last long. It may work for some time, but it will lead to burnout and high fluctuation in the long-term".

Generally speaking, it was beneficial to have some rivalries between the players for the RE elicitation phase when considering the production output. Competition allowed to open people's mind and therefore, explore more creative and specific requirements for the proposed business case. Nonetheless, individual leaderboards or activity feeds might not always be the right choice. They should be avoided in other RE phases where success relies on team work. For instance, estimating user stories usually requires the involvement of the entire development team (Cohn, 2004). Team leaderboards or team challenges should preferably be implemented to enhance cooperation between players.

7.1.2 EXTRINSIC REWARDS

Most implemented game elements were selected by exploring the literature for best practices and with respect to human motives. Next to PBL, which are included in most standardized gamification platforms, we also customized some game elements to the particular RE task and business case. A set of questions in the post survey measured how well these game elements were appreciated by the gamification group. To our surprise, the results from Figure 6-1 in previous chapter showed that most elements that work out of the box were lower rated than our own customized ones. Furthermore, external rewards (e.g. points, badges, levels, likes and prizes) were rather poorly accepted by the players. In contrast, intrinsically motivating game elements (e.g. storytelling, onboarding and videos), which are fun and exciting without any reward, obtained the highest of all scores. Facial animation had the lowest scores of all element, but from the post experiment discussion we know that the text-to-speech avatar mainly missed the target audience.

The main advantage of PBL is that they can be quickly implemented in any web application and with minimal investments (Lilyquist, 2014). Many industry companies like *Bunchball*, *Badgeville* and *GamEffective* saw the opportunity behind gamification and developed commercial solutions that can be scaled to any company's size and budget (Herger, 2015). While these pre-packed solutions might be effective in the short-term, they can quickly lose value and become dull, as the story of Foursquare has taught us (C. Williams, 2013). These basic game elements might be one reason why Brian Burke claimed that many gamified solutions will fail in 2014. He stated that "in many cases, organizations are simply counting points, slapping meaningless badges on activities and creating gamified applications that are simply not engaging for the target audience" (Gartner Inc., 2012). Furthermore, PBL might work better with large crowds, where points lead to social recognition and make players become a respected individual in the crowd.

7.2 USER ENGAGEMENT

In our research we observed much higher performance in the gamification group, although both groups were emotionally and cognitively engaged in similar manner. Positive emotions were reported around 75% by both groups. Experienced flow was approximately 70% for the control group and 60% for the treatment group.

7.2.1 THEORY OF FLOW

The results of our analysis closely resemble the research findings from Mekler et al. (2013a). The authors conducted a similar experiment to examine the effect of points on participant's intrinsic motivation and performance. With more than 170 participants they measured productivity and quality of online image annotation. Their data revealed that participants using the point-based system significantly produced more tags. Despite the higher performance level, the reported intrinsic motivation remained similar in both groups. As a result, the authors concluded that the rewarding point-system did not undermine intrinsic motivation. In subsequent work, the researchers conducted a further controlled experiment, but this time involving levels and leaderboards as well (Mekler, Brühlmann, Opwis, & Tuch, 2013b). The treatment groups that were exposed to points, levels or the leaderboard produced again more tags, but still with no differences concerning intrinsic motivation.

The researchers are still trying to find answers to this phenomenon and used a temporary explanation based on an article from Robertson (2010): Basic game elements such as points, badges and leaderboards are beneficial to show progress and reward user actions, but they alone don't provide a holistic game experience yet. According to Przybylski, Rigby, & Ryan (2010) a good game only becomes intrinsically motivating when players are optimal challenged and can seek for mastery (E. L. Deci & Ryan, 2010). For this very reason, we presume that an optimal flow wasn't present in our game play, because the players were not challenged enough throughout the game.

7.2.2 EXTRINSIC MOTIVATION

A second possible explanation is that the main game play in our experiment was too focused on extrinsic motivation. After completing the onboarding program, most game elements where designed to reward players with points, badges and levels for their actions. Based on a meta-analysis of 128 experiments, E. L. Deci, Koestner, & Ryan, (1999) found out that rewards can control people's behavior, but can also decrease intrinsic motivation for interesting activities. Hecker (2010) suggested that extrinsic motivation can undermine intrinsic interest for a certain task. The author and video game programmer stated that "tangible, expected and contingent rewards reduce free-choice intrinsic motivation; and verbal, unexpected, informational feedback, increases free-choice and self-reported intrinsic motivation". From these findings we assume that the implemented achievement system, which was highly driven by external rewards, might have caused an emotional and cognitive decrease within the gamification group.

7.2.3 SELF-DETERMINATION THEORY

Finally, a further conjecture might be that the control group was engaged by a social dimension. From the literature section we know that autonomy, competence and relatedness are the three basic psychological needs for self-motivation (E. L. Deci & Ryan, 2010). Autonomy was equivalent in both experimental groups. Besides a few guidelines that were provided at the beginning of the experiment,

all subjects had equal amount of freedom to act. The differences between the groups were found in the latter two motivational needs. Many of the implemented game elements such as the challenges, levels and leaderboard allowed players in the treatment group to enhance their competences with the integrated progression and feedback mechanism. On the other side, relatedness was highly satisfied by the users interacting with the non-gamification prototype. As a result, while one group was primarily progressing in a virtual game and enhancing their competences, the other group was socially engaged in the requirements elicitation process.

7.3 PERFORMANCE

7.3.1 USER STORIES & SCENARIOS

User Stories combined with the scenario technique from behavior-driven development were selected as a mean to gather requirements in structured form from stakeholders. The results in "Appendix E6: Experienced Difficulty" have shown that user stories together with personas were relatively simple to apply. In contrast, participants reported that the use of the scenario template was more complex. From the post-discussion we discovered that the subjects often confused scenarios with user stories. People occasionally found themselves writing new user stories rather than acceptance criteria. This confusion did not necessarily entail any negative consequences. Scenarios required in-depth investigation and analysis of a certain user story and therefore, new requirements could be discovered. Nonetheless, after first attempts with the scenario template most participants became confident to apply this technique.

7.3.2 INVEST & KANO

The findings from the study revealed that user story and scenario production was distinctively longer lasting in the gamification group. In addition to increased quantity, the intrinsic and extrinsic quality of expressed requirements were better too. Intrinsic quality was measured by five Scrum experts with the INVEST model (Wake, 2003). The key difference was found in the two characteristics "estimable" and "small". Due to the fact that the treatment group had written more requirements, their level of detail was much more specific and concrete. On the contrary, the comparable group primarily expressed their requirements on a much higher level of abstraction. These abstract requirements were mainly embodying epics rather than user stories.

Kano (Kano et al., 1984) was used to weight the requirements based on stakeholder preferences. The returned questionnaires from 13 future end users showed that nearly three quarters of all user stories from the control group were classified as must-be requirements. The last quarter was compound with attractive and indifferent requirements. In sharp contrast, the user story sample of the treatment group contained equal amount of must-be requirements. However, nearly halve of their requirements were categorized as attractive requirements. Implementing these would cause excitement and higher satisfaction for the end user. A few indifferent user stories were present too, but only very limited.

Both INVEST and Kano have shown to be applicable frameworks to measure intrinsic and extrinsic quality of user stories. All Scrum experts were already familiar with the INVEST model. Kano was also easy to apply, even though a few people found it sometimes tricky to answer the dysfunctional questions. Overall, we think that these assessment tools are helpful to evaluate the quality of user stories.

7.3.3 COGNITIVE THEORY OF IDEA GENERATION

Competition had a positive effect on production and quality, but also creativity was noticeably higher rated in the treatment group. The intensive discussion in the control group probably had absorbed people's attention, whereas the gamification group was primarily engaged in the production of new requirements. The integrated rewarding system and leaderboard presumably pushed these people to be more productive. In an early phase where new requirements have to be elicitated from scratch, it might be desirable to keep collaboration at a minimum level. We support this recommendation with the cognitive theory of idea generation.

In the cognitive theory of idea generation (SIAM: Search for Ideas in Associative Memory) Stroebe & Nijstad (2004) provided sufficient evidence that creativity in brainstorming groups can decrease under certain circumstances. According to their theory, ideas are produced in two independent phases. In the first phase problem-relevant knowledge is being activated, followed by idea-generation in a second phase. Hereby, both phases can be negatively affected by a phenomenon known as production blocking (Diehl & Stroebe, 1987). Production blocking is an inevitable consequence of waiting periods caused by verbal exchange between group members. In their hypothesis they state that long waiting periods can interrupt the train of thought and therefore, inhibit the generation of new ideas.

To prevent production blocking the authors proposed methods such as "brainwriting" and "electronic brainstorming". These methods allow ideas to be written down and shared without any verbal communication. Over many years, experiments revealed that individual brainstorming reduced the effect of production blocking on group performance (Aiken, Krosp, Shirani, & Martin, 1994; Gallupe et al., 1992; Lamm & Trommsdorff, 1973; Lewis, Sadosky, & Connolly, 1975).

In summary, we assume that the gamification group dived into a much deeper thinking phase and were more open minded, whereas the control group's train of thoughts were inhibited by the intense discussion and criticism.

7.4 HYPOTHESES EVALUATION

Based on the quantitative findings and inferential statistics from previous chapter, it was possible to determine if our given hypotheses were true or false. The two hypothesis were stated as follow:

- ***H1*:** "If a diversified gamification RE platform aligned with stakeholder motivation is deployed, then user engagement is significantly increased."
- ***H2***: If stakeholders are more engaged in requirements engineering with respect to their expertise, then the overall performance of the process and outcomes is significantly increased.

Figure 7-1 illustrates our hypothetical framework that was used to test the gamification theory. The red colored concepts represent dimensions that did not report any statistical differences between the two groups. The blue boxes represent the concepts that did report a strong difference.

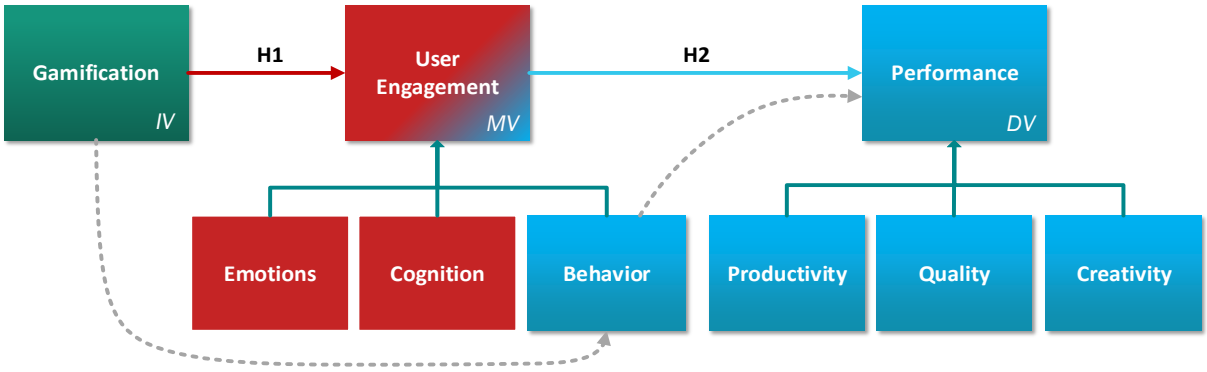

FIGURE 7-1. CONCEPTUAL FRAMEWORK

7.4.1 HYPOTHESIS 1

Both groups reported a similar level of positive emotions that was measured with the PA from the PANAS questionnaire (Watson et al., 1988). Cognition was measured with the FSS questionnaire (Rheinberg, 1987) and did also not expose any variation between the two groups. Although no statistical difference could be found between these two sub-concepts, the measured behavior clearly did, but in two completely different ways.

Participants interacting with the gamification prototype significantly caused more page visits (M = 161.0, SD = 40.367) than did those from the control group (M = 88.833, SD = 38.338), $t(9) = 3.036$, $p < 0.05$. In contrast, the control group significantly wrote more text messages (M = 24.167, SD = 1.732) than did the treatment group (M = 1.0, SD = 1.732), $t(9) = -2.65$, $p < 0.05$).

For the first hypothesis it was not possible to reject the null hypothesis, because both groups were equally engaged by the artifact and very actively involved in the requirements elicitation process.

7.4.2 HYPOTHESIS 2

The statistical results that were used to measure our second hypothesis do support our theory derived from the literature. The gathered data from both experimental groups reported significant variations in all sub-dimensions of the performance concept.

First of all, the average number of written user stories within the treatment group was significantly higher (M = 10.0, SD = 2.345) than those of the control group (M = 3.5, SD = 2.258), $t(9) = 4.673$, $p < 0.05$. The number of added scenarios was also higher in the treatment group (M = 13.4, SD = 5.727) than in the control group (M = 3.0, SD = 3.847), $t(9) = 3.597$, $p < 0.05$.

Next, user stories gathered by the treatment group enabled for better estimations (E) (M = 3.504, SD = 1.177) compared to those of the control group (M = 2.418, SD = 1.213), $t(188) = 5.714$, $p < 0.001$. In addition, user stories of the treatment group were smaller (S) (M = 3.244, SD = 1.187) than those of the control group (M = 2.364, SD = 1.007), t(188) = 4.837, p < 0.001.

Finally, Creativity was significantly higher for the treatment group (M = 3.044, SD = 1.085) compared to the control group (M = 2.236, SD = 0.922), t(188) = 4.853, p < 0.001.

Performance was indirectly impacted by gamification, which caused a change in the behavior dimension. Consequently, our second hypothesis seems to be true and therefore, we rejected the null hypothesis for H2.

7.4.3 REVISED CONCEPTUAL MODEL

To better explain the indirect impact that gamification had on performance, we had to remove some obscurities and describe our hypothetical framework more precisely. In our experiment the implemented competitive game elements caused higher performance in the treatment group, but also diminished collaboration between the group subjects. In requirements elicitation less communication might be an advantage, most notably when the process requires creative thinking. However, in other phases more cooperation might be required for high quality results. Therefore, two new hypothetical models hatched from our initial framework and should be deployed in alignment to the RE tasks and goals.

RE ELICITATION

Brainstorming is quite often used to explore requirements of a system (Nuseibeh & Easterbrook, 2000). Sharing ideas in groups can further stimulate thinking, but verbal communication should be prevented in a divergent phase (Stroebe & Nijstad, 2004). When new requirements must be gathered by stakeholders and creativity is a prerequisite, then we suggest to use the conceptual framework as shown in Figure 7-2. For illustration purpose the concepts "motivation" and "stakeholder expertise" were omitted.

FIGURE 7-2. PROPOSED CONCEPTUAL MODEL FOR RE ELICITATION

This framework describes the term of gamification more precisely by framing a competitive environment, which leads to following two hypotheses:

- **H1.** If a competing gamification RE platform is deployed, then user engagement in RE elicitation is significantly increased.
- **H2.** If stakeholders are more engaged in RE elicitation, then the overall performance of the process and outcomes is significantly increased.

RE ANALYSIS, SPECIFICATION AND VALIDATION

In later stages of the RE process where requirements are converged, more cooperation between the stakeholders is required (Nuseibeh & Easterbrook, 2000). For these phases, an additional framework is proposed, but with the emphasis on social game elements. For instance, a team leaderboard or a team quests that can only be solved in groups could enhance performance. This framework is based on assumptions and not supported with empirical evidence.

This second framework is based on assumptions and its purpose is to provide orientation for future research. Figure 7-3 shows the model for RE analysis, specification and validation, with support of the following hypotheses:

- **H1.** If a social gamification RE platform is deployed, then user engagement in RE analysis, specification and validation is significantly increased.

- **H2.** If stakeholders are more engaged in RE analysis, specification and validation, then the overall performance of the process and outcomes is significantly increased.

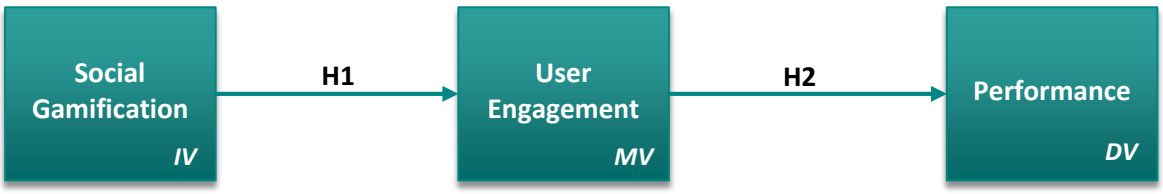

FIGURE 7-3. PROPOSED CONCEPTUAL MODEL FOR RE ANALYSIS, SPECIFICATION AND VALIDATION

7.5 CONCLUSION

This research project demonstrated that gamification has the ability to influence human behavior to achieve a desired output. The success however, highly depends on the choice of game mechanics and game elements, as they can affect different psychological needs. The experimental findings showed that an individual leaderboard can incentivize competition and turn team players into rivals. The group who was exposed to the gamified prototype mainly produced user stories and scenarios on its own behalf. Nonetheless, a competitive environment led to an increased requirements production in the treatment group, but the measured level of quality and creativity was higher too.

In contrast, most user stories gathered by the control group basically covered the expected requirements (must-be). The measured level of creativity and quality was considerably low compared to the treatment group. At the same time this group was collaborating to a greater degree and working as a team towards the predefined experimental objective. Their behavior was apparently not influenced by any virtual competition.

The following facts summarize the most important findings between the two experimental groups:

- The treatment group was primarily working for themselves.
- The treatment group produced more user stories.
- The treatment group produced more scenarios.
- The treatment group delivered higher quality results.
- The treatment group delivered more creative ideas
- The control group was well organized.
- The control group worked as a team.
- The control group conducted an intensive discussion.
- The control group was very inhibited by the discussion.
- Most user stories from the control group were also present in the treatment group.
- User Stories from the control group were written on a much higher level of abstraction.
- The user story template and personas were easy to apply in both groups.
- The scenario template was more difficult to apply and often caused some confusion.
- The overall satisfaction with the gamified prototype was higher than the non-gamified prototype, but not significantly.

Simulating competition can be supportive in requirements elicitation where creative thinking is required to collect basic, but also innovative requirements. We have seen that too much collaboration can cause production blocking and therefore, inhibit creativity (Stroebe & Nijstad, 2004). In later stages of the requirements engineering process more cooperation might be required to analyze, specify and validate the collected requirements (Sommerville, 2010). Hereby, more social game

mechanics and elements, such as team leaderboards or team challenges are possible means to stimulate cooperation between players within the same group (Werbach & Hunter, 2012). A further approach is to distribute different roles to players who then must work together as a team. For example, Fernandes et al. (2012) demonstrated in their research how the six hats of thinking can be used to enhance collaboration and participation in RE workshops.

From our literature study we know that gamification is more than PBL (Chou, 2015), and with our experiment we were able to add some empirical evidence to this claim. We are not stating that PBL is unfavorable, but should rather be complementary to intrinsically motivating game elements that satisfy the three psychological needs from SDT (E. L. Deci & Ryan, 2010). In more simple terms, "in a good game, progress metrics and rewards are part of the package - not the core experience. Well-crafted games are an artful blend of intrinsic pleasure and extrinsic scaffolding" (Jo Kim, 2014).

In summary, we conclude that the selection of game mechanics and game elements for RE highly depends on the engineering activities and desired performance results. Individual leaderboards are a favorable mean when requirements need to be gathered in an initial phase and creative thinking is demanded. In later stages of the RE process it might be wise to implement more cooperative and social game mechanics that facilitate team work. Furthermore, the focus of gamification should be less on external rewards, but rather on strategies to satisfy intrinsic motivation such as autonomy, competence and relatedness (E. L. Deci & Ryan, 2010).

7.6 RESEARCH QUESTIONS ANSWERED

The aim of this research was to find answers to our main research question (MRQ), which was stated as follow:

MRQ: **How can gamification be applied to scenario-based requirements engineering in order to increase stakeholder engagement and improve the quality of requirements engineering?**

In order to answer this main question, four sub-questions were formulated each addressing a different problem area. Based on the literature study and results derived from the conducted experiment, we were able to answer them with following statements.

