Evaluating the Benefits of Gamification both in Business Environment and for Business Purposes

UNIVERSITY OF TURKU
Faculty of Science and Engineering
Master's Thesis, Technology
November 2019
Konsta Sinisalo

Abstract

Contents

1. Introduction 1
1.1. Background for the research 2
1.2. Environment for the research 3
2. Previous research and theory 6
2.1. Reasons for gamification 6
2.1.1. Motivation psychology relevant for gamification 7
2.1.2. Other perceivable benefits of gamification 11
2.1.3 Gamification design 13
2.2. Evaluation of gamification benefits 14
2.2.1. Quantitative versus qualitative evaluation 15
2.2.2. Gamification evaluation in practice 16
2.2.3. Evaluation methods for the research 17
2.2.3. Recognized limitations of the study 19
3. The gamified solutions and how to evaluate them 20
3.1. Customer project 20
3.1.1. Requirements and constraints for the project 20
3.1.2. The development of the project 21
3.2. LOGE 24
3.2.1. The admin view – creating a discussion and observing feedback in LOGE 25
3.2.2. The gamified discussions in LOGE 27
3.3. Research plan 31
3.3.1. Customer project 31
3.3.2. LOGE 34
3.4. Hypotheses for the research 37
3.5. Conducting the research 41
4. Results 43

4.1. Project results 43
4.2. Relevance of results 43
4.3. Observations about used methods and applied theories 43
5. Analysis 44
5.1. Effectiveness of used methods and applied theories 44
5.2. How to improve gamification evaluation methods 44
6. Summary 45
References 46
Appendix A: The qualitative questions in the test 49
The training material 51
The ending test 52
Questions about the gamified training (only for the test group) 53

1. Introduction

In modern industry and especially software industry optimization, influence and effective and intelligent design have become imperative for economic competition and progress. In these fields, there is always a demand for ways to enhance productivity and current processes. In recent years, gamification has become a way to achieve this through UI and UX design.

Gamification is a term originally coined by Nick Pelling in 2002 [1]. The term refers to the application of game design elements in non-game environments and contexts [1][2] [11]. It is often used with the intention to engage users, solve problems, and to purposefully change how people behave [1][2].

Gamification as a concept has risen to public popularity in less than a decade (see Figure 1) and its concept has in that short time become a point of interest for both academy [3] and industry alike. For academy, there is interest to understand its effect and study its application; for industry, to optimize performance and influence with its application. Both to the psychological and the behavioural effects of gamification have been examined in a multitude of studies in the past decade [3][5]. These studies vary greatly in their experiments' timeframe length, methods, and the aspect of gamification they focus on [3][5].

When considering gamification to optimize their business, companies are mainly interested in its ability to produce value to the company. This can happen either by gamification having a direct calculatable effect on the company's efficiency or by it