RQ 1: *What are the common methods, techniques and tools in scenario-based requirements engineering?*

Several techniques and methods were encountered during the literature review to describe and model scenarios. Use cases are very popular in requirements engineering, either documented in textual form or modelled with UML diagrams. In agile methodologies, user stories combined with acceptance criteria are very common to express stakeholders' requirements. The user story (Cohn, 2005) and scenario template from BDD (North, 2006) provide a textual structure to document requirements and can greatly benefit any aspect of the development process.

RQ 2: *What is the potential and the benefits of gamification in the context of requirements engineering?*

Gamification has the potential to drive user behavior and influence user engagement. *iThink* was one of the first gamified approaches in requirements engineering to improve collaboration and participation with gamification (Fernandes et al., 2012). The researchers demonstrated in two case studies that it was possible to discover new requirements and discus existing ones with the aid of a playful experience for stakeholders.

In our experiment, we observed that a competitive gamification environment can lead to more creative requirements. These creative requirements were also more specific formulated, which resulted in a higher level of quality.

RQ 3: *Which activities of scenario-based requirements engineering can be effectively gamified?*

Essentially, all requirements engineering activities can be gamified, when the underlying problems are related to human behavior. For instance, gathering new requirements, prioritizing and estimating requirements, specifying requirements and validating requirements can all be supported by various types of mechanics known from classical video games.

However, the selection of game mechanics and game elements depend on the underlying RE activities. Requirements elicitation can contain more competitive game elements, while the other phases require more social gamification.

RQ 4: *What game elements do positively support scenario-based requirements engineering?*

Building a virtual competition with points, badges, leaderboards and levels can cause people to perform better in requirements elicitation. Social game elements like a chat or team challenges can be supportive in other RE phases where cooperation is demanded. Furthermore, intrinsic motivating game elements that do not reward players for their activities were most enjoyed. These are for instance, storytelling, onboarding, videos and challenges.

RQ 5: *How can the quality of scenario-based RE be measured?*

The intrinsic quality of user stories was measured with the INVEST approach (Wake, 2003). INVEST allows to determine the characteristics of a good quality user story and is often used in a Scrum backlog. Kano (Kano et al., 1984) has shown to be a good tool to measure customer satisfaction of requirements from an end user perspective. Originally developed as a marketing tool, the Kano model allows to classify customer preferences into different categories.

Furthermore, creativity can have a direct impact on the quality of requirements. Requirements that contain novel ideas can lead to higher customer satisfaction and therefore, improve the overall product quality.

RQ 6: *How effective is gamified scenario-based requirements engineering in practice?*

The design of our experiment demonstrated that a gameful experience in requirements elicitation is able to effectively influence users' behavior. Not only did the gamification group produce more unique requirements, but also their quality and creativity was much higher compared to the control group. Interpersonal communication between subject members in the gamification group was hardly present by reason of the implemented competitive game mechanics.

7.7 LIMITATIONS

The small sample size and lack of probability sampling prevented us from making generalizations about the population we studied. The feedback from the pilot phase had shown that it can get confusing to keep track of user stories when too many people are involved. This prompted us to keep the group sizes rather small. Moreover, because the research project was conducted within a software

engineering company, we were bound to the available resources. Interested employees that were willing to invest their time on voluntary basis were asked to take part in the experiment.

Next, the total game time of two hours in the experiment inhibited us to draw conclusion about the long-term effect of gamification. Extrinsic rewards have shown to be effective in the short-term, but their long-term consequences remain unknown. Aside from that, the composition of our gamified system was evaluated as a holistic framework, making it impossible to predict the impact size of the individual game elements on players' behavior.

Finally, we were aware that many of the participants' Reiss profile results had mean values around the center. This means that the basic desires depend upon the context of a particular situation and would have required an in-depth analysis in a face-to-face meeting with the participants (Reiss & Havercamp, 1998).

7.8 FUTURE RESEARCH

In the context of our work, we recommend future research to run several more experiments to better generalize the results to the requirements engineering discipline. 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 (Stroebe & Nijstad, 2004). Next, 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 influential impacts on motivation and behavior. Mekler et al. (2013a) have already opened a door in this direction by conducting experiments with points, levels and leaderboards. Furthermore, we propose to test the gamification platform in a software engineering project over several days or weeks to detect long-term trends with respect to user engagement.

Emotions and cognition should also be measured at different stages in the experiment. For instance, the FSS and PANAS questionnaires should be answered after the onboarding program, then at half time and once again at the end of the experiment. This would allow to show a trend line of how players are emotionally and cognitively engaged over time.

7.8.1 GAMIFICATION IS MORE THAN JUST PBL

PBLs are the most visible and tangible game elements in gamification, but they are only just the tip of the iceberg (Werbach & Hunter, 2012). The evaluation of our selected game elements has shown that the gamified solution should focus more on intrinsically motivating game mechanics. From the results we saw that the storyline was most appreciated by the players. An opportunity to increase intrinsic motivation could be supported by an unfolding storyline where new features are unlocked. For instance, after the onboarding program, the player might only have one persona available to write user stories and scenarios. However, after reaching a certain level or solving a specific quest, the storyline continues and new personas and quests are unlocked.

An unfolding storyline with unlocking features has the potential to provoke curiosity and increase people's status level by providing access to certain features (Zichermann & Cunningham, 2011). Moreover, the level of flow (Csikszentmihalyi, 1991) could be increased by gradually adding harder quests as the narrative background story proceeds.

7.8.2 END-TO-END GAMIFIED APPROACH

Next to RE elicitation it would be valuable to gamify other phases of requirements engineering and therefore, provide new scientific input. The user story game, as we developed it, could be further enhanced and automatically integrate a Kanban system (Sugimori, Kusunoki, Cho, & Uchikawa, 1977). As a result, new user stories would be automatically added to the product backlog. A gamification layer could then be stacked on top of project management activities with the aim to increase motivation within the development team. For instance, burn down charts could be used to show remaining work (Cohn, 2005), similar to a progress bar. Online planning poker and user story points could be used to estimate implementation efforts (Grenning, 2002).

On the stakeholder frontend side, the like function could be exchanged with the Kano questionnaire. Instead of hitting the like button when a user favors a requirement, she or he can answer paired questions with two dropdowns lists. The user's or customer's answers would then be synchronized with the product backlog and automatically prioritize the requirements.

Future product releases should consider to directly listen to the voice of the customer (VoC) and actively include the "crowd" in the requirements engineering process (Snijders et al., 2014). Project progression in the backend could then be displayed to stakeholders in the frontend. Furthermore, a loyalty program built around the platform could reward stakeholders for their submitted requirements after a successful product release. This would result in a comprehensive end-to-end gamified solution for requirements engineering, software engineering and project management.

7.8.3 MOTIVATIONAL DESIGN

In near future we will probably see many advancements on the topic of gamification within economics and behavioral science. Kuo (2013) already stated in his blog post on Gamification.co that the next wave of gamification will not only include game elements, but actual game plays, complemented with web 2.0 elements to solve real business problems.

Jesse Schell even mentioned in one of his talk at Dice Summit that "if companies who make things want to increase people's use of these things, the trick isn't to add more points and bonuses and badges for the sake of adding them, but to find the context of the psychological needs that can be met". In other words, instead of gamification, companies should focus on the motivational design of the experience itself (LeFebvre, 2011).

8 REFERENCES

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, *37*(4), 445–456.

Abreu, F. B., & Carapuça, R. (1994). Object-oriented software engineering: Measuring and controlling the development process. In *Proceedings of the 4th international conference on software quality* (Vol. 186).

Adrion, W. R., Branstad, M. A., & Cherniavsky, J. C. (1982). Validation, verification, and testing of computer software. *ACM Computing Surveys (CSUR)*, *14*(2), 159–192.

Agnarson, A. (2015). Hueman. Retrieved March 2, 2015, from http://alxmedia.se/themes/hueman/

Aiken, M., Krosp, J., Shirani, A., & Martin, J. (1994). Electronic brainstorming in small and large groups. *Information & Management*, *27*(3), 141–149.

Albrecht, A. J. (1979). Measuring application development productivity. In *Proceedings of the Joint SHARE/GUIDE/IBM Application Development Symposium* (Vol. 10, pp. 83–92).

Alexander, I. F., & Maiden, N. (2005). *Scenarios, stories, use cases: through the systems development life-cycle*. John Wiley & Sons.

Alterio, M., & McDrury, J. (2003). *Learning through storytelling in higher education: Using reflection and experience to improve learning*. Routledge.

Amazon. (2014). Retrieved from http://www.amazon.de/

Ambler, S. W. (2014). The Non-Existent Software Crisis: Debunking the Chaos Report. Retrieved from http://www.drdobbs.com/architecture-and-design/the-non-existent-software-crisis-debunki/240165910

Andrew, H. (1998). Demographic Methods. *London, Arnold*, 329.

Armour, P. G. (2001). The laws of software process-Getting from here to there: Putting a boundary around process. ASSOC COMPUTING MACHINERY 1515 BROADWAY, NEW YORK, NY 10036 USA.

Attfield, S., Kazai, G., Lalmas, M., & Piwowarski, B. (2011). Towards a science of user engagement (Position Paper). In *WSDM Workshop on User Modelling for Web Applications*.

Attridge, M. (2009). Measuring and managing employee work engagement: A review of the research and business literature. *Journal of Workplace Behavioral Health*, *24*(4), 383–398.

Barbacci, M., Klein, M. H., Longstaff, T. A., & Weinstock, C. B. (1995). *Quality Attributes.*

Barki, H., & Hartwick, J. (1989). Rethinking the concept of user involvement. *MIS Quarterly*, 53–63.

Baroudi, J. J., Olson, M. H., & Ives, B. (1986). An empirical study of the impact of user involvement on system usage and information satisfaction. *Communications of the ACM*, *29*(3), 232–238.

Bartle, R. (1996). Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD Research*, *1*(1), 19.

Basili, V. R., & Rombach, H. D. (1988). The TAME project: Towards improvement-oriented software environments. *Software Engineering, IEEE Transactions on*, *14*(6), 758–773.

Bass, L. (2007). *Software architecture in practice*. Pearson Education India.

Bateman, C. (2006). Mathematics of XP. Retrieved April 23, 2015, from http://onlyagame.typepad.com/only_a_game/2006/08/mathematics_of_.html

Bauer, T. N., & Erdogan, B. (2011). Organizational socialization: The effective onboarding of new employees.

Beail, N. (1985). *Repertory grid technique and personal constructs: Applications in clinical & educational settings*. Routledge TS - BibTeX.

Beardsley, M. C. (1982). The aesthetic point of view: Selected essays.

Beck, K. (2000). *Extreme programming explained: embrace change*. Addison-Wesley Professional TS - BibTeX.

Beck, K. (2003). *Test-driven development: by example*. Addison-Wesley Professional.

Becker, H. A. (1983). The role of gaming and simulation in scenario project. *Stahl, Ed., Operational Gaming: An International Approach, International Institute for Applied Systems Analysis, Laxenburg, Austria*, 187–202.

Becker, S. A., & Berkemeyer, A. (2002). Rapid application design and testing of web usability. *IEEE Multimedia*, *9*(4), 38–46.

Bell, M. W. (2008). Toward a definition of "virtual worlds." *Journal For Virtual Worlds Research*, *1*(1).

Benabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The Review of Economic Studies*, *70*(3), 489–520.

Be-novative. (2015). Retrieved from http://be-novative.com/

Berander, P. (2007). Evolving prioritization for software product management.

Berger, C., Blauth, R., Boger, D., Bolster, C., Burchill, G., DuMouchel, W., … others. (1993). Kano's methods for understanding customer-defined quality. *Center for Quality Management Journal*, *2*(4), 3–35.

Berry, D., Gacitua, R., Sawyer, P., & Tjong, S. F. (2012). The case for dumb requirements engineering tools. In *Requirements Engineering: Foundation for Software Quality* (pp. 211–217). Springer.

Bhattacherjee, A. (2012). Social science research: principles, methods, and practices.

Bittner, K. (2002). *Use case modeling*. Addison-Wesley Longman Publishing Co., Inc.

Boehm, B. W. (1981). *Software engineering economics*. Prentice-Hall: Englewood Cliffs.

Boehm, B. W. (1984). Verifying and validating software requirements and design specifications. In *IEEE software*.

Boehm, B. W., Brown, J. R., & Kaspar, H. (1978). Characteristics of software quality.

Boehm, B. W., & Ross, R. (1989). Theory-W software project management principles and examples. *Software Engineering, IEEE Transactions on*, *15*(7), 902–916.

Booch, G., Jacobson, I., & Rumbaugh, J. (2000). OMG unified modeling language specification. *Object Management Group Ed: Object Management Group*, *1034*.

Booch, G., Rumbaugh, J., & Jacobson, I. (1999). *The unified modeling language user guide*. Pearson Education India TS - BibTeX.

Boulila, N., Hoffmann, A., & Herrmann, A. (2011). Using Storytelling to record requirements: Elements for an effective requirements elicitation approach. In *Multimedia and Enjoyable Requirements Engineering-Beyond Mere Descriptions and with More Fun and Games (MERE), 2011 Fourth International Workshop on* (pp. 9–16).

Boulos, M. N. K., Hetherington, L., & Wheeler, S. (2007). Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. *Health Information & Libraries Journal*, *24*(4), 233–245.

Breitman, K., & Leite, J. (2002). Managing user stories. In *International Workshop on Time-Constrained Requirements Engineering* (p. 168).

Brennan, K. (2009). *A guide to the business analysis body of knowledge (Babok Guide)*. International Institute of business analysis.

Brickey, J., Walczak, S., & Burgess, T. (2012). Comparing semi-automated clustering methods for persona development. *Software Engineering, IEEE Transactions on*, *38*(3), 537–546.

Brown, D. M. (2010). *Communicating design: developing web site documentation for design and planning*. New Riders.

Brown, J., Lindgaard, G., & Biddle, R. (2008). Stories, sketches, and lists: Developers and interaction designers interacting through artefacts. In *Agile, 2008. AGILE'08. Conference* (pp. 39–50).

Burke, B. (2014). *Redefine Gamification to Understand Its Opportunities and Limitations*. Retrieved from https://www.gartner.com/doc/2699119/redefine-gamification-understand-opportunities-limitations

Byrne, S. (2014). A massive success: 10 years of World of Warcraft (Q&A). Retrieved from http://www.cnet.com/news/world-of-warcraft-ten-year-anniversary-interview-ion-hazzikostas/

Caillos, R. (1958). Les jeux et les hommes. *Le Masque et Le Vertige*.

Campbell, J. P. (2008). Modeling the performance prediction problem in industrial and organizational psychology. *Key Theoretical Issues and Concerns in ESPM*. Sage.

Captain Up. (2015). Retrieved March 2, 2015, from https://captainup.com/

Carroll, E. (2005). Estimating software based on use case points. In *Companion to the 20th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications* (pp. 257–265 TS – BibTeX).

Carroll, J. (2000). Five reasons for scenario-based design. *Interacting with Computers*, *13*(1), 43–60.

Chapman, P. M. (1997). *Models of engagement: Intrinsically motivated interaction with multimedia learning software*. University of Waterloo.

Charette, R. N. (2005). Why Software Fails [Software Failure]. *IEEE Spectr*, *42*(9), 42–49. doi:10.1109/MSPEC.2005.1502528

Chen, J. (2007). Flow in games (and everything else). *Communications of the ACM*, *50*(4), 31–34.

Cheng, R., & Vassileva, J. (2006). Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modeling and User-Adapted Interaction*, *16*(3-4), 321–348.

Chidamber, S. R., & Kemerer, C. F. (1994). A metrics suite for object oriented design. *Software Engineering, IEEE Transactions on*, *20*(6), 476–493.

Chou, Y.-K. (2015). Points, Badges, and Leaderboards: The Gamification Fallacy. Retrieved February 11, 2015, from http://www.yukaichou.com/gamification-study/points-badges-and-leaderboards-the-gamification-fallacy/#.VNs9rfmG_UU

Chris, M. (2013). User Journeys – The Beginner's Guide. Retrieved February 19, 2015, from http://theuxreview.co.uk/user-journeys-beginners-guide/

Christel, M. G., & Kang, K. C. (1992). *Issues in requirements elicitation*.

Chung, L., Nixon, B., Yu, E., & Mylopoulos, J. (2000). Non-functional requirements. *Software Engineering*.

Cimperman, R. (2006). *Uat defined: a guide to practical user acceptance testing*. Addison-Wesley Professional.

CMS Usage Statistics. (2015). Retrieved March 2, 2015, from http://trends.builtwith.com/cms

Cockburn, A. (1997). Using goal-based use cases. *Journal of Object-Oriented Programming*, *10*(7), 56.

Cohn, M. (2004). *User stories applied: For agile software development*. Addison-Wesley Professional.

Cohn, M. (2005). *Agile estimating and planning*. Pearson Education.

Colbert, A. E., Mount, M. K., Harter, J. K., Witt, L. A., & Barrick, M. R. (2004). Interactive effects of personality and perceptions of the work situation on workplace deviance. *Journal of Applied Psychology*, *89*(4), 599.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes.

Coon, D., & Mitterer, J. (2008). *Introduction to psychology: Gateways to mind and behavior*. Cengage Learning.

Cooper, A. (1999). *The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]* (Vol. 261). Sams Indianapolis.

Csikszentmihalyi, M. (1991). *Flow: The psychology of optimal experience* (Vol. 41). HarperPerennial New York.

Damian, D. (2007). Stakeholders in global requirements engineering: Lessons learned from practice. *Software, IEEE*, *24*(2), 21–27.

Davis, A. M. (1993). *Software requirements: objects, functions, and states*. Prentice-Hall, Inc TS - BibTeX.

De Bono, E. (1999). *Six thinking hats*. Taylor & Francis.

Deci, E. (1975). L.(1975) Intrinsic motivation. *New York, London*.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627.

Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. *Advances in Experimental Social Psychology*, *13*, 39–80.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality*, *19*(2), 109–134.

Deci, E. L., & Ryan, R. M. (2010). *Self-Determination*. Wiley Online Library.

Desanctis, G., & Gallupe, R. B. (1987). A foundation for the study of group decision support systems. *Management Science*, *33*(5), 589–609.

Deterding, S. (2012). Gamification: designing for motivation. *Interactions*, *19*(4), 14–17.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15).

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428).

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, *53*(3), 497.

Do Prado Leite, J. C. S., Doorn, J. H., Hadad, G. D. S., & Kaplan, G. N. (2005). Scenario inspections. *Requirements Engineering*, *10*(1), 1–21.

Dobing, B., & Parsons, J. (2006). How UML is used. *Communications of the ACM*, *49*(5), 109–113.

Dorfman, M. (1997). *System and software requirements engineering*. *IEEE Computer Society Press Tutorial* (2nd ed.). IEEE Computer Society Press.

Dreger, J. B. (1989). *Function point analysis*. Prentice-Hall, Inc TS - BibTeX.

Dubois, D. J., & Tamburrelli, G. (2013). Understanding gamification mechanisms for software development. In *Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering* (pp. 659–662).

El Emam, K., & Madhavji, N. H. (1995). Measuring the success of requirements engineering processes. In *Requirements Engineering, 1995., Proceedings of the Second IEEE International Symposium on* (pp. 204–211).

Ellis, C. M., & Sorensen, A. (2007). Assessing employee engagement: the key to improving productivity. *Perspectives*, *15*(1), 1–9.

Engeser, S., & Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, *32*(3), 158–172.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis*. MIT-press TS - BibTeX.

ESA. (2014). Essential Facts about the Computer and Video Game Industry. Retrieved from http://www.theesa.com/facts/pdfs/esa_ef_2014.pdf

Fabbrini, F., Fusani, M., Gnesi, S., & Lami, G. (2001). The linguistic approach to the natural language requirements quality: benefit of the use of an automatic tool. In *Software Engineering Workshop, 2001. Proceedings. 26th Annual NASA Goddard* (pp. 97–105).

Facts and Figures. (2015). Retrieved April 1, 2015, from http://www.maibornwolff.de/en/facts-and-figures

Fagan, M. E. (2001). Design and code inspections to reduce errors in program development. In *Pioneers and Their Contributions to Software Engineering* (pp. 301–334). Springer.

Fernandes, J., Duarte, D., Ribeiro, C., Farinha, C., Pereira, J. M., & Silva, M. M. da. (2012). iThink: A Game-Based Approach Towards Improving Collaboration and Participation in Requirement Elicitation. *Procedia Computer Science*, *15*, 66–77. doi:10.1016/j.procs.2012.10.059

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement.

Finkel, I. L. (2007). *Ancient board games in perspective: papers from the 1990 British Museum colloquium, with additional contributions*. British Museum Pubns Ltd.

Firesmith, D. (2005). Are your requirements complete? *Journal of Object Technology*, *4*(1), 27–44.

Fisher, R. A. (1936). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Fluegge, E. R. (2008). *Who put the fun in functional? Fun at work and its effects on job performance*. University of Florida.

Fong, D. (1996). Using the self-stated importance questionnaire to interpret Kano questionnaire results. *The Center for Quality Management Journal*, *5*(3), 21 – 24.

Galitz, W. O. (2007). *The essential guide to user interface design: an introduction to GUI design principles and techniques*. John Wiley & Sons.

Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, *35*(2), 350–369.

Game Mechanics. (2014). Retrieved February 11, 2015, from http://badgeville.com/wiki/Game_Mechanics

Garrett, J. J. (2010). *Elements of User Experience, The: User-Centered Design for the Web and Beyond*. Pearson Education.

Gartner Inc. (2012). Gartner Says By 2015, More Than 50 Percent of Organizations That Manage Innovation Processes Will Gamify Those Processes. Retrieved September 14, 2014, from http://www.gartner.com/newsroom/id/1629214

Glinz, M. (2000). Improving the quality of requirements with scenarios. *Proceedings of the Second World Congress on Software Quality*.

Goguen, J. A., & Linde, C. (1993). Techniques for requirements elicitation. *RE*, *93*, 152–164.

Granic, I., Lobel, A., & Engels, R. C. M. E. (2013). The benefits of playing video games.

Gravetter, F., & Wallnau, L. (2013). *Essentials of statistics for the behavioral sciences*. Cengage Learning.

Gray, D., Brown, S., & Macanufo, J. (2010). *Gamestorming: A playbook for innovators, rulebreakers, and changemakers*. " O'Reilly Media, Inc."

Greist, J. H., Klein, M. H., Eischens, R. R., Faris, J., Gurman, A. S., & Morgan, W. P. (1979). Running as treatment for depression. *Comprehensive Psychiatry*, *20*(1), 41–54.

Grenning, J. (2002). Planning poker or how to avoid analysis paralysis while release planning. *Hawthorn Woods: Renaissance Software Consulting*, *3*.

Griffiths, M. D. (2008). Internet and video-game addiction. *Adolescent Addiction: Epidemiology, Assessment and Treatment*, 231–267.

Gunter, C. A., Gunter, E. L., Jackson, M., & Zave, P. (2000). A reference model for requirements and specifications. In *Requirements engineering, 2000. Proceedings. 4th International Conference on* (p. 189).

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? A Literature Review of Empirical Studies on Gamification. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 3025–3034).

Harel, D. (1987). Statecharts: A visual formalism for complex systems. *Science of Computer Programming*, *8*(3), 231–274.

Harris, M. A., & Weistroffer, H. R. (2008). Does User Participation Lead to System Success? In *Proceedings of the Southern Association for Information Systems Conference, Richmond, VA, USA*.

Harter, J. K., Schmidt, F. L., Agrawal, S., & Plowman, S. K. (2013). The Relationship Between Engagement At Work And Organizational Ooutcomes. *Q12 Meta-Analysis*.

Harter, J. K., Schmidt, F. L., & Hayes, T. L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: a meta-analysis. *Journal of Applied Psychology*, *87*(2), 268.

Harter, J. K., Schmidt, F. L., & Keyes, C. L. M. (2003). Well-being in the workplace and its relationship to business outcomes: A review of the Gallup studies. *Flourishing: Positive Psychology and the Life Well-Lived*, *2*, 205–224.

Haven, B., Vittal, S., Spivey, O. C., Favier, J., & Cokeh, E. (2008). Measuring Engagement: Four Steps To Making Engagement Measurement A Reality. *Forrester*.

Havercamp, S. M., & Steven Reiss, S. (2003). A comprehensive assessment of human strivings: Test-retest reliability and validity of the Reiss profile. *Journal of Personality Assessment*, *81*(2), 123–132.

Heaton, R. K. (1993). Wisconsin card sorting test: computer version 2. *Odessa: Psychological Assessment Resources*.

Hecker, C. (2010). Achievements considered harmful. *GDC 2010*.

Heller, S. (2001). BACK TALK. *PRINT: AMERICA'S GRAPHIC DESIGN MAGAZINE*, *55*(2).

Herger, M. (2014). *Enterprise Gamification: Engaging people by letting them have fun (Volume 1)* (1st ed.). CreateSpace Independent Publishing Platform.

Herger, M. (2015). Results: Gamification Platform Leader Matrix 2015. Retrieved June 24, 2015, from http://enterprise-gamification.com/index.php?option=com_content&view=article&id=256

Hobbs, J. (2005). An introduction to user journeys. Retrieved July 14, 2015, from http://boxesandarrows.com/an-introduction-to-user-journeys/

Hofmann, H. F., & Lehner, F. (2001). Requirements engineering as a success factor in software projects. *IEEE Software*, *18*(4), 58–66. Retrieved from Hofmann, Lehner 2001 - Requirements engineering as a success.pdf TS - BibTeX

Horowitz, E., Sahni, S., & Rajasekaran, S. (1997). *Computer algorithms C++: C++ and pseudocode versions*. Macmillan TS - BibTeX.

Hotspots Analytics - Heatmaps, User Activity & Custom Event Tracking. (2015). Retrieved April 15, 2015, from https://wordpress.org/plugins/hotspots/

Hueman Theme. (2015). Retrieved February 10, 2015, from https://wordpress.org/themes/hueman/

Hull, E., Jackson, K., & Dick, J. (2005). *Requirements engineering* (Vol. 3). Springer.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (p. 4).

Hutchins, B. (2008). Signs of meta-change in second modernity: the growth of e-sport and the World Cyber Games. *New Media & Society*, *10*(6), 851–869.

ISO/IEC/IEEE. (2011a). Systems and software engineering — Architecture description. *ISO/IEC/IEEE 42010:2011(E)*.

ISO/IEC/IEEE. (2011b). Systems and software engineering — Life cycle processes — Requirements engineering. *ISO/IEC/IEEE 29148:2011(E)*, 1–46.

ISO/IEC/IEEE. (2011c). Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. *ISO/IEC 25010:2011*.

Izard, C. E. (1993). *The Differential Emotions Scale: DES IV-A;[a Method of Measuring the Meaning of Subjective Experience of Discrete Emotions]*. University of Delaware.

Jackson, W. (2013). What Characteristics Make Good Agile Acceptance Criteria? Retrieved from http://www.seguetech.com/blog/2013/03/25/characteristics-good-agile-acceptance-criteria

Jacobson, I. (1992). Object oriented software engineering: a use case driven approach.

Jacobson, I., Booch, G., & Rumbaugh, J. (1999). *The unified software development process* (Vol. 1). Addison-Wesley Reading.

Jacobson, I., Spence, I., & Bittner, K. (2011). *Use-Case 2.0: The Guide to Succeeding with Use Cases*. Ivar Jadobson International. Retrieved from http://www.ivarjacobson.com/Use_Case2.0_ebook/

Jacques, R. (1995). Engagement as a Design Concept for Multimedia. *Canadian Journal of Educational Communication*, *24*(1), 49–59.

Jarke, M., Bui, X. T., & Carroll, J. M. (1998). Scenario management: An interdisciplinary approach. *Requirements Engineering*, *3*(3-4), 155–173.

Jedlitschka, A., Ciolkowski, M., & Pfahl, D. (2008). Reporting experiments in software engineering. In *Guide to advanced empirical software engineering* (pp. 201–228). Springer.

Jeffries, R. (2001). Essential XP: Card, Conversation, Confirmation. *XP Magazine*, *30*.

Jo Kim, A. (2014). Game Design and Extrinsic Motivation. Retrieved June 24, 2015, from http://amyjokim.com/blog/2014/03/09/game-design-and-extrinsic-motivation/

Jung, J. H., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management Science*, *56*(4), 724–742.

Juul, J. (2003). The game, the player, the world: looking for a heart of gameness. In *DIGRA Conf.*

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724.

Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control*, *14*(2), 147–156.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education*. John Wiley & Sons TS - BibTeX.

Kappelman, L. A., McKeeman, R., & Zhang, L. (2006). Early warning signs of IT project failure: The dominant dozen. *Information Systems Management*, *23*(4), 31–36.

Karlsson, J., & Ryan, K. (1997). A cost-value approach for prioritizing requirements. *Software, IEEE*, *14*(5), 67–74.

Karlsson, J., Wohlin, C., & Regnell, B. (1998). An evaluation of methods for prioritizing software requirements. *Information and Software Technology*, *39*(14), 939–947.

Karppinen, P. (2005). Meaningful learning with digital and online videos: Theoretical perspectives. *AACE Journal*, *13*(3), 233–250.

Kearney, T. J. (1990). Frequent Flyer Programs: A Failure in competitive strategy, with lessons for management. *Journal of Consumer Marketing*, *7*(1), 31–40.

Kemppilä, S., & Lönnqvist, A. (2003). Subjective productivity measurement. *The Journal of American Academy of Business*, *2*(2), 531–537.

Kent, S. (2010). *The Ultimate History of Video Games: from Pong to Pokemon and beyond... the story behind the craze that touched our li ves and changed the world*. Random House LLC.

Klopp, E. (2014). Die Kano-Methode.

Koepp, M. J., Gunn, R. N., Lawrence, A. D., Cunningham, V. J., Dagher, A., Jones, T., … Grasby, P. M. (1998). Evidence for striatal dopamine release during a video game. *Nature*, *393*(6682), 266–268.

Kompaso, S. M., & Sridevi, M. S. (2010). Employee engagement: The key to improving performance. *International Journal of Business and Management*, *5*(12), p89.

Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., Schaufeli, W. B., de Vet Henrica, C. W., & van der Beek, A. J. (2011). Conceptual frameworks of individual work performance: a systematic review. *Journal of Occupational and Environmental Medicine*, *53*(8), 856–866.

Kruchten, P. B. (1995). The 4+ 1 view model of architecture. *Software, IEEE*, *12*(6), 42–50.

Krueger, J., & Killham, E. (2007). The Innovation Equation. *Gallup Management Journal*, 2007–2008.

Kuhn, D. L. (1989). Selecting and effectively using a computer aided software engineering tool. Retrieved from http://www.osti.gov/scitech/biblio/5611931

Kujala, S., Kauppinen, M., Lehtola, L., & Kojo, T. (2005). The role of user involvement in requirements quality and project success. *Requirements Engineering, 2005. Proceedings. 13th IEEE International Conference on*.

Kumar, J. (2013). *Gamification at work: designing engaging business software*. Springer.

Kuo, I. (2013). Moving Beyond Points and Badges: Gamification 2.0. Retrieved June 26, 2015, from http://gco.mystagingwebsite.com/2013/04/03/moving-beyond-points-and-badges-gamification-2-0/

Kyng, M. (1995). Creating contexts for design. In *Scenario-based design* (pp. 85–107).

La Guardia, J. G., Ryan, R. M., Couchman, C. E., & Deci, E. L. (2000). Within-person variation in security of attachment: a self-determination theory perspective on attachment, need fulfillment, and well-being. *Journal of Personality and Social Psychology*, *79*(3), 367.

Laitenberger, O., & DeBaud, J.-M. (2000). An encompassing life cycle centric survey of software inspection. *Journal of Systems and Software*, *50*(1), 5–31.

Lamm, H., & Trommsdorff, G. (1973). Group versus individual performance on tasks requiring ideational proficiency (brainstorming): A review.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications.

Lapouchnian, A. (2005). Goal-oriented requirements engineering: An overview of the current research. *University of Toronto*.

Laurel, B. (2013). *Computers as theatre*. Pearson Education.

Lazzaro, N. (2004). Why we play games: Four keys to more emotion without story. In *Game Developers Conference*.

LeFebvre, R. (2011). Game visionary Jesse Schell talks about the pleasure revolution. Retrieved June 26, 2015, from http://venturebeat.com/2011/12/05/jesse-schell-talks-about-the-pleasure-revolution/

Leffingwell, D., & Widrig, D. (2003). *Managing software requirements: a use case approach*. Pearson Education TS - BibTeX.

Lehmann, J., Lalmas, M., Yom-Tov, E., & Dupret, G. (2012). Models of user engagement. In *User Modeling, Adaptation, and Personalization* (pp. 164–175). Springer.

Leon, A. (1991). The respective roles of user participation and user involvement in information system implementation success.

Lerman, K., & Ghosh, R. (2010). Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks. *ICWSM*, *10*, 90–97.

Lewis, A. C., Sadosky, T. L., & Connolly, T. (1975). The effectiveness of group brainstorming in engineering problem solving. *Engineering Management, IEEE Transactions on*, (3), 119–124.

Li, F.-L., Horkoff, J., Mylopoulos, J., Guizzardi, R. S. S., Guizzardi, G., Borgida, A., & Liu, L. (2014). Non-functional Requirements as Qualities, with a Spice of Ontology. In *Requirements Engineering Conference (RE), 2014 IEEE 22nd International* (pp. 293–302).

Li, W., & Henry, S. (1993). Object-oriented metrics that predict maintainability. *Journal of Systems and Software*, *23*(2), 111–122.

Lilyquist, L. (2014). Three Considerations for Business Gamification. Retrieved June 26, 2015, from https://badgeville.com/blog/3-considerations-for-business-gamification

Löfgren, M., & Witell, L. (2005). Kano's theory of attractive quality and packaging. *Quality Management Journal*, *12*(3), 7–20.

Loftus, G. R., & Loftus, E. F. (1983). *Mind at play: The psychology of video games* (Vol. 14). Basic Books New York.

Lucassen, G., Dalpiaz, F., Brinkkemper, S., & van der Werf, J. M. E. M. (2015). Forging High-Quality User Stories: Towards a Discipline for Agile Requirements. In *Proceedings of the IEEE International Requirements Engineering Conference*. Ottawa.

Lui, T.-W., Piccoli, G., & Ives, B. (2007). Marketing strategies in virtual worlds. *ACM SIGMIS Database*, *38*(4), 77–80.

Macey, W. H., & Schneider, B. (2008). The meaning of employee engagement. *Industrial and Organizational Psychology*, *1*(1), 3–30.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266.

Marczewski, A. (2013a). Gamasutra: Andrzej Marczewski's Blog - What's the difference between Gamification and Serious Games? Retrieved September 14, 2014, from http://www.gamasutra.com/blogs/AndrzejMarczewski/20130311/188218/Whats_the_difference_between_Gamification_and_Serious_Games.php

Marczewski, A. (2013b). *Gamification: a simple introduction*. Andrzej Marczewski.

Martin, F. (2003). Use Cases And Stories. Retrieved from http://martinfowler.com/bliki/UseCasesAndStories.html

Martin, J., & Tsai, W. T. (1990). N-fold inspection: A requirements analysis technique. *Communications of the ACM*, *33*(2), 225–232.

Martin, R. C. (1998). UML Tutorial: Sequence Diagrams. *Engineering Notebook Column*, 1–5.

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*(4), 370.

Mason, D. H. (1994). Scenario-based planning: decision model for the learning organization. *Strategy & Leadership*, *22*(2), 6–11.

McCall, J. A., Richards, P. K., & Walters, G. F. (1977). *Factors in software quality*. General Electric, National Technical Information Service.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

McManus, J. (2004). A stakeholder perspective within software engineering projects. *Engineering Management Conference, 2004. Proceedings. 2004 IEEE International*.

Meet the leading web analytics plugin for WordPress. (2015). Retrieved April 8, 2015, from Meet the leading web analytics plugin for WordPress

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013a). Disassembling gamification: the effects of points and meaning on user motivation and performance. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1137–1142).

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013b). Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 66–73).

Miller, T. W. (2005). *Data and text mining: A business applications approach*. Pearson Prentice Hall.

Mogyorodi, G. (2001). Requirements-based testing: an overview. *Technology of Object-Oriented Languages, International Conference on*.

Mora M., & Denger C. (2003). *Requirement metrics. An initial literature survey on measurement approaches for requirement specifications*. Brüssel.

Moretón, J. (2014). Stop the Clock! Time in Games and Gamification. Retrieved March 10, 2015, from http://gecon.es/2014/09/02/paren-el-reloj-el-tiempo-en-los-juegos/

Mullenweg, M., Boren, R., Jaquith, M., Ozz, A., & Westwood, P. (2011). WordPress. Wordpress.

Myers, B. A. (1985). The importance of percent-done progress indicators for computer-human interfaces. In *ACM SIGCHI Bulletin* (Vol. 16, pp. 11–17).

Mylopoulos, J., Chung, L., & Nixon, B. (1992). Representing and using nonfunctional requirements: A process-oriented approach. *Software Engineering, IEEE Transactions on*, *18*(6), 483–497.

Mylopoulos, J., Chung, L., & Yu, E. (1999). From object-oriented to goal-oriented requirements analysis. *Communications of the ACM*, *42*(1), 31–37.

Nah, F. F.-H. (2004). A study on tolerable waiting time: how long are Web users willing to wait? *Behaviour & Information Technology*, *23*(3), 153–163.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. *Handbook of Positive Psychology*, 89–105.

Naveed, M. (2014). The INVEST Scale. Retrieved April 10, 2015, from https://www.scrumalliance.org/community/articles/2014/january/the-invest-scale

Nenonen, S., Rasila, H., Junnonen, J.-M., & Kärnä, S. (2008). Customer Journey--a method to investigate user experience. In *Proceedings of the Euro FM Conference Manchester* (pp. 54–63).

Nestler, E. J., & Carlezon Jr, W. A. (2006). The mesolimbic dopamine reward circuit in depression. *Biological Psychiatry*, *59*(12), 1151–1159. Retrieved from http://www.oxforddictionaries.com/definition/english/goal

Nguyen, D. Q. (1998). The essential skills and attributes of an engineer: A comparative study of academics, industry personnel and engineering students. *Global J. of Engng. Educ*, *2*(1), 65–75.

Nguyen, L., & Shanks, G. (2009). A framework for understanding creativity in requirements engineering. *Information and Software Technology*, *51*(3), 655–662.

Nicholson, J. A., Nicholson, D. B., Coyle, P., Hardin, A., & Krishen, A. S. (2014). Exploring the use of virtual world technology for idea-generation tasks. *International Journal of E-Collaboration (IJeC)*, *10*(3), 44–62.

Nielsen, J. (2006). F-shaped pattern for reading web content. *Alertbox: Current Issues in Web Usability (retrieved 31 Oct 2008 from Http://www. Useit. Com/alertbox/reading_ Pattern. Html)*.

Nike+. (2014). Retrieved from http://www.nike.com/us/en_us/c/running/nikeplus/gps-app

North, D. (2006). Behavior Modification: The evolution of behavior-driven development. *Better Software*, *8*(3).

North, D. (2007). WHAT'S IN A STORY? Retrieved from http://dannorth.net/whats-in-a-story/

Nuseibeh, B., & Easterbrook, S. (2000). Requirements engineering: a roadmap. *Proceedings of the Conference on the Future of Software Engineering*.