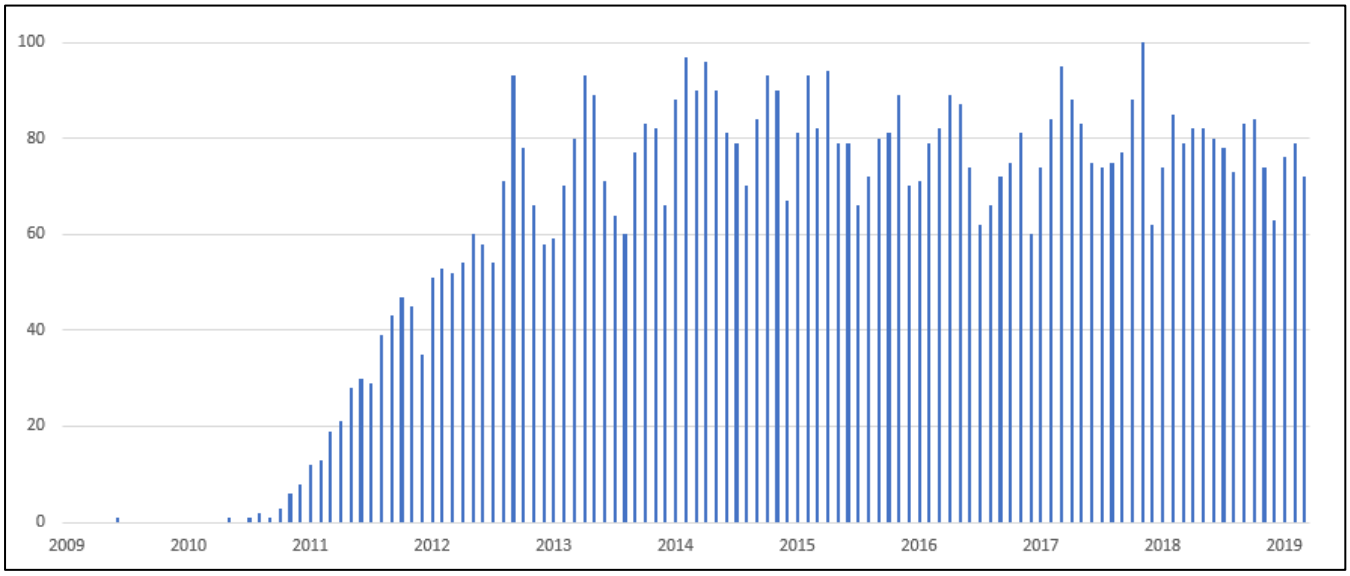

Figure 1 Trend data for "gamification" between 2009 and early 2019. Time represented by the x-axis, relative value by the y-axis. In the relative value 100 is the greatest amount of searches and all the other values are in relation to that. [4]

affecting the mood of the workers positively. Evaluation and quantification of these benefits has a clear value for companies that implement gamification.

In this thesis, we will investigate the current methods of measuring gamification benefits. We will take a look at how effective they are as tools in an industrial setting, and how they could be further improved. Chapter 2 discusses the theory of gamification and its benefits and how it has so far been evaluated in academic study. Chapter 3 presents the outlines of the research and how the evaluation of gamification benefits was implemented in the study. Chapter 4 presents the results received from the research and how useful and creditable they were in showing the gamification benefits of the research object. Chapter 5 analyses the effectiveness of the used methods and the applied theories and looks for ways to improve them. Chapter 6 presents the conclusion of the thesis along with discussion on possibilities for future gamification research methods.

1.1. Background for the research

Gamification in the current industry media is presented as a silver bullet that can be used for almost anything in any field.

"*Inspire a more focused, proficient, efficient and engaged workforce with real impact on your business*"
- Front web page of the gamification company Gameffective [6]

"*Gamification has the ability to drive new and returning customers to your store*"
- Blog text of the e-commerce company Shopify [7]

These quotes reflect the way gamification is commonly presented to the industry by gamification companies and consults. There appear in fact to be multiple different benefits in gamification depending on what one wants to focus on. The benefits can be examined focusing on the experience of an individual user that the gamification targets, the effect of gamification on the task that gamification is applied to, or the cumulative effects that gamification has on a company level.

Gamifying a task (or a process) also has a direct impact on its individual executor. Thus, it seems reasonable to assume that these two aspects are most heavily influenced by the gamification. The commonly supposed benefits of gamification usually involve the task becoming optimized and easier to execute; the implication is that bringing game elements to a task makes the flow of the task more intuitive and

engaging, keeps the users more focused, cuts the execution time, and might produce other unforeseeable benefits. For example, a simple task of filling in a form might become easier for the customer, if the form or the filling activity were to be gamified. Thus, gamification would by extension make the task faster to execute while also removing a lot of the company's need for customer support

For the purposes of this thesis, an individual's perspective in gamification means the perspective of an end user of the gamified task (if something else is meant it will be made clear in the context). When considering an individual in gamification, the perceived benefits are twofold: the gamified task becomes less taxing to mentally engage in, and gamification makes the task more intuitive and aids its efficient completion. This has also been acknowledged in the academical research, which categorizes gamification outcomes into psychological and behavioural [3]. A common public conception of gamification is not quite clear: on the one hand, gamification is a way to make mundane and repetitive tasks more engaging and tolerable. On the other hand, especially if done poorly, gamification is forced, uninteresting, and a waste of money. Another concern, albeit not as big one, is that if gamification is mixed with some sort of real-world reward system it will encourage people to cheat in order to win, to abuse system, or to instil other kinds of animosity among users.

Lastly, gamification can have great significance to the company. As stated above, it has been described almost like a silver bullet that can engage any workforce and make any task more effective. This has obvious and great benefits for a company that seeks to optimize their tasks in order to maximise profit or simply to minimize the employee turnover.

With these interests in mind, commercial gamification would greatly benefit from an efficient way to measure the actual benefits of gamification to a task and companies that sell gamification could better present their effectiveness. Companies buying gamification could have a better understanding of what gamification can achieve. Both sides would also be able to better negotiate the criteria that the gamified solution should fulfil.