O'Brien, H. L., & Toms, E. G. (2008). 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.

Otway, H. (2012). Voki avatars in the classroom. *More than Just Knowing Stuff*.

Ow, S. H., & Yaacob, M. H. (1997). A study on the requirements review process in software development: problems and solutions. In *2013 IEEE 37th Annual Computer Software and Applications Conference* (p. 542).

Parnas, D. L., & Weiss, D. M. (1985). Active design reviews: principles and practices. In *Proceedings of the 8th international conference on Software engineering* (pp. 132–136).

Paulus, P. B., & Yang, H.-C. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior and Human Decision Processes*, *82*(1), 76–87.

Pavlov, I. P., & Gantt, W. (1928). Lectures on conditioned reflexes: Twenty-five years of objective study of the higher nervous activity (behaviour) of animals.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*(3), 45–77.

Piccione, P. A. (1980). *In search of the meaning of Senet*. Archaeological Institute of America.

Pichler, R. (2010). *Agile product management with scrum: Creating products that customers love*. Addison-Wesley Professional.

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

Pisarek, M. (2011). Creating a SharePoint Vision using the Gamestorming Cover Story Technique. Retrieved February 23, 2015, from http://www.sharepointanalysthq.com/2011/12/creating-a-sharepoint-vision-using-the-gamestorming-cover-story-technique/

Podeswa, H. (2009). *UML For The IT Business Analyst* (2nd ed.). Boston: Cengage Learning PTR.

Pohl, K. (2010). *Requirements Engineering: Fundamentals, Principles, and Techniques* (1st TS - .). Springer Publishing Company, Incorporated.

Pohl, K., & Rupp, C. (2011). *Basiswissen Requirements Engineering: Aus-und Weiterbildung zum" Certified Professional for Requirements Engineering"; Foundation Level nach IREB-Standard*. dpunkt-Verlag.

PowToon. (2015). Retrieved March 26, 2015, from http://www.powtoon.com/

Pratt, C. T. (2011). *dopmaine*. *American Heritage® Dictionary of the English Language, Fifth Edition* (5th ed.). American Heritage® Dictionary of the English Language. Retrieved from http://www.thefreedictionary.com/dopamine

PrEmo Tool. (2015). Retrieved March 24, 2015, from http://www.premotool.com/

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology*, *14*(2), 154.

Pujol, N. (2010). Freemium: Attributes of an Emerging Business Model. *SSRN Electronic Journal*. doi:10.2139/ssrn.1718663

Quesenbery, W. (2003). The five dimensions of usability. *Content and Complexity: Information Design in Technical Communication*, 81–102.

Quesenbery, W., & Brooks, K. (2010). *Storytelling for user experience: Crafting stories for better design*. Rosenfeld Media.

Quiñones, M. A., Ford, J. K., & Teachout, M. S. (1995). The relationship between work experience and job performance: a conceptual and meta-analytic review. *Personnel Psychology*, *48*(4), 887–910.

Radatz, J., Geraci, A., & Katki, F. (1990). IEEE standard glossary of software engineering terminology. *IEEE Std*, *610121990*, 121990.

Radoff, J., & Kidhardt, E. (2011). *Game on: energize your business with social media games*. Wiley.

Read, J. C., MacFarlane, S. J., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children* (Vol. 2, pp. 1–23).

Regnell, B., & Brinkkemper, S. (2005). Market-driven requirements engineering for software products. In *Engineering and managing software requirements* (pp. 287–308). Springer.

Reiss Motivation Profile. (2015). Retrieved March 27, 2015, from http://reissprofile.pl/the-reiss-profile/

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, *8*(3), 179.

Reiss, S., & Havercamp, S. M. (1998). Toward a comprehensive assessment of fundamental motivation: Factor structure of the Reiss Profiles. *Psychological Assessment*, *10*(2), 97.

Reynolds, T. J., & Gutman, J. (1988). Laddering theory, method, analysis, and interpretation. *Journal of Advertising Research*, *28*(1), 11–31.

Rheinberg, F. (1987). Fragen zum Erleben von Tätigkeiten. *Ein Fragebogen Zum Erfassen von Flow-Erleben Im Alltag*.

Rheinberg, F., Vollmeyer, R., & Engeser, S. (2003). *Die Erfassung des Flow-Erlebens*. na.

Rivera, J., & der Meulen, R. (2013). Gartner Says Worldwide PC, Tablet and Mobile Phone Combined Shipments to Reach 2.4 Billion Units in 2013. *Tablet*, *116*(197,202), 265–731.

Robertson, M. (2010). Can't play, won't play. Retrieved June 19, 2015, from http://hideandseek.net/2010/10/06/cant-play-wont-play/

Robinson, D., Perryman, S., & Hayday, S. (2004). The drivers of employee engagement. *Report-Institute for Employment Studies*.

Roger, S. P. (2005). Software engineering: a practitioner's approach. *McGrow-Hill International Edition*.

Role. (n.d.). Retrieved February 25, 2015, from http://www.britannica.com/EBchecked/topic/507038/role

Rolland, C., Achour, C. Ben, Cauvet, C., Ralyté, J., Sutcliffe, A., Maiden, N., … Dubois, E. (1998). A proposal for a scenario classification framework. *Requirements Engineering*, *3*(1), 23–47.

Rolland, C., Souveyet, C., & Achour, C. Ben. (1998). Guiding goal modeling using scenarios. *Software Engineering, IEEE Transactions on*, *24*(12), 1055–1071.

Rosser, J. C., Lynch, P. J., Cuddihy, L., Gentile, D. A., Klonsky, J., & Merrell, R. (2007). The impact of video games on training surgeons in the 21st century. *Archives of Surgery*, *142*(2), 181–186.

Runeson, P., & Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, *14*(2), 131–164.

Rupp, C., Simon, M., & Hocker, F. (2009). *Requirements-Engineering und -Management* (5th ed.). Nürnberg: Hanser Fachbuchverlag.

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: a single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology*, *57*(3), 493.

Russell, S., & Creighton, O. (2011). Virtual world tools for Requirements Engineering. In *Multimedia and Enjoyable Requirements Engineering-Beyond Mere Descriptions and with More Fun and Games (MERE), 2011 Fourth International Workshop on* (pp. 17–20).

Ryan, R. M., & Deci, E. (1985). Intrinsic motivation and selfdetermination in human behavior. *New York*.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360.

Ryser, J., & Glinz, M. (1999). A practical approach to validating and testing software systems using scenarios. In *QWE'99, 3 rd International Software Quality Week Europe*.

Saaty, T. L. (1988). *What is the analytic hierarchy process?* Springer TS - BibTeX.

Salamone, J. D., & Correa, M. (2012). The mysterious motivational functions of mesolimbic dopamine. *Neuron*, *76*(3), 470–485.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT press.

Schaufeli, W. B., Salanova, M., González-Romá, V., & Bakker, A. B. (2002). The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *Journal of Happiness Studies*, *3*(1), 71–92.

Scheffe, H. (1999). *The analysis of variance* (Vol. 72 TS - B). John Wiley & Sons.

Schmidt, D. C. (2006). Model-driven engineering. *COMPUTER-IEEE COMPUTER SOCIETY-*, *39*(2), 25.

Schreiber, I. (2009). *Challenges for game designers*. Cengage Learning.

Second Life. (2015). Retrieved February 23, 2015, from http://secondlife.com/

Segal, D. (2014). Behind League of Legends, E-Sports's Main Attraction. Retrieved from http://www.nytimes.com/2014/10/12/technology/riot-games-league-of-legends-main-attraction-esports.html

Sharif, B., Khan, S. A., & Bhatti, M. W. (2012). Measuring the Impact of Changing Requirements on Software Project Cost: An Empirical Investigation. *International Journal of Computer Science Issues (IJCSI)*, *9*(3).

Shull, F., Rus, I., & Basili, V. (2000). How perspective-based reading can improve requirements inspections. *Computer*, *33*(7), 73–79.

Simon, H. A. (1969). *The sciences of the artificial* (Vol. 136 TS -). MIT press.

Smith, M. F. (1991). *Software prototyping: adoption, practice and management*. McGraw-Hill, Inc TS - BibTeX.

Snijders, R., Dalpiaz, F., Brinkkemper, S., Hosseini, M., Ali, R., & Özüm, A. (2015). REfine: A Gamified Platform for Participatory Requirements Engineering. In *1st International Workshop on Crowd-Based Requirements Engineering (CrowdRE 2015)*.

Snijders, R., Dalpiaz, F., Hosseini, M., Shahri, A., Ali, R., & others. (2014). Crowd-Centric Requirements Engineering. In *Proceedings of the Second International Workshop on Crowdsourcing and Gamification in the Cloud (CGCloud 2014)*.

Sommerville, I. (2010). *Software Engineering*. (M. Horton, M. Hirsch, M. Goldstein, & J. Holcom, Eds.) (9th ed.). Boston: Addison-Wesley.

Sommerville, I., & Kotonya, G. (1998). *Requirements engineering: processes and techniques*. John Wiley & Sons, Inc.

SpigitEngage. (2015). Retrieved from http://www.mindjet.ch/spigitengage/

Stedman, T. L. (2011). *Motivation* (7th ed.). Stedman's Medical Dictionary for the Health Professions and Nursing. Retrieved from http://medical-dictionary.thefreedictionary.com/motivation

Stephenson, W. (1964). *The play theory of mass communication*. Transaction Publishers.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. *Handbook of Creativity*, *1*, 3–15.

Stevens, P., & Pooley, R. J. (2006). *Using UML: software engineering with objects and components*. Pearson Education.

Stevenson, A. (Ed.). (2010). *Oxford Dictionary of English - Oxford Reference* (3rd ed.). Oxford: Oxford University Press.

Stroebe, W., & Nijstad, B. A. (2004). Warum Brainstorming in Gruppen Kreativit{ä}t vermindert. *Psychologische Rundschau*, *55*(1), 2–10.

Sugimori, Y., Kusunoki, K., Cho, F., & Uchikawa, S. (1977). Toyota production system and kanban system materialization of just-in-time and respect-for-human system. *The International Journal of Production Research*, *15*(6), 553–564.

Sun, Y., & Kantor, P. B. (2006). Cross-Evaluation: A new model for information system evaluation. *Journal of the American Society for Information Science and Technology*, *57*(5), 614–628.

Sutcliffe, A. (1998). Scenario-based requirements analysis. *Requirements Engineering*, *3*(1), 48–65.

Sutcliffe, A. (2002). *User-centred requirements engineering*. Springer.

Sutcliffe, A. (2003). Scenario-based requirements engineering. *Requirements Engineering Conference, 2003. Proceedings. 11th IEEE International*.

Sutcliffe, A. G., Maiden, N. A. M., Minocha, S., & Manuel, D. (1998). Supporting scenario-based requirements engineering. *Software Engineering, IEEE Transactions on*, *24*(12), 1072–1088.

Sutherland, M., & Maiden, N. (2010). Storyboarding requirements. *Software, IEEE*, *27*(6), 9–11.

Takeda, H., Veerkamp, P., & Yoshikawa, H. (1990). Modeling design process. *AI Magazine*, *11*(4), 37.

Thom, J., Millen, D., & DiMicco, J. (2012). Removing gamification from an enterprise SNS. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1067–1070).

Tichy, W. F. (1998). Should computer scientists experiment more? *Computer*, *31*(5), 32–40.

Treadway, M. T., Buckholtz, J. W., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., … Zald, D. H. (2012). Dopaminergic mechanisms of individual differences in human effort-based decision-making. *The Journal of Neuroscience*, *32*(18), 6170–6176.

Tremblay, M. C., Hevner, A. R., & Berndt, D. J. (2010). Focus groups for artifact refinement and evaluation in design research. *Communications of the Association for Information Systems*, *26*, 1.

Trott, P. (2008). *Innovation management and new product development*. Pearson education.

Turco, K. (2013). It's Not All Fun and Games; Potential Risks of Gamification. Retrieved June 23, 2015, from http://technologyadvice.com/gamification/blog/its-not-all-fun-and-games-potential-risks-of-gamification/

Tychsen, A., Hitchens, M., Brolund, T., & Kavakli, M. (2005). The game master. In *Proceedings of the second Australasian conference on Interactive entertainment* (pp. 215–222).

Uemura, T., Kusumoto, S., & Inoue, K. (1999). Function point measurement tool for UML design specification. In *Software Metrics Symposium, 1999. Proceedings. Sixth International* (pp. 62–69).

Van Aken, J. E. (2005). Management research as a design science: Articulating the research products of mode 2 knowledge production in management. *British Journal of Management*, *16*(1), 19–36.

Van Lamsweerde, A. (2000). Requirements engineering in the year 00: A research perspective. In *Proceedings of the 22nd international conference on Software engineering* (pp. 5–19).

Van Lamsweerde, A. (2001). Goal-oriented requirements engineering: a guided tour. In *Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on* (pp. 249–262). Toronto, Ont.

Van Lamsweerde, A. (2009). Requirements engineering: from system goals to UML models to software specifications.

Van Lamsweerde, A., & Willemet, L. (1998). Inferring declarative requirements specifications from operational scenarios. *Software Engineering, IEEE Transactions on*, *24*(12), 1089–1114.

Venture Spirit. (2015). Retrieved from http://www.venturespirit.com/

Verschuren, P., Doorewaard, H., & Mellion, M. J. (2010). *Designing a research project*. Eleven International Publishing TS - BibTeX.

Viswesvaran, C. (1993). *Modeling job performance: Is there a general factor?*

Viswesvaran, C. (2001). Assessment of individual job performance: A review of the past century and a look ahead. *Handbook of Industrial, Work and Organizational Psychology*, *1*, 110–126.

Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, *8*(4), 216–226.

Von Alan, R. H., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Wake, B. (2003). Invest in Good Stories, and SMART Tasks. Retrieved January 20, 2015, from http://xp123.com/articles/invest-in-good-stories-and-smart-tasks/

Ward, P. T. (1986). *Structured Development for Real-Time Systems: Vol. I: Introduction and Tools*. Pearson Education.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*(2), 219.

Webster, J., & Ahuja, J. S. (2006). Enhancing the design of web navigation systems: the influence of user disorientation on engagement and performance. *Mis Quarterly*, 661–678.

Webster, J., & Ho, H. (1997). Audience engagement in multimedia presentations. *ACM SIGMIS Database*, *28*(2), 63–77.

Weinberg, G. M. (1992). *Quality software management* (Vol. 1). Dorset House Publ.

Wellins, R., & Concelman, J. (2005). Creating a culture for engagement. *Workforce Performance Solutions*, *4*, 1–4.

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Wharton Digital Press.

Werneck, V. M. B., Oliveira, A. de P. A., & do Prado Leite, J. C. S. (2009). Comparing GORE Frameworks: i-star and KAOS. In *WER*.

Whatley, S. (2011). 43 Things That Customers Think Are Fun. Retrieved March 19, 2015, from http://www.simonwhatley.co.uk/43-things-that-customers-think-are-fun

Wiegers, K. (2001a). Improving quality through software inspections. *Software Development*, *3*(4), 1–15.

Wiegers, K. (2001b). When two eyes aren't enough. *Software Development*, *9*(10), 58–61.

Wiegers, K. (2002). *Peer reviews in software: A practical guide*. Addison-Wesley Boston TS - BibTeX.

Wiegers, K., & Beatty, J. (2013). *Software requirements* (3rd ed.). Redmond, Washington: Microsoft Press.

Wieringa, R., & Dubois, E. (1998). Integrating semi-formal and formal software specification techniques. *Information Systems*, *23*(3), 159–178.

Williams, C. (2013). Despite funding, Foursquare faces battle against apathy. Retrieved from http://www.telegraph.co.uk/technology/social-media/9990426/Despite-funding-Foursquare-faces-battle-against-apathy.html

Williams, J., & Ross, S. (2010). *Score One for Quality! Using Games to Improve Product Quality*. Redmond.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer TS - BibTeX.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & Behavior*, *9*(6), 772–775.

Young, R. M., & Barnard, P. (1986). *The use of scenarios in human-computer interaction research: Turbocharging the tortoise of cumulative science* (Vol. 17). ACM.

Yu, E. S. K. (1993). Modeling organizations for information systems requirements engineering. In *Requirements Engineering, 1993., Proceedings of IEEE International Symposium on* (pp. 34–41).

Zahran, S. (1998). *Software process improvement*. Addison-Wesley.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps*. " O'Reilly Media, Inc."

9 APPENDIX

APPENDIX A: GAME ELEMENTS

APPENDIX A1: JON RADOFF'S 42 FUN FACTORS

Motivation / Fun Factor	*Power*	*Curiosity*	*Independence*	*Acceptance*	*Order*	*Saving*	*Honor*	*Idealism*	*Social contact*	*Family*	*Status*	*Vengeance*	*Romance*	*Physical Activity*	*Tranquility*	TOTAL
Recognizing Patterns		x			x											2
Collecting	x				x	x					x					4
Finding Random Treasures	x					x					x					3
Achieving a Sense of Completion	x		x		x										x	4
Gaining Recognition for Achievements				x					x		x					3
Creating Order out of Chaos					x										x	2
Customizing Virtual Worlds			x						x		x					3
Gathering Knowledge		x							x		x					3
Organizing Groups of People					x				x	x	x		x			5
Noting Insider References				x					x							2
Being the Centre of Attention	x										x		x			3
Experiencing Beauty and Culture					x								x		x	3
Romance							x			x			x			3
Exchanging Gifts				x			x			x			x			4
Being a Hero	x		x				x	x	x		x	x	x			8
Being a Villain	x		x									x				3
Being a Wise Old Man	x		x	x			x			x	x					6
Being a Rebel	x		x						x		x	x	x			6
Being the Ruler	x				x		x			x	x		x			6
Pretending to Live in a Magical Place		x						x					x		x	4
Listening to a Story		x							x							2
Telling Stories	x			x							x					3
Predicting the Future					x											1
Competition	x										x	x		x		4
Psychoanalyzing		x			x											2
Mystery		x											x			2

Mastering a Skill	x		x											x		3
Exacting Justice and Revenge	x											x	x			3
Nurturing			x				x			x	x		x			5
Excitement														x		1
Triumph over Conflict	x											x				2
Relaxing															x	1
Experiencing the Freakish or Bizarre		x														1
Being Silly				x											x	2
Laughing														x	x	2
Being Scared														x		1
Strengthening a Family Relationship						x		x	x							3
Improving One's Health														x		1
Imagining a Connection with the Past		x			x											2
Exploring a World		x	x													2
Improving Society							x	x			x					3
Enlightenment		x	x												x	3
TOTAL	14	10	10	6	10	3	7	4	9	6	15	6	12	6	8	

TABLE 9-1. 42 FUN FACTORS (RADOFF & KIDHARDT, 2011)

1. ***Recognizing Patterns*** – Anything from visual patterns, motion patterns, strategic patterns or mathematical patterns.
2. ***Collecting*** – Collections communicate status, suggest organization, lead to rewards, represent wealth and are mementos.
3. ***Finding Random Treasures*** – Like winning a jackpot or slot machine, finding shells at the beach or opening Cracker Jacks to find a surprise.
4. ***Achieving a Sense of Completion*** – Giving players a constant sense of finishing something like progress bars, to-do lists, achievements and levels.
5. ***Gaining Recognition for Achievements*** – Achievement systems provide a sense of accomplishment and a chance to be recognized.
6. ***Creating Order out of Chaos*** – Sorting, lining things up and classifying give players a sense of control over their environment.
7. ***Customizing Virtual Worlds*** – People enjoy leaving their mark and place great value on things they've made.
8. ***Gathering Knowledge*** – Studying and being taught are not fun, but learning is fun because we are naturally curious.
9. ***Organizing Groups of People*** – Organizing groups of people to achieve shared goals is a source of enjoyment.
10. ***Noting Insider References*** – Allows players to be part of and belong to an in-group.
11. ***Being the Centre of Attention*** – Satisfy the human need for attention by putting the player at the center of the universe.

12. ***Experiencing Beauty and Culture*** – Games feature artwork, music and designs that appeal to the human senses.
13. ***Romance*** – Games can provide opportunities for flirting, wooing and building relationships with the opposite sex.
14. ***Exchanging Gifts*** – Players enjoy giving gifts to their friends and the act of giving triggers reciprocity.
15. ***Being a Hero*** – Playing as the hero appeals to the human desire for power.
16. ***Being a Villain*** – It's about the fantasy of having power without consequences.
17. ***Being a Wise Old Man*** – This is typically a high status role that may also touch on the motivator of family.
18. ***Being a Rebel*** – The opportunity to flaunt society's rules while remaining basically good.
19. ***Being the Ruler*** – The chance to be a person with considerable power over other people.
20. ***Pretending to Live in a Magical Place*** – Players enjoy imaging being in worlds different than their own.
21. ***Listening to a Story*** – Stories appeal to our curiosity about people, places and things.
22. ***Telling Stories*** – Games provide an opportunity for players to construct and tell their own unique stories.
23. ***Predicting the Future*** – Predicting the future makes people feel smart, in-control and influential.
24. ***Competition*** – People enjoy the sense of power that comes from winning.
25. ***Psychoanalyzing*** – Predicting, guessing or understanding the motivations of others can be a source of fun.
26. ***Mystery*** – Striking a balance between revealing a little while holding back the rest can create a fun experience.
27. ***Mastering a Skill*** – Increasing one's mastery without becoming frustrated gives people a sense of flow.
28. ***Exacting Justice and Revenge*** – Justice and revenge provide a sense of idealism and tranquility when wrongs are righted.
29. ***Nurturing*** – Growing things stems from your motivations for family, saving and power.
30. ***Excitement*** – Suspense, horror, competitive action and anticipation help create an addictive, exciting experience.
31. ***Triumph over Conflict*** – Resolving conflict provides the player with a sense of victory.
32. ***Relaxing*** – Games can create a mental vacation which can lead to tranquility.
33. ***Experiencing the Freakish or Bizarre*** – People crave new and unique experiences that are different from their everyday lives.
34. ***Being Silly*** – Players enjoy an escape from the serious and mundane.
35. ***Laughing*** – People love to laugh, especially with their friends.
36. ***Being Scared*** – People enjoy the sensation of danger without the actual danger.
37. ***Strengthening a Family Relationship*** – Players enjoy feeling companionship with members of their family.
38. ***Improving One's Health*** – People dislike exercise, but love to feel fit.
39. ***Imagining a Connection with the Past*** – Nostalgia is a powerful emotional trigger for good and bad emotions.
40. ***Exploring a World*** – Understanding your environment gives you a sense of power and control.
41. ***Improving Society*** – Players can satisfy their need to leave the world a better place than when they came into it.

42. ***Enlightenment*** – Games provide a way for players to explore decisions and their consequences, leading to greater knowledge.

(Whatley, 2011)

APPENDIX A2: MAPPING GAME ELEMENTS TO FUN FACTORS

Motivation / Fun Factor	*Points*	*Badges*	*Leaderboard*	*Levels*	*Challenges*	*Activity Feed*	*Avatar*	*Onboarding*	*Game Master*	*Storytelling*	*Videos*	*Facial animations*	*Progress bar*	*Quiz*	*Timer*	*Like*	*Prize*	TOTAL
Recognizing Patterns	x					x								x				3
Collecting	x	x															x	3
Finding Random Treasures		x															x	2
Achieving a Sense of Completion	x	x	x	x	x			x		x			x	x	x		x	11
Gaining Recognition for Achievements	x	x	x	x		x			x	x						x	x	9
Creating Order out of Chaos	x	x	x	x	x	x		x	x	x	x	x	x		x	x	x	15
Customizing Virtual Worlds							x		x									2
Gathering Knowledge					x	x		x		x	x	x		x				7
Organizing Groups of People			x			x			x									3
Noting Insider References									x									1
Being the Centre of Attention			x			x				x							x	4
Experiencing Beauty and Culture		x		x			x				x	x						5
Romance							x					x						2
Exchanging Gifts																		0
Being a Hero			x			x	x									x	x	5
Being a Villain							x											1
Being a Wise Old Man			x	x		x	x									x	x	6
Being a Rebel							x											1
Being the Ruler			x			x	x											3
Pretending to Live in a Magical Place									x	x								2

Listening to a Story									x	x		x						3
Telling Stories									x									1
Predicting the Future						x												1
Competition			x			x										x	x	4
Psychoanalyzing			x			x			x									3
Mystery					x			x	x	x	x	x		x				7
Mastering a Skill					x			x						x				3
Exacting Justice and Revenge																x		1
Nurturing																		0
Excitement		x	x	x	x			x		x	x	x		x			x	10
Triumph over Conflict			x														x	2
Relaxing																		0
Experiencing the Freakish or Bizarre	x	x	x	x	x	x				x	x	x						9
Being Silly																		0
Laughing																		0
Being Scared					x									x				2
Strengthening a Family Relationship																		0
Improving One's Health																		0
Imagining a Connection with the Past																		0
Exploring a World								x	x	x								3
Improving Society									x									1
Enlightenment	x	x	x	x	x	x		x	x	x	x	x		x			x	13
TOTAL	7	9	14	8	9	14	8	8	13	12	7	9	2	8	2	6	12	

TABLE 9-2. MAPPING GAME ELEMENTS TO FUN FACTORS

APPENDIX A3: MAPPING GAME ELEMENTS TO MOTIVATION

Motivation / Game Element	***Power***	***Curiosity***	***Independence***	***Acceptance***	***Order***	***Saving***	***Honor***	***Idealism***	***Social contact***	***Family***	***Status***	***Vengeance***	***Romance***	***Physical Activity***	***Tranquility***	**TOTAL**
Points	2	3	2	1	4	1	0	0	1	0	2	0	0	0	3	19
Badges	3	2	2	1	4	2	0	0	1	0	3	0	1	1	4	24
Leaderboard	7	3	4	2	5	0	3	1	3	3	7	3	4	2	3	50
Levels	2	2	3	2	3	0	1	0	1	1	2	0	1	1	4	23
Challenges	2	4	3	0	2	0	0	0	1	0	1	0	1	3	3	20

Activity Feed	5	5	3	2	6	0	3	1	4	3	8	2	4	1	2	49
Avatar	5	0	5	1	2	0	4	1	3	3	5	3	5	0	1	38
Onboarding	2	4	4	0	2	0	0	0	1	0	1	0	1	2	3	20
Game Master	1	1	0	3	3	1	1	1	4	2	3	0	3	2	3	28
Storytelling	2	7	3	1	2	0	0	1	3	0	3	0	3	0	4	29
Video	0	4	1	0	2	0	0	0	1	0	1	0	2	1	3	15
Facial Animations	0	5	1	0	2	0	1	0	2	1	1	0	3	1	3	20
Progress Bar	1	0	1	0	2	0	0	0	0	0	0	0	0	0	2	6
Quiz	2	4	3	0	2	0	0	0	1	0	1	0	1	3	2	19
Timer	1	0	1	0	2	0	0	0	0	0	0	0	0	0	2	6
Like	4	0	2	2	1		2	1	2	1	4	3	2	1	1	26
Prize	8	1	4	2	3	2	2	1	2	1	7	3	2	2	3	43
TOTAL	47	45	42	17	47	6	17	7	30	15	49	14	33	20	46	

TABLE 9-3. MAPPING GAME ELEMENTS TO MOTIVATION

APPENDIX B: PILOT STUDY

APPENDIX B1: QUESTIONNAIRE

Q1: What did you like about the prototype?
Q2: What did you dislike about the prototype?
Q3: What should be improved for the prototype?
Q4: How did you like the game elements?
Q5: Was the goal of the requirements elicitation workshop clear to you?
Q6: Do you think other employees of MaibornWolff are capable to independently use the tool?
Q7: How much time would be necessary to gather requirements for this business case?
Q8: Do you think such a tool could be used for an RE workshop?
Q9: Are the questions in the survey appropriately and clearly formulated?
Q10: Is the video conferencing tool an appropriate case study to experiment with?

APPENDIX B2: ANSWERS

PILOT TESTER 1

Q1	▪ *"I very much enjoyed the activity feed. This gave me an ambition to want me to write the second user story before pilot tester 2."* ▪ *"Being able to see the activities of pilot tester 2 motivated me to work more."* ▪ *"I enjoyed the animated introduction movie with Tom."* ▪ *"I also liked the personas and how they were represented."* ▪ *"The prototype was easy to use and self-explaining."*
Q2	▪ *"The roles of the personas were not completely clear to me. Furthermore, I always first selected the persona which was similar to my personal role."* ▪ *"The quiz was for me a negative surprise and caused negative emotions, because I was not expecting it."* ▪ *"The Explanation of the Captain Up game elements (points, badges, levels) were to extensive in the onboarding program."* ▪ *"The scenario template is not quite easy to apply."*
Q3	▪ *"The roles of the personas must be better explained with more background information. This information could for instance be included in the spoken part."* ▪ *"I recommend to omit the point distribution for page visits. However, giving some points for visiting the page for the very first time is acceptable."*
Q4	▪ *"Some badges did not contain a description of how they can be achieved. This was making me more frustrating that curios."* ▪ *"The badges should be more outstanding."* ▪ *"I mainly perceived the game elements that addressed my personal interests. For instance, the activity feed or the levels."* ▪ *"Leveling up in the game is feasible and well implemented."*
Q5	▪ *"The goal of the practice and its outcome was not fully clear to me and should be stated in more detail. There should some instructions explaining how people must work together in such a tool."*
Q6	
Q7	▪ *"The time range for using the prototype should be set between 1 and 1.5 hours."*

Q8	▪ *"I can imagine that such a prototype might have a positive effect and cause some fun for the stakeholders."* ▪ *"The point system is an interesting approach for rewarding people."* ▪ *"There shouldn't be too many people on the platform at the same time. Otherwise it can become very confusing to keep track the activities and user stories."*
Q9	▪ *"There should be a note at the beginning of every questionnaire stating that the personal data will be anonymized and not publicly disclosed."* ▪ *"Participants should be able to reject the access of their Reiss profile."* ▪ *"There was a question regarding behavior-driven development in the first questionnaire. This is more of a buzz word and should be removed."*
Q10	▪ *"The video conferencing tool is a good case study, because the problem exists today. The three personas are sufficient as well."*

PILOT TESTER 2

Q1	▪ *"The leaderboard was for me more interesting and motivating to outperform my competitors."* ▪ *"I liked how the personas were able to talk and the inclusion of a profile description in written textual form."* ▪ *"One answer in the quiz regarding persona should was not correctly formulated."* ▪ *"Adding a quiz to the onboarding program is a good idea so that users pay attention and learn relevant knowledge."*
Q2	▪ *"The personas were for me also too scarce presented, and there were some contradicting information within these personas."* ▪ *"The user stories were well explained in the onboarding program, but the scenarios were lacking some examples."* ▪ *"I once received an error message stating that I am writing to many comments at a time (comment flood function in Wordpress)."* ▪ *"The chat was not noticeable. There wasn't any notification when I wrote something to pilot tester 2."*
Q3	▪ *"The personas should contain more information about what they do and what their interests are."* ▪ *"At the beginning of the onboarding program there should be a note stating that the new acquired knowledge will be tested with a quiz at the end."* ▪ *"The different game elements of the Captain Up plugin should be explained on maximum one page. One sentence for each game element should be enough."* ▪ *"There are still some mistakes in the texts of the onboarding program that should be checked."* ▪ *"An example of a scenario using the template should be added to the onboarding program."*
Q4	▪ *Receiving points for every page load was too disturbing. Furthermore, it was not clear to me why I received points for page visits without doing anything."* ▪ *"Setting a limited amount of time is fine by me, but I didn't pay attention to the timer. However, this might be because of the fact that I wasn't under any time pressure."* ▪ *"At first glance I didn't perceive the importance of the badges."* ▪ *"The goal of the badges shouldn't be to achieve as many as possible. However, they are still a nice supplementary mean to increase participation"*

	▪ *"The user story video was too long and far beyond its purpose. The focus should only be on user story cards."*
Q5	
Q6	▪ *"The prototype should be easy to use for any type of people, including also non IT-experts."* ▪ *"Writing scenarios was very easy, especially when the template is already given."* ▪ *"The scenario template was a bit more difficult to use, but I got used to it after applying it a few times."*
Q7	▪ *"With a real case I assume that two hours should be sufficient to gather user stories and scenarios."*
Q8	▪ *"I think such a solution is useful because of its location-independency. I would try it out in a real RE workshop to see the effect of the game elements."*
Q9	▪ *"It is obvious that the questionnaire are standardized."* ▪ *"Answering the questions regarding my emotions were difficult to answer."* ▪ *"There should be a note that people should answer the questions with gut instinct."*
Q10	

APPENDIX C: HEATMAP EXAMPLE

Heatmaps are graphical overlays of your website which points out what content is hot and what not. This is mainly done by tracking the mouse clicks of the visitor. The click-heatmap makes it possible to analyze the clicking behavior of the visitor (Nielsen, 2006).

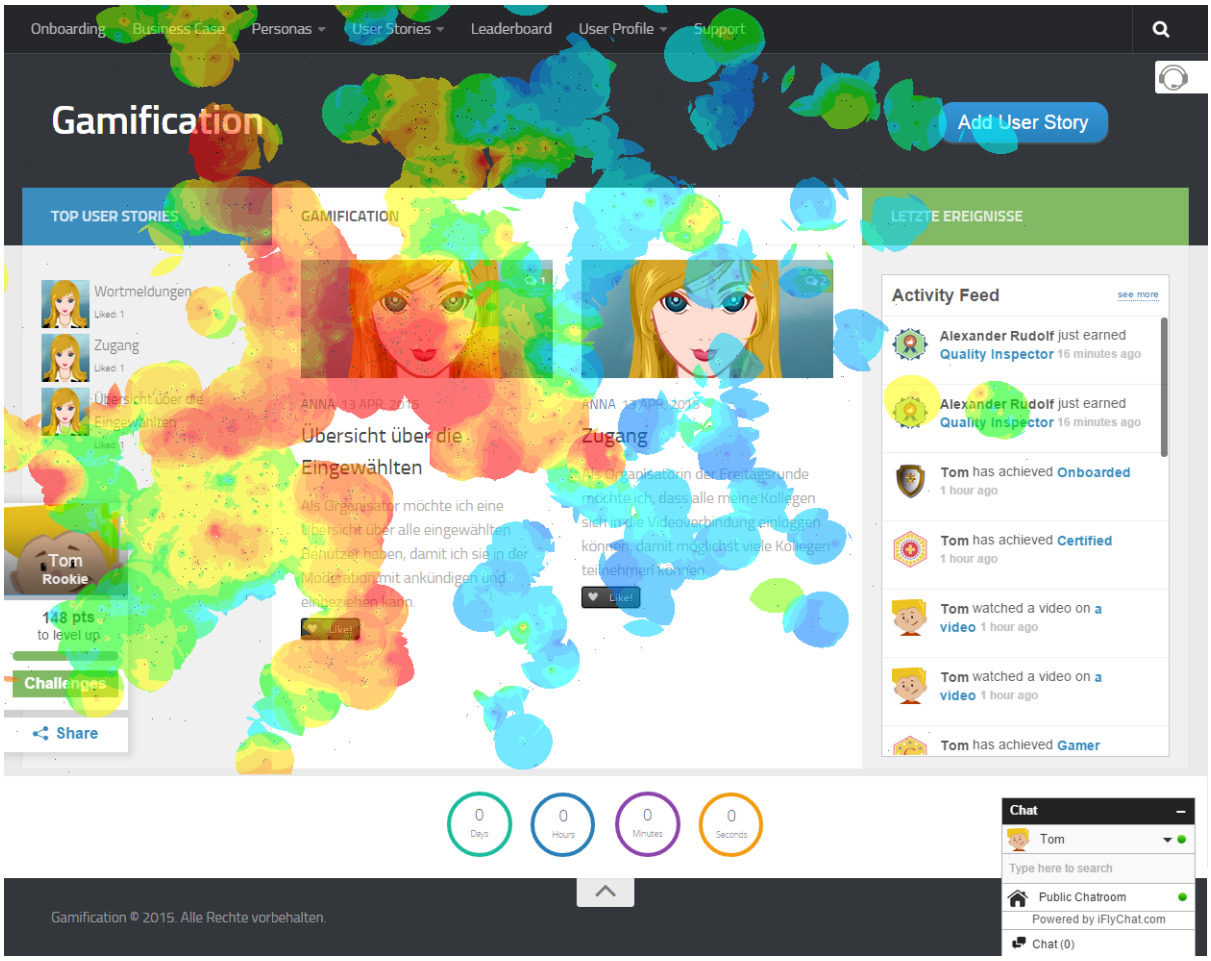

FIGURE 9-1. HEATMAP EXAMPLE

APPENDIX D: SURVEY RESULTS

APPENDIX D1: PRETEST

User ID	Initials	Age	Gender	Profession	IT Experience	RE Experience	User Story Experience	RE + User Story Experience (SUM)
1	P.N.	28,00	f	C	2	2	2	4
2	F.W.	30,00	f	C	7	7	7	14
3	M.B.	46,00	m	C	15	8	3	11
4	R.H.	26,00	m	SE	2	3	5	8
5	S.M.	35,00	m	C	7	6	6	12
6	N.K.	34,00	f	C	8	5	5	10
7	P.M.	38,00	f	E	12	1	2	3
8	G.B.	31,00	f	C	5	9	8	17
9	H.M.	35,00	m	C	6	7	7	14
10	D.U.	32,00	m	SE	2	2	3	5
11	J.T.	32,00	m	E	15	2	3	5
12	S.B.	25,00	f	C	1	1	5	6

Profession:

- C = IT Consultant
- SE = Software Engineer
- E = Enabler (employee for company internal operations)

TABLE 9-4. PRETEST RESULTS

APPENDIX D2: EXPERIMENTAL GROUPS

TREATMENT GROUP

User ID	Name	Gender	RE + User Story Experience	Profession	IT Experience
2	F.W.	f	14	C	7
4	R.H.	m	8	SE	2
5	S.M.	m	12	C	7
6	N.K.	f	10	C	8
11	J.T.	m	5	E	15
12	S.B.	f	6	C	1
Total	**6**	**3 x f 3 x m**	**55 (Ø = 9,2)**	**4 x C 1 x SE 1 x E**	**40 (Ø = 6,7)**

TABLE 9-5. TREATMENT GROUP

CONTROL GROUP

User ID	Name	Gender	RE + User Story Experience	Profession	IT Experience
1	N.P.	f	4	C	2
3	M.B.	m	11	C	15
7	P.M.	f	3	E	12
8	G.B.	f	17	C	5
9	H.M.	m	14	C	6
10	D.U.	m	5	SE	2
Total	**6**	**3 x f** **3 x m**	**54** **(Ø = 9)**	**4 x C** **1 x SE** **1 x E**	**42** **(Ø = 7)**

TABLE 9-6. CONTROL GROUP

APPENDIX D3: REISS PROFILE SCORES

User ID	1	2	3	4	5	6	7	8	9	10	11	12
User	N.P.	F.W.	M.B.	R.H.	S.M.	N.K.	P.M.	G.B.	H.M.	D.U.	J.T.	S.B.
Power	-1,01	0,25	-0,25	-0,76	0,25	-0,14	-0,76	-0,13	0,25	-1,65	-0,50	-0,89
Independence	0,24	0,56	-0,56	-0,24	1,23	2,00	-0,56	0,24	0,71	0,87	0,68	-0,08
Curiosity	0,23	0,83	0,08	1,59	-0,16	0,08	-0,53	0,98	0,23	0,38	0,78	0,98
Acceptance	-0,40	-0,51	-0,17	-1,65	-0,59	-2,00	-0,28	-1,08	-0,62	0,74	-0,35	-0,40
Order	0,38	0,75	0,88	0,00	0,33	0,12	-0,12	0,50	0,12	-0,38	-0,66	1,25
Saving	0,06	0,42	-0,42	1,02	0,06	-1,87	-0,78	-0,18	-1,51	-0,78	-1,47	-0,90
Honor	0,91	0,76	0,00	0,76	2,00	1,21	1,06	0,91	-0,15	-0,15	1,21	-0,15
Idealism	-1,95	-0,24	-1,83	-0,85	-2,00	-0,49	-0,98	-0,37	-0,37	-0,73	-1,30	0,49
Social Contact	1,27	0,42	-1,13	-1,55	0,09	-0,56	-0,70	0,70	-0,56	-1,13	0,63	-0,42
Family	0,32	-0,08	-0,22	-0,22	0,47	0,19	0,19	0,32	0,59	0,59	0,07	-0,35
Status	-2,00	-0,68	-0,16	-2,00	0,34	-0,78	-1,41	-0,68	0,16	-0,57	-0,34	-0,26
Vengeance	-0,25	-1,03	-1,52	-0,34	-0,54	-1,62	-0,83	-0,74	-0,05	-1,42	0,34	-1,72
Romance	0,17	-0,11	-0,96	-1,10	0,08	-0,82	-2,00	0,59	0,31	-1,10	0,43	-0,68
Eating	1,06	-0,56	0,69	-2,00	-0,41	-1,81	-0,31	-0,94	-0,06	-0,94	0,52	0,44
Physical Activity	0,75	1,45	0,65	0,65	1,10	-1,05	-1,95	0,35	1,65	-0,65	-0,10	-0,55
Tranquility	-0,43	-1,08	-0,43	-1,08	-0,84	-0,75	-0,97	0,97	-0,54	-0,11	-0,74	-0,43

TABLE 9-7. REISS PROFILE SCORES

APPENDIX D4: CONTROLLING MOTIVATION

Based on the Reiss profile data, there were no significant differences found in human motivations between the experimental group (A) and the control group (B). Therefore, it can be assumed that both groups are well-balanced without any significant differences in personal desires.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Power	A	6	-,2983	,49604	,20251
	B	6	-,5917	,68727	,28058

Independence	A	6	,6917	,83456	,34071
	B	6	,1567	,60928	,24874
Curiosity	A	6	,6833	,63500	,25924
	B	6	,2283	,48684	,19875
Acceptance	A	6	-,9167	,71715	,29278
	B	6	-,3017	,60327	,24628
Order	A	6	,2983	,65573	,26770
	B	6	,2300	,45268	,18481
Saving	A	6	-,4567	1,13463	,46321
	B	6	-,6017	,55449	,22637
Honor	A	6	,9650	,70973	,28974
	B	6	,4300	,58573	,23912
Idealism	A	6	-,7317	,86467	,35300
	B	6	-1,0383	,69995	,28576
Social Contact	A	6	-,2317	,79396	,32413
	B	6	-,2583	1,00597	,41069
Family	A	6	,0133	,29602	,12085
	B	6	,2983	,30049	,12268
Status	A	6	-,6200	,78271	,31954
	B	6	-,7767	,80092	,32697
Vengeance	A	6	-,8183	,79353	,32396
	B	6	-,8017	,59536	,24306
Romance	A	6	-,3667	,59018	,24094
	B	6	-,4983	1,01140	,41290
Eating	A	6	-,6367	1,07608	,43931
	B	6	-,0833	,82764	,33788
Physical Activity	A	6	,2500	,97724	,39896
	B	6	,1333	1,26082	,51473
Tranquility	A	6	-,8200	,24454	,09983
	B	6	-,2517	,65959	,26928

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Power	,848	10	,416	,29333	,34602	-,47765	1,06432
Independence	1,268	10	,233	,53500	,42185	-,40493	1,47493
Curiosity	1,393	10	,194	,45500	,32666	-,27285	1,18285
Acceptance	-1,607	10	,139	-,61500	,38259	-1,46746	,23746
Order	,210	10	,838	,06833	,32529	-,65647	,79313

Saving	,281	10	,784	,14500	,51556	-1,00375	1,29375
Honor	1,424	10	,185	,53500	,37567	-,30206	1,37206
Idealism	,675	10	,515	,30667	,45416	-,70528	1,31861
Social Contact	,051	10	,960	,02667	,52319	-1,13907	1,19240
Family	-1,655	10	,129	-,28500	,17220	-,66869	,09869
Status	,343	10	,739	,15667	,45718	-,86200	1,17534
Vengeance	-,041	10	,968	-,01667	,40500	-,91906	,88573
Romance	,275	10	,789	,13167	,47806	-,93352	1,19685
Eating	-,998	10	,342	-,55333	,55422	-1,78820	,68154
Physical Activity	,179	10	,861	,11667	,65124	-1,33438	1,56772
Tranquility	-1,979	10	,076	-,56833	,28719	-1,20822	,07156

TABLE 9-8. INDEPENDENT T-TEST FOR MOTIVATION

APPENDIX E: EXPERIMENT RESULTS

APPENDIX E1: GAMIFICATION

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
Storytelling	5	4,00	5,00	4,6000	,54772	,300
Onboarding	5	2,00	5,00	4,4000	1,34164	1,800
Videos	5	4,00	5,00	4,4000	,54772	,300
Progress Bar	5	4,00	5,00	4,4000	,54772	,300
Challenges	3	4,00	5,00	4,3333	,57735	,333
Game Master	4	3,00	5,00	4,2500	,95743	,917
Timer	5	3,00	5,00	4,2000	,83666	,700
Quiz	5	2,00	5,00	4,0000	1,22474	1,500
Activity Feed	5	3,00	4,00	3,8000	,44721	,200
Leaderboard	4	3,00	4,00	3,7500	,50000	,250
Likes	5	2,00	5,00	3,6000	1,34164	1,800
Prize	5	2,00	5,00	3,6000	1,14018	1,300
Badges	5	3,00	4,00	3,4000	,54772	,300
Points	5	2,00	5,00	3,2000	1,30384	1,700
Avatar	5	2,00	4,00	3,2000	,83666	,700
Levels	5	2,00	3,00	2,8000	,44721	,200
Facial Animation	4	2,00	3,00	2,5000	,57735	,333

TABLE 9-9. GAME ELEMENTS STATISTICS

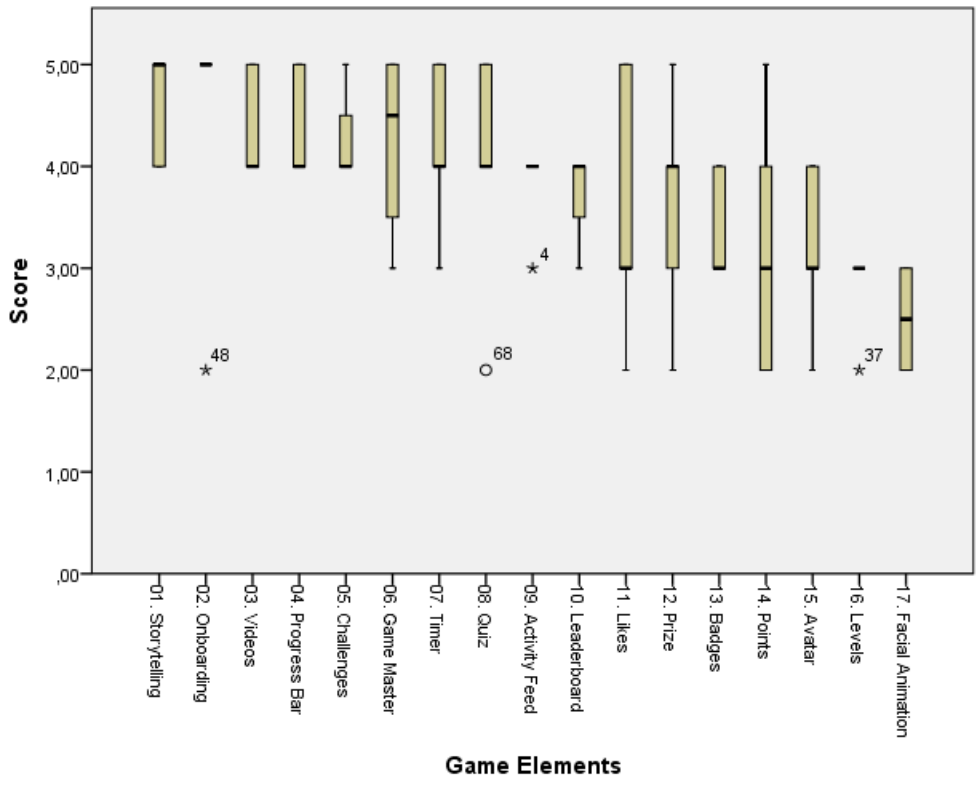

FIGURE 9-2. BOX PLOT FOR GAME ELEMENTS

GROUPED GAME ELEMENTS

	Motivation	N	Mean	Std. Deviation	Std. Error Mean
Score	Customized Elements	48	4,0208	1,04147	,15032
	Standardized Elements	32	3,4375	,80071	,14155

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Score	2,682	78	,009	,58333	,21751	,15030	1,01637

TABLE 9-10. INDEPENDENT T-TEST FOR TYPES OF GAME ELEMENTS

FIGURE 9-3. PLOT BOX FOR TYPES OF GAME ELEMENTS

APPENDIX E2: MOTIVATION

Table 9-11 shows some significant strong correlation results between motivation and game elements.

		Levels	Avatar	Likes
Power	Pearson Correlation	,439	,938*	,469
	Sig. (2-tailed)	,460	,018	,425
	N	5	5	5
Acceptance	Pearson Correlation	,985**	,000	,214
	Sig. (2-tailed)	,002	1,000	,729
	N	5	5	5
Honor	Pearson Correlation	,112	,762	,911*

	Sig. (2-tailed)	,858	,134	,032
	N	5	5	5
Idealism	Pearson Correlation	,041	-,608	-,924*
	Sig. (2-tailed)	,948	,276	,025
	N	5	5	5
Social Contact	Pearson Correlation	,890*	,377	,505
	Sig. (2-tailed)	,043	,532	,385
	N	5	5	5
Family	Pearson Correlation	,350	,719	,881*
	Sig. (2-tailed)	,564	,171	,049
	N	5	5	5
Status	Pearson Correlation	,907*	,174	,433
	Sig. (2-tailed)	,034	,779	,466
	N	5	5	5

Correlation is significant at the 0.05 level (2-tailed).

Correlation is significant at the 0.01 level (2-tailed).

TABLE 9-11. CORRELATION ANALYSIS

APPENDIX E3: EMOTIONS

	Group	N	Mean	Std. Deviation	Std. Error Mean
Emotions	A	5	36,8000	4,02492	1,80000
	B	6	37,0000	4,00000	1,63299

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Emotions	-,082	9	,936	-,20000	2,42884	-5,69442	5,29442

TABLE 9-12. INDEPENDENT T-TEST FOR EMOTIONS

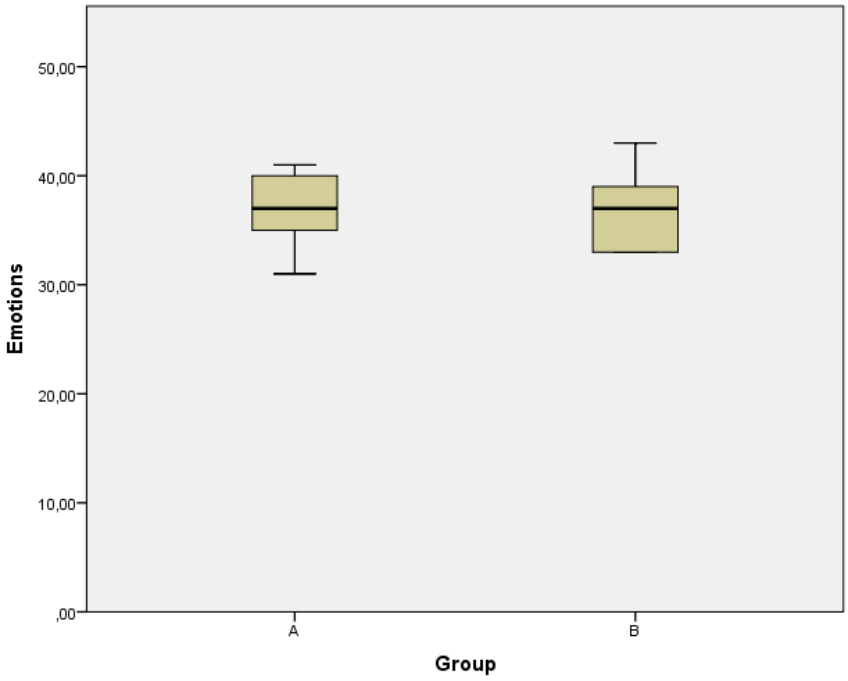

FIGURE 9-4. BOX PLOT FOR EMOTIONS

APPENDIX E4: COGNITION

	Group	N	Mean	Std. Deviation	Std. Error Mean
Cognition	A	5	50,4000	7,63544	3,41467
	B	6	43,3333	5,64506	2,30458

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Cognition	1,767	9	,111	7,06667	3,99901	-1,97973	16,11306

TABLE 9-13. INDEPENDENT T-TEST FOR COGNITION

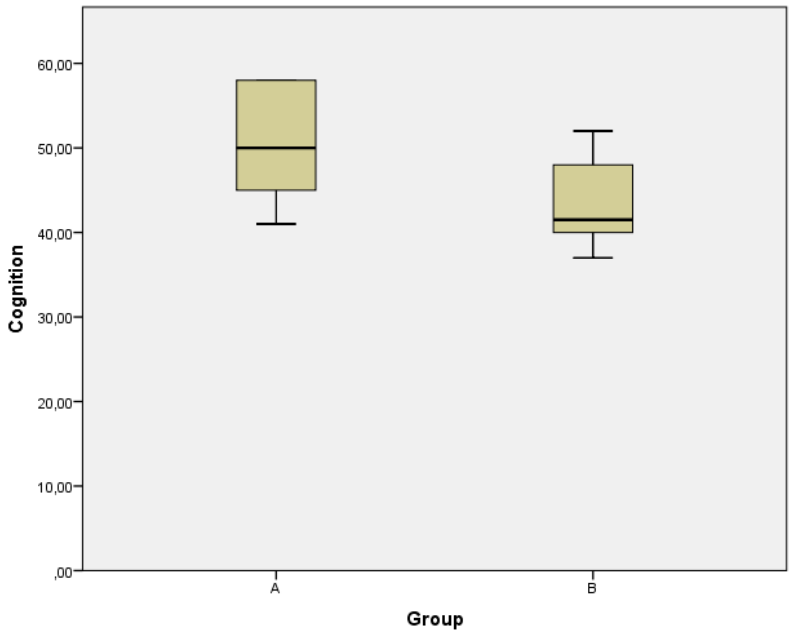

FIGURE 9-5. BOX PLOT FOR COGNITION

APPENDIX E5: CORRELATION BETWEEN EMOTIONS & COGNITION

Emotions and Cognition were strongly positively correlated, $r(11) = 0.623$, $p < 0.05$ as shown in the matrix below.

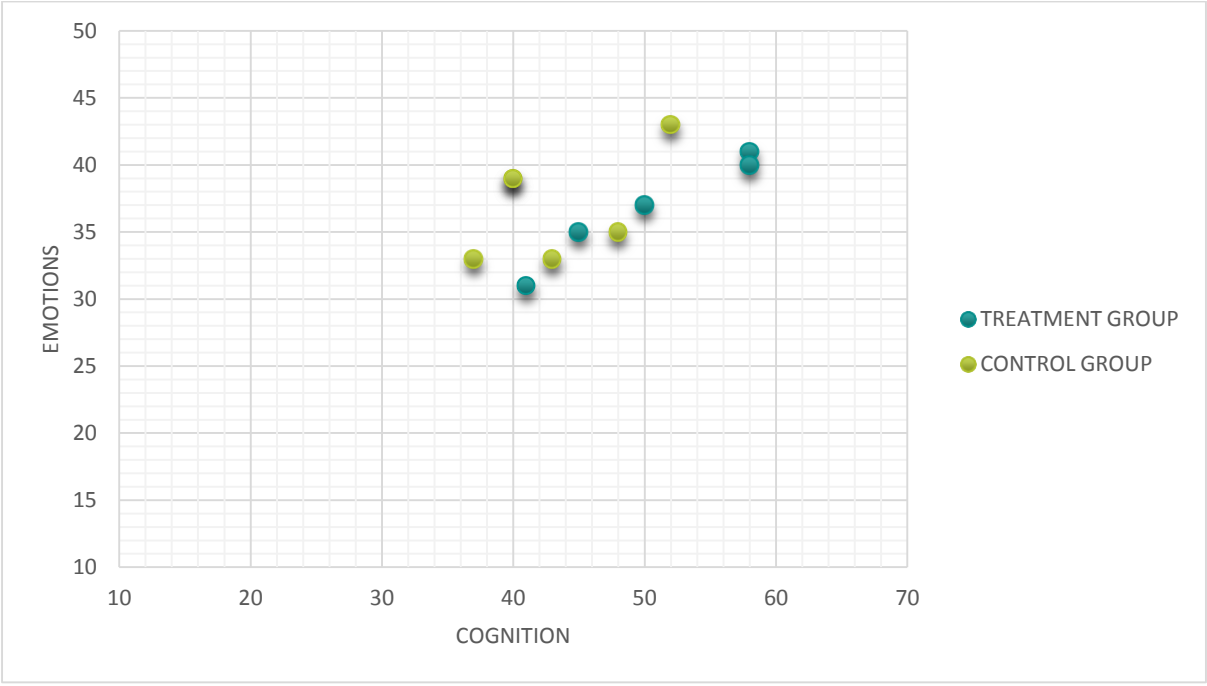

FIGURE 9-6. CORRELATION BETWEEN EMOTIONS & COGNITION

		Emotions	Cognition
Emotions	Pearson Correlation	1	,623
	Sig. (2-tailed)		,040
	N	11	11

Cognition	Pearson Correlation	,623	1
	Sig. (2-tailed)	,040	
	N	11	11

TABLE 9-14. CORRELATION BETWEEN EMOTIONS AND COGNITION

APPENDIX E6: EXPERIENCED DIFFICULTY

Participants of the gamification group did not significantly experience a lower level of difficulty for writing user stories (M = 2.0, SD = 0.707) than those of the control group (M = 3.168, SD = 1.722), $t(9)$ = -1.409, $p > 0.05$.

Users of the treatment group did not significantly find writing scenarios easier (M = 5.4, SD = 2.51) than those of the control group (M = 4.667, SD = 2.066), $t(9)$ = 0.533, $p > 0.05$.

Users of the gamification group did not perceive the level of difficulty for using personas significantly lower (M = 2.4, SD = 1.673) than those of the control group (M = 3.0, SD = 2.0), $t(9)$ = -0.532, $p > 0.05$.

The results of these three statistical conclusions are shown in Figure 9-7 and listed in Table 9-15.

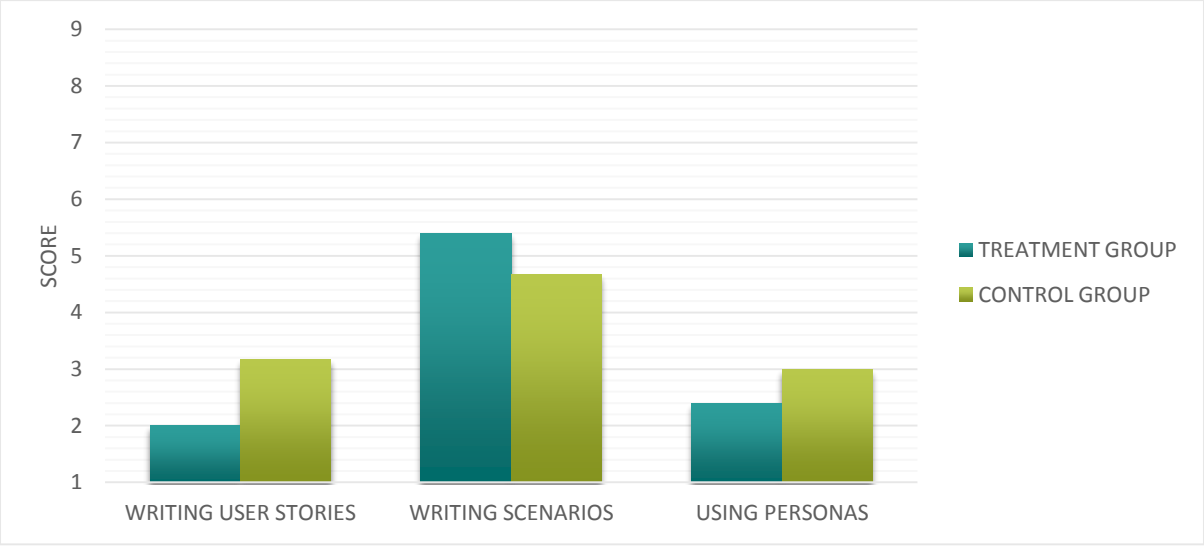

FIGURE 9-7. EXPERIENCED DIFFICULTY

	Group	N	Mean	Std. Deviation	Std. Error Mean
Writing User Stories	A	5	2,0000	,70711	,31623
	B	6	3,1667	1,72240	,70317
Writing Scenarios	A	5	5,4000	2,50998	1,12250
	B	6	4,6667	2,06559	,84327
Using Personas	A	5	2,4000	1,67332	,74833
	B	6	3,0000	2,00000	,81650

	t-test for Equality of Means					
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference

						Lower	Upper
Writing User Stories	-1,409	9	,192	-1,16667	,82813	-3,04003	,70670
Writing Scenarios	,533	9	,607	,73333	1,37688	-2,38139	3,84806
Using Personas	-,532	9	,607	-,60000	1,12744	-3,15044	1,95044

TABLE 9-15. INDEPENDENT T-TEST FOR SEVERITY

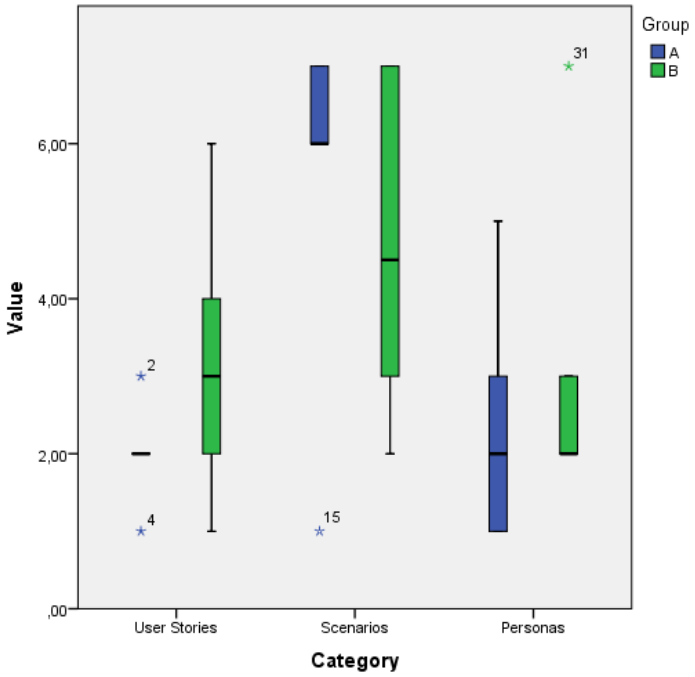

FIGURE 9-8. BOX PLOT FOR SEVERITY

APPENDIX E7: OVERALL SATISFACTION

An analysis of the overall satisfaction between the treatment group (M = 7.4, SD = 1.342) and the control group (M = 5.333, SD = 2.338) did not show any significant difference, $t(9) = 1.742$, $p > 0.05$. The following compares the results in a bar chart, followed by the statistical data in Table 9-16.

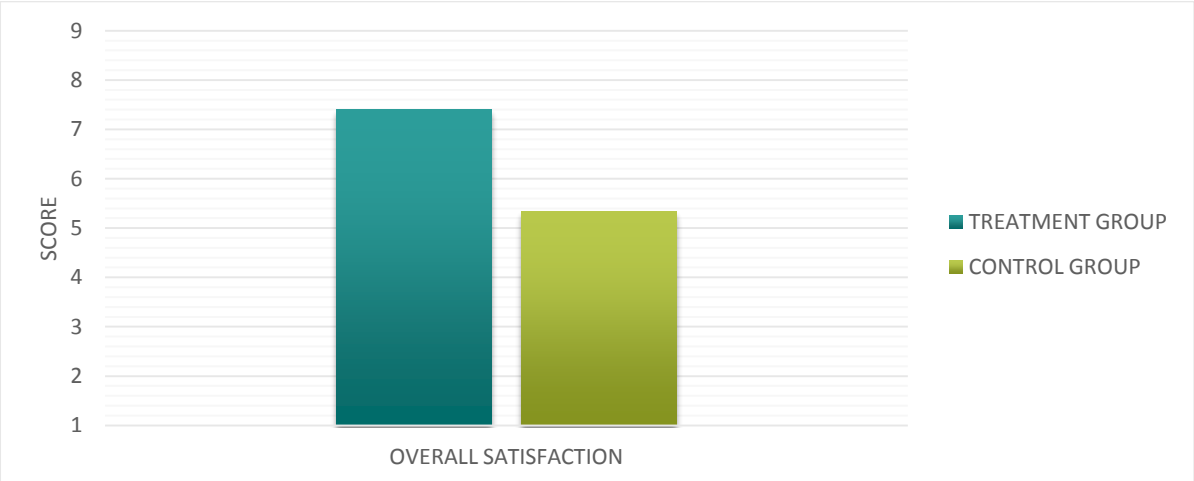

FIGURE 9-9. OVERALL SATISFACTION

	Group	N	Mean	Std. Deviation	Std. Error Mean
Satisfaction	A	5	7,4000	1,34164	,60000
	B	6	5,3333	2,33809	,95452

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Satisfaction	1,742	9	,115	2,06667	1,18613	-,61655	4,74989

TABLE 9-16. INDEPENDENT T-TEST FOR OVERALL SATISFACTION

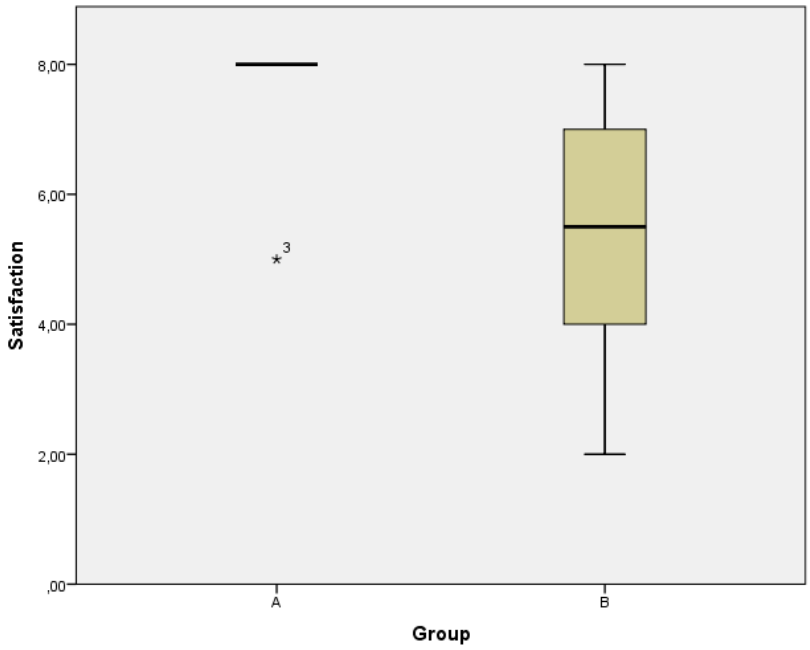

FIGURE 9-10. BOX PLOT FOR OVERALL SATISFACTION

APPENDIX E8: BEHAVIOR

	Group	N	Mean	Std. Deviation	Std. Error Mean
Page Visits	A	5	161,0000	40,36707	18,05270
	B	6	88,8333	38,33754	15,65123

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Page Visits	3,036	9	,014	72,16667	23,76857	18,39842	125,93491

TABLE 9-17. INDEPENDENT T-TEST FOR BEHAVIORAL DATA

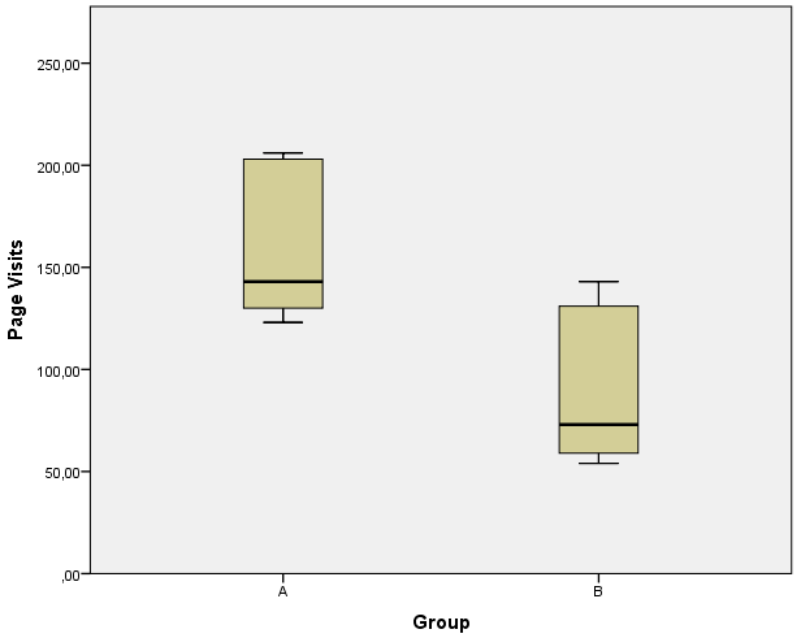

FIGURE 9-11. BOX PLOT FOR PAGE VISITS

APPENDIX E9: CHAT MESSAGES

	Group	N	Mean	Std. Deviation	Std. Error Mean
Chat Messages	A	5	1,0000	1,73205	,77460
	B	6	24,1667	19,99417	8,16258

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Chat Messages	-2,560	9	,031	-23,16667	9,05112	-43,64173	-2,69160

TABLE 9-18. INDEPENDENT T-TEST FOR CHAT MESSAGES

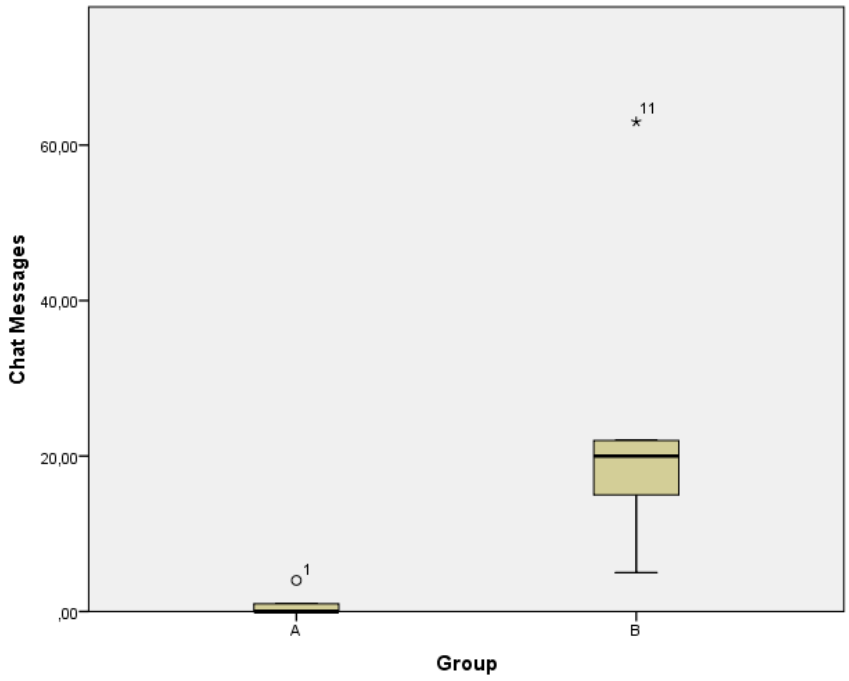

FIGURE 9-12. BOX PLOT FOR CHAT MESSAGES

MESSAGE FREQUENCY

The following diagram illustrates a time line showing at what time chat messages occurred during the experimental phase. From the image it is obvious that the control group was intensively collaborating, while the treatment group was not. From Table 9-19 we can see that the control group 1.2 wrote messages per minute, while 0.04 messages were written by the treatment group.

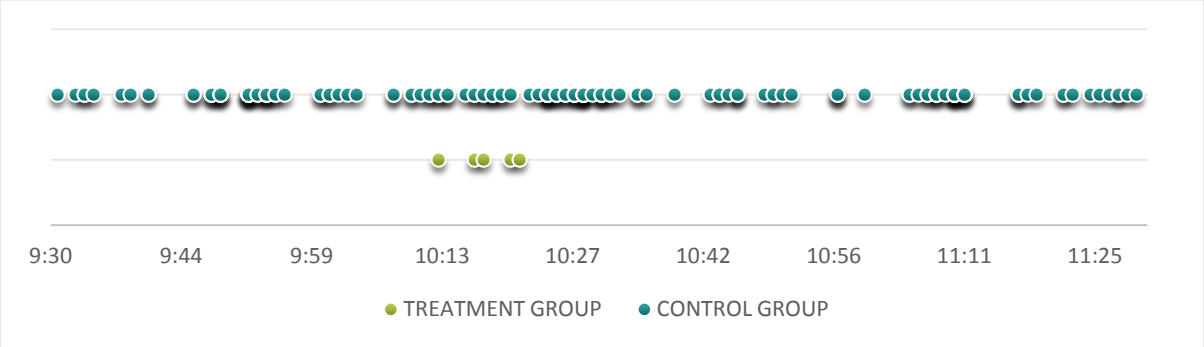

FIGURE 9-13. MESSAGE FREQUENCY

Treatment Group	Control Group
0,04 Messages / Minute	1,2 Messages / Minute

TABLE 9-19. MESSAGE FREQUENCY

APPENDIX E10: PRODUCTIVITY

	Group	N	Mean	Std. Deviation	Std. Error Mean
User Stories	A	5	10,0000	2,34521	1,04881
	B	6	3,5000	2,25832	,92195

Scenarios	A	5	13,4000	5,72713	2,56125
	B	6	3,0000	3,84708	1,57056

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
User Stories	4,673	9	,001	6,50000	1,39111	3,35309	9,64691
Scenarios	3,597	9	,006	10,40000	2,89137	3,85927	16,94073

TABLE 9-20. INDEPENDENT T-TEST FOR PRODUCTIVITY

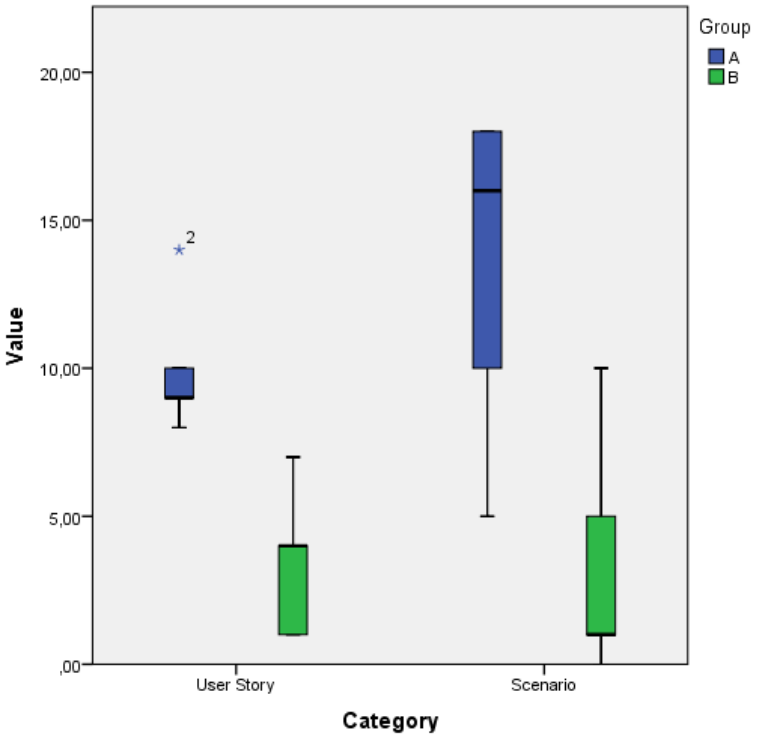

FIGURE 9-14. BOX PLOT FOR PRODUCTIVITY

REQUIREMENTS PRODUCTION FREQUENCY

The following image represents a timeline plot of when user stories and scenarios were written by the two experimental groups. From Table 9-21 we see that content creation frequency was higher in the gamification group.

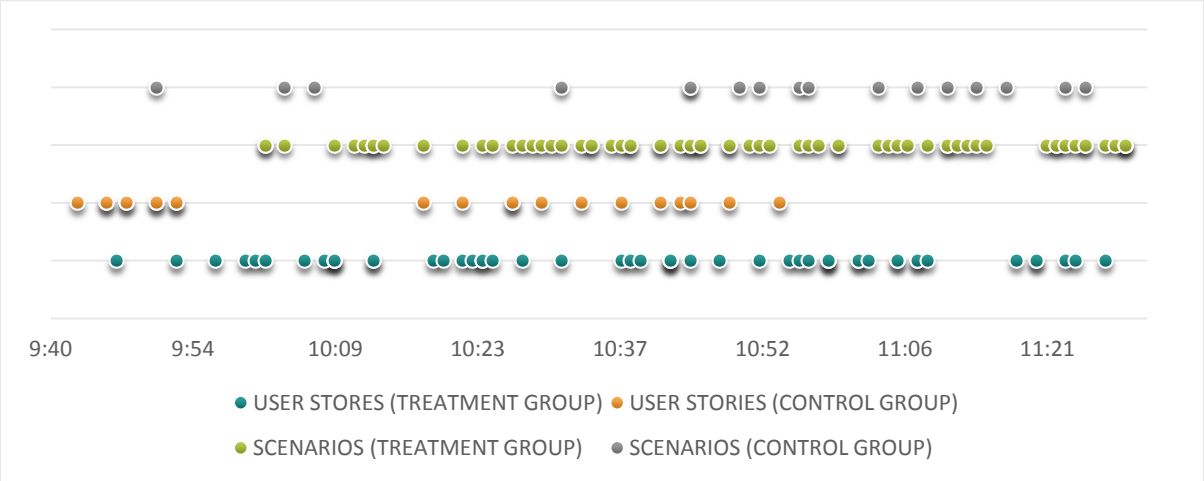

FIGURE 9-15. CONTENT PRODUCTION FREQUENCY

Treatment Group	Control Group
0,425 User Stories / Minute	0,175 User Stories / Minute
0,57 Scenarios / Minute	0,15 Scenarios / Minute
1,33 Scenarios / User Story	0,86 Scenarios / User Story

TABLE 9-21. CONTENT PRODUCTION FREQUENCY

APPENDIX E11: SOCIAL INTEREST

	Group	N	Mean	Std. Deviation	Std. Error Mean
Independence	A	5	,4300	,59758	,26725
	B	6	,1567	,60928	,24874
Social Contact	A	5	-,1660	,86927	,38875
	B	6	-,2583	1,00597	,41069

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Independence	,747	9	,474	,27333	,36581	-,55418	1,10084
Social Contact	,161	9	,876	,09233	,57383	-1,20576	1,39043

TABLE 9-22. INDEPENDENT T-TEST FOR SOCIAL INTEREST

APPENDIX E12: QUALITY

INVEST MODEL

	Group	N	Mean	Std. Deviation	Std. Error Mean
Independent	A	135	4,0222	,94999	,08176

	B	55	3,4364	1,30190	,17555
Negotiable	A	135	3,9852	1,09942	,09462
	B	55	3,8909	1,04833	,14136
Valuable	A	135	3,9333	1,05236	,09057
	B	55	4,0545	1,06141	,14312
Estimable	A	135	3,5037	1,17736	,10133
	B	55	2,4182	1,21245	,16349
Small	A	135	3,2444	1,18741	,10220
	B	55	2,3636	1,00671	,13574
Testable	A	135	4,1926	1,04021	,08953
	B	55	3,4182	1,37020	,18476

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Independent (equal variance not assumed)	3,025	78,481	,003	,58586	,19366	,20036	,97136
Negotiable	,543	188	,588	,09428	,17356	-,24810	,43666
Valuable	-,718	188	,473	-,12121	,16876	-,45412	,21169
Estimable	5,714	188	,000	1,08552	,18997	,71078	1,46026
Small	4,837	188	,000	,88081	,18211	,52156	1,24006
Testable (equal variance not assumed)	3,772	80,546	,000	,77441	,20531	,36588	1,18294

TABLE 9-23. INDEPENDENT T-TEST FOR INVEST

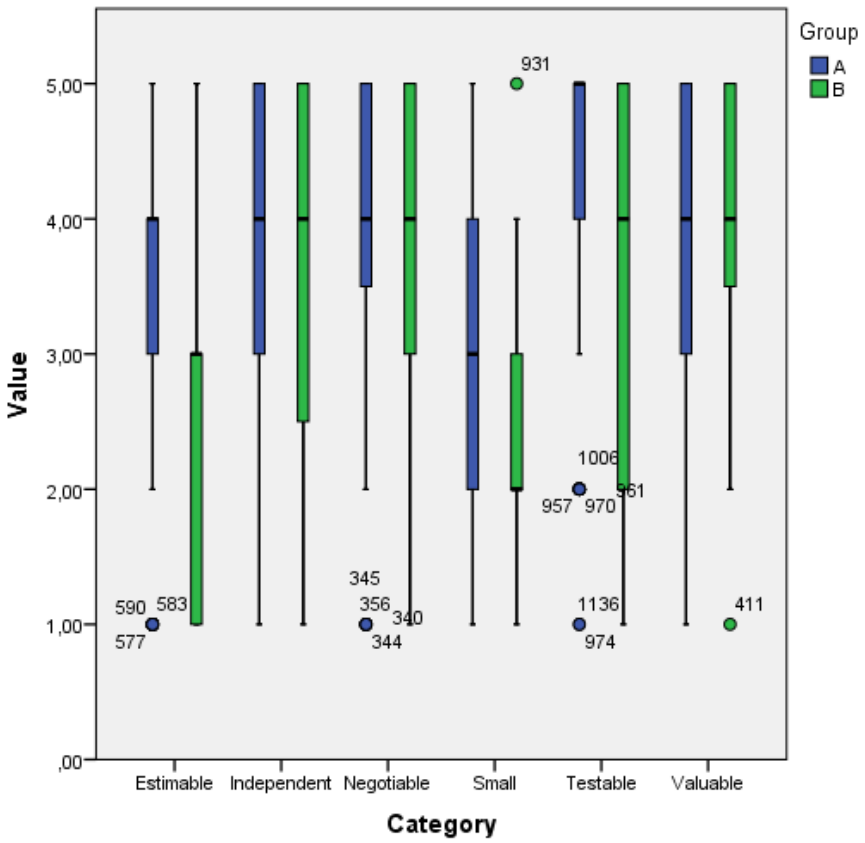

FIGURE 9-16. BOX PLOT FOR INVEST

PEARSON CORRELATION

		Estimatable	Small	Testable
Estimable	Pearson Correlation	1	,706	,708
	Sig. (2-tailed)		,000	,000
	N	180	180	180
Small	Pearson Correlation	,706	1	,556
	Sig. (2-tailed)	,000		,000
	N	180	180	180
Testable	Pearson Correlation	,708	,556	1
	Sig. (2-tailed)	,000	,000	
	N	180	180	180

TABLE 9-24. CORRELATION BETWEEN ESTIMABLE, SMALL AND TESTABLE

KANO MODEL

ID	Story Title		Attractive	Performance	Must-Be	Indifferent	Reverse	Questionnable
A01	**Vortrag über iPad-App**	Responses	9	1	0	3	0	0
		Percentage	69%	8%	0%	23%	0%	0%
A05	**Mitschnitt später ansehen**	Responses	8	1	2	2	0	0
		Percentage	62%	8%	15%	15%	0%	0%
A08	**Sprachqualität**	Responses	0	0	12	1	0	0
		Percentage	0%	0%	92%	8%	0%	0%
A11	**Sehen und Verstehen**	Responses	3	2	7	1	0	0
		Percentage	23%	15%	54%	8%	0%	0%
A12	**Flexibel Videokonferenzinhalte steuern**	Responses	5	2	3	3	0	0
		Percentage	38%	15%	23%	23%	0%	0%
A14	**Andere Konferenzteilnehmer stummschalten**	Responses	5	1	5	2	0	0
		Percentage	38%	8%	38%	15%	0%	0%
A15	**Flexibler Ortswechsel**	Responses	4	4	4	1	0	0
		Percentage	31%	31%	31%	8%	0%	0%
A17	**Chat mit anderen Standorten**	Responses	3	1	6	3	0	0
		Percentage	23%	8%	46%	23%	0%	0%
A18	**Protokollierung**	Responses	6	1	2	4	0	0
		Percentage	46%	8%	15%	31%	0%	0%
A19	**Powerpoints anzeigen**	Responses	3	1	7	2	0	0
		Percentage	23%	8%	54%	15%	0%	0%
A21	**Präsentationen herunterladen**	Responses	6	0	0	7	0	0
		Percentage	46%	0%	0%	54%	0%	0%
A23	**Mehrere Mikrofone**	Responses	6	2	2	3	0	0
		Percentage	46%	15%	15%	23%	0%	0%
A27	**Freitagsrunde nur an einem Rechner**	Responses	5	2	4	2	0	0
		Percentage	38%	15%	31%	15%	0%	0%
A31	**Videofunktion**	Responses	4	2	4	3	0	0

		Percentage	31%	15%	31%	23%	0%	0%
A32	**Vortrag über Rollenspiele**	Responses	1	0	0	12	0	0
		Percentage	8%	0%	0%	92%	0%	0%
A33	**Alle Konferenzteilnehmer sollen aufgelistet werden**	Responses	3	0	3	7	0	0
		Percentage	23%	0%	23%	54%	0%	0%
A34	**Dialog mit Video**	Responses	9	0	0	4	0	0
		Percentage	69%	0%	0%	31%	0%	0%
A35	**Chat Einträge markieren**	Responses	5	0	0	8	0	0
		Percentage	38%	0%	0%	62%	0%	0%
A38	**Nachbereitung der aufgezeichneten Veranstaltung**	Responses	6	0	1	5	1	0
		Percentage	46%	0%	8%	38%	8%	0%
A41	**Wichtige Chats direkt ins Online Protokoll übertragen**	Responses	6	0	1	6	0	0
		Percentage	46%	0%	8%	46%	0%	0%
A42	**Vorträge vorher aufnehmen einspielen**	Responses	7	0	0	5	1	0
		Percentage	54%	0%	0%	38%	8%	0%
A43	**Servicedienst**	Responses	6	1	0	6	0	0
		Percentage	46%	8%	0%	46%	0%	0%
A44	**Videos einspielen**	Responses	5	2	4	2	0	0
		Percentage	38%	15%	31%	15%	0%	0%
A46	**ICE von Köln nach München**	Responses	1	0	5	7	0	0
		Percentage	8%	0%	38%	54%	0%	0%
A47	**Agendathemen für den nächsten Termin im Tool sammeln/per Mail an das Tool schicken**	Responses	3	0	2	9	0	0
		Percentage	21%	0%	14%	64%	0%	0%
A48	**Intuitive Bedienbarkeit**	Responses	4	2	6	1	0	0
		Percentage	31%	15%	46%	8%	0%	0%
A50	**Publikum sehen**	Responses	5	1	4	3	0	0
		Percentage	38%	8%	31%	23%	0%	0%
B03	**Vorbereitung Freitagsrunde**	Responses	2	4	2	5	0	0
		Percentage	15%	31%	15%	38%	0%	0%
B04	**Aufzeichnung anschauen**	Responses	8	1	2	2	0	0
		Percentage	62%	8%	15%	15%	0%	0%

B05	**sich bewegen beim präsentieren**	Responses	3	3	6	2	0	0
		Percentage	21%	21%	43%	14%	0%	0%
B07	**Fragen remote stellen**	Responses	3	1	9	0	0	0
		Percentage	23%	8%	69%	0%	0%	0%
B10	**Moderation**	Responses	6	1	4	2	0	0
		Percentage	46%	8%	31%	15%	0%	0%
B14	**Nachbearbeitungsaufwand**	Responses	4	4	4	1	0	0
		Percentage	31%	31%	31%	8%	0%	0%
B15	**Themen remote präsentieren**	Responses	2	1	9	1	0	0
		Percentage	15%	8%	69%	8%	0%	0%
B16	**Einfache und schnelle Organisation**	Responses	4	4	4	1	0	0
		Percentage	31%	31%	31%	8%	0%	0%
B17	**Fragen beantworten**	Responses	2	1	7	3	0	0
		Percentage	15%	8%	54%	23%	0%	0%
B18	**Präsentationen in Echtzeit mitverfolgen**	Responses	1	1	8	3	0	0
		Percentage	8%	8%	62%	23%	0%	0%
B20	**Unabhängig vom Standort die Präsentation mitverfolgen**	Responses	0	1	11	1	0	0
		Percentage	0%	8%	85%	8%	0%	0%
A42	**Vorträge vorher aufnehmen einspielen**	Responses	7	0	0	5	1	0
		Percentage	54%	0%	0%	38%	8%	0%
A43	**Servicedienst**	Responses	6	1	0	6	0	0
		Percentage	46%	8%	0%	46%	0%	0%

TABLE 9-25. KANO EVALUATION TABLE

	Group	N	Mean	Std. Deviation	Std. Error Mean
CS+	A	27	,4411	,19266	,03708
	B	11	,3964	,20872	,06293
CS-	A	27	-,3130	,25588	,04924
	B	11	-,6109	,19357	,05836

	t-test for Equality of Means					
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference

						Lower	Upper
CS+	,634	36	,530	,04475	,07056	-,09835	,18784
CS-	3,468	36	,001	,29795	,08592	,12370	,47220

TABLE 9-26. INDEPENDENT T-TEST FOR CUSTOMER SATISFACTION

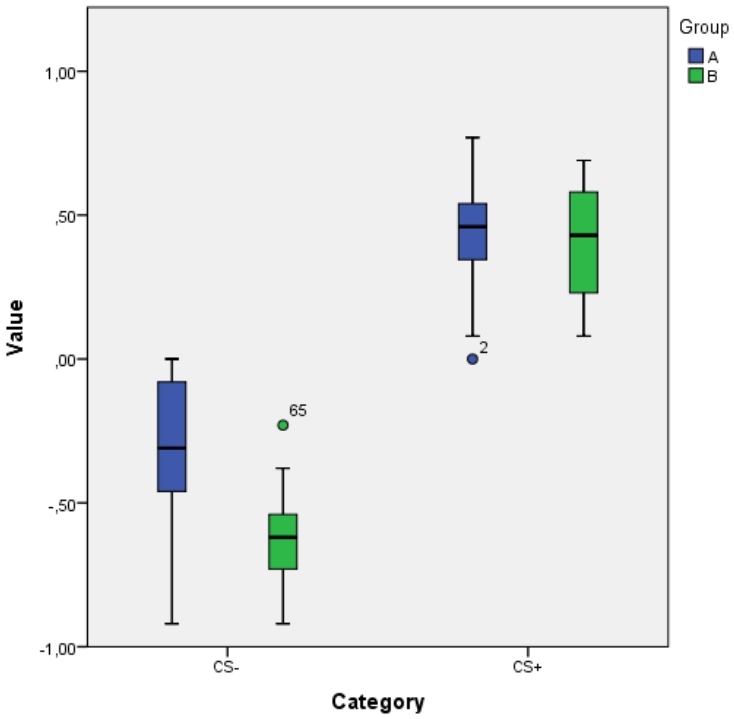

FIGURE 9-17. BOX PLOT FOR CUSTOMER SATISFACTION

APPENDIX E13: CREATIVITY

	Group	N	Mean	Std. Deviation	Std. Error Mean
Creativity	A	135	3,0444	1,08495	,09338
	B	55	2,2364	,92223	,12435

	t-test for Equality of Means						
						95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Creativity	4,853	188	,000	,80808	,16650	,47964	1,13652

TABLE 9-27. INDEPENDENT T-TEST FOR CREATIVITY

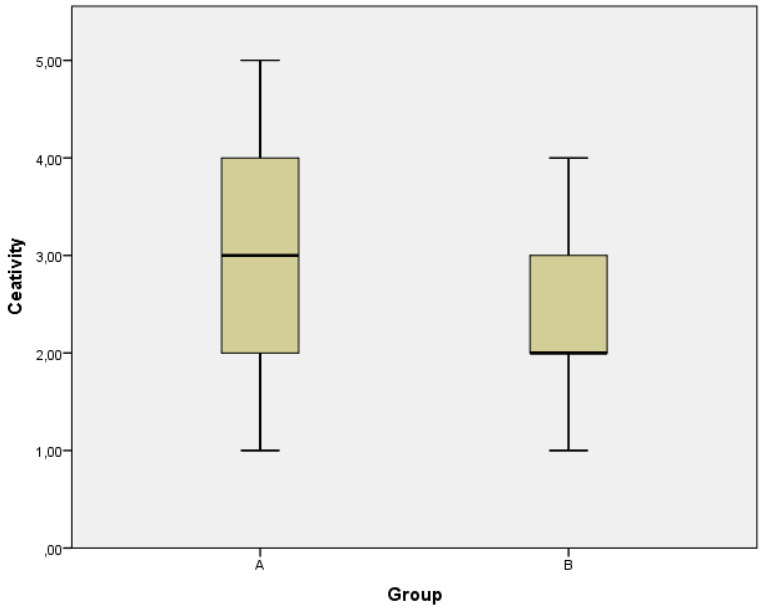

FIGURE 9-18. BOX PLOT FOR CREATIVITY

CREATIVITY PER USER STROY

Story ID	Mean	N	Std. Deviation	Variance
A01	3,8000	5	,83666	,700
A05	3,0000	5	,70711	,500
A08	1,6000	5	,89443	,800
A11	3,2000	5	1,30384	1,700
A12	3,4000	5	,89443	,800
A14	2,2000	5	1,09545	1,200
A15	2,4000	5	,89443	,800
A17	1,8000	5	,44721	,200
A18	3,4000	5	,89443	,800
A19	3,2000	5	,83666	,700
A21	3,2000	5	,83666	,700
A23	3,2000	5	1,64317	2,700
A27	1,8000	5	,83666	,700
A31	2,0000	5	1,00000	1,000
A32	4,0000	5	,00000	,000
A33	3,4000	5	,89443	,800
A34	3,8000	5	1,09545	1,200
A35	3,2000	5	,83666	,700
A38	3,6000	5	,54772	,300
A41	4,2000	5	,83666	,700
A42	3,8000	5	,44721	,200
A43	3,2000	5	1,09545	1,200
A44	3,0000	5	,70711	,500
A46	3,0000	5	,00000	,000

A47	3,2000	5	,83666	,700
A48	2,8000	5	1,30384	1,700
A50	2,8000	5	1,64317	2,700
B03	2,6000	5	1,51658	2,300
B04	3,0000	5	,70711	,500
B05	2,4000	5	,54772	,300
B07	1,8000	5	,44721	,200
B10	3,0000	5	1,00000	1,000
B14	2,0000	5	,70711	,500
B15	2,2000	5	1,09545	1,200
B16	2,4000	5	,89443	,800
B17	1,8000	5	,83666	,700
B18	1,8000	5	,83666	,700

TABLE 9-28. STATISTICS FOR CREATIVITY PER USER STORY

APPENDIX E14: RELATIONSHIP BETWEEN CREATIVITY & KANO CATEGORIES

ANOVA

Dependent Variable: Creativity

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	10,308	2	5,154	18,934	,000
Within Groups	9,528	35	,272		
Total	19,836	37			

			Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	Must-Be	Attractive	-1,06667*	,18751	,000	-1,5256	-,6078
		Indifferent	-1,02857*	,23644	,000	-1,6072	-,4499
	Attractive	Must-Be	1,06667*	,18751	,000	,6078	1,5256
		Indifferent	,03810	,23882	,986	-,5464	,6226
	Indifferent	Must-Be	1,02857*	,23644	,000	,4499	1,6072
		Attractive	-,03810	,23882	,986	-,6226	,5464

TABLE 9-29. ANOVA TEST FOR CREATIVITY & KANO CATEGORIES

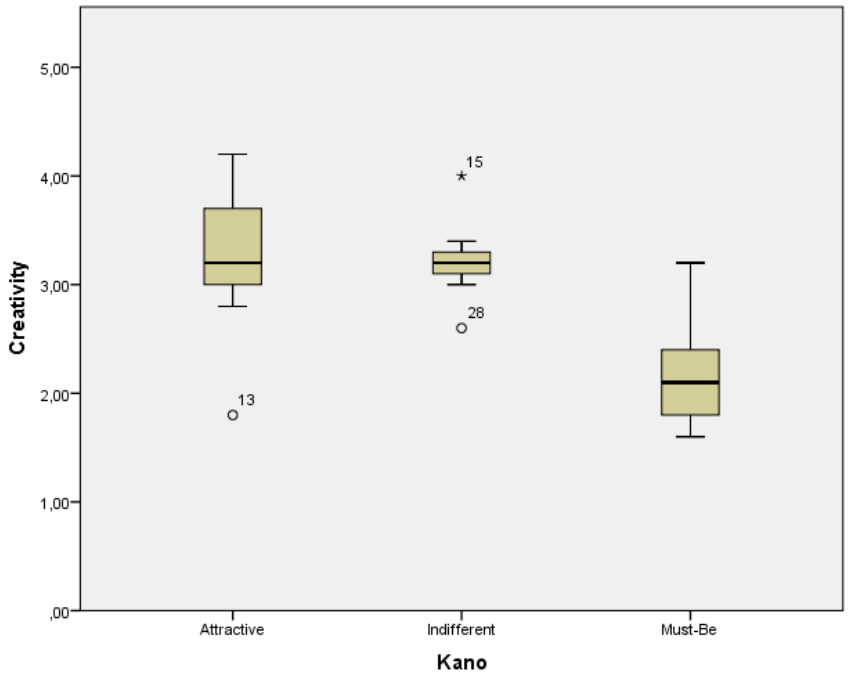

FIGURE 9-19. BOX PLOT FOR CREATIVITY PER KANO CATEGORY

SPEARMAN'S RHO CORRELATION

			Creativity	Kano Categories
Spearman's rho	Creativity	Correlation Coefficient	1,000	,632
		Sig. (2-tailed)	.	,000
		N	36	36
	Kano Categories	Correlation Coefficient	,632	1,000
		Sig. (2-tailed)	,000	.
		N	36	36

TABLE 9-30. SPEARMAN'S RHO TEST FOR CREATIVITY & KANO CATEGORIES