1.2. Environment for the research

NordicEdu is a Finnish company founded in 2011 that specialises in gamification and educational games. [8] As of March 2019. NordicEdu has grown to employ over a

dozen people and has developed tens of software products, out of which 30 have been gamified software. [9]

In practice all NordicEdu's profits come from ordered and tailormade software projects. [9] This thesis tests and measures two gamified products within the developed by NordicEdu: a gamification project ordered from them by one of their customers and a product that NordicEdu is actively developing. NordicEdu has approved this testing of the products they developed. The company that ordered the second gamification project has also approved and accommodated this testing on their part but wishes to remain anonymous in this thesis. It will be referred to as *the customer* or *the customer company*.

The customer company that ordered the first gamification project is a large enterprise (over 250 employees) and their objective lies in retail. They have multiple outlets all around Finland and a steady economical standing. As a retail-oriented enterprise, the customer wanted to further develop their customer service by gamifying their introductory training of new retail workers. The customer was also in the process of renewing their e-commerce system. As a result, they ordered another gamification project with a goal to introduce their retail workers to the system and teach them its proper presentation and utilization in a retail environment. This latter e-commerce system training for their retail workers is the first study case for this thesis.

The second gamification project is a product developed by NordicEdu called LOGE. LOGE is based on a board game called Good Decision Game. The board game was developed by Talent Vectia and it aimed to gamify the interaction between people in corporate setting. To make Good Decision Game available to use even when working remotely Talent Vectia and NordicEdu begun to work on a digital version of the game in 2016. The end result, now called LOGE, is a browser-based tool developed using MEAN stack technologies (**M**ongoDB, **E**xpress.js, **A**ngular, and **N**ode.js). It is still being actively developed by Talent Vectia and NordicEdu, former of which offers it to its customers as a SaaS. [34][35]

This thesis focuses on properly evaluating how successful and effective gamification has been for both the project of the customer and LOGE. We will go through the gamified solutions, the empirical tests that were used to evaluate the success of the gamification and how well they functioned to measure gamification in a commercial setting. The research questions are as follows:

1. What current gamification theory and metrics are most suitable to evaluate commercial gamification?

2. How can the evaluation of gamification benefits best be implemented in a commercial environment?

2. Previous research and theory

Gamification is a concept that emerged after the growing popularity of gaming. Especially video games have in a few decades grown into a massive economic influence that is consumed more every day. For a long time, it has been a common part of people's lives and life styles. Games are typically extremely good at getting the end user (i.e. the player) interested in playing them and in wanting to keep playing them. In some cases, the time is not in any way pre-determined. This is a highly sought out goal in practically any professional work environment.

Gamification has received a lot of academic interest in the field after its emergence to public knowledge in the past decade and, especially, gamification of education has been the subject of many studies [3] [10]. Research on the subject varies widely from the psychological reasoning for gamification's effectiveness [13] to optimal designs for singular game elements for gamification purposes [14]. To understand gamification and to evaluate its implementation, one must first understand how gamification works and what can be achieved with it. [15]

2.1. Reasons for gamification

"…it is expected that gamification can foster the initiation or continuation of goal-directed behavior, i.e. motivation."

- Michael Sailer et al. [15]

Gamification is, in essence, a way to make tasks more appealing to the end user and to encourage active engagement in them. By extension, intelligent gamification can also be used to solve different UX (user experience) problems and to guide the user through the task without a need for tutorials that can most often be described as unappealing. One of the most used examples of this is the social networking platform LinkedIn [16] in which a new user is prompted to fill in potential missing information as they are simultaneously shown an abstract title of how well maintained their personal profile is. This is highly engaging and gets the users quickly invested in the platform without holding their hand through a painstaking tutorial.

User motivation is often overlooked but has a far graver effect on an enterprise as it becomes a constant and steady factor in corporate world. Unmotivated and

disengaged workers have been evaluated to cost 300 billion dollars overall per year in the US alone [27]. Research also indicates that employee engagement can reduce staff turnover by up to 87% and boost employee performance up until 44% [27]. Although motivation and engagement are hard to measure and it can be argued that these kinds of numbers are signs of positive subsequent secondary results.

Although gamification is used to motivate users and make a task feel less tedious, it has also been used to achieve other benefits as a result of these. For example, it has been used to engage customers [17], to help students achieve better grades [18] and to promote health activities [19]. In all of the previously mentioned cases, gamification has been evaluated to have a positive influence on the subject research subjects.

However, quantitative studies on gamification that have reported more than descriptive statistics have also discovered evidence to indicate results that are not entirely positive; some studies have showed that the results of gamification may not be long term and that, once implemented, removing gamification might in fact have "detrimental effects" [3] [10]. When discussing qualitative results, gamification studies tend to be more mixed and often the gamified aspects that appeal to one individual are explicitly disliked by another [3].

In the end, these studies seem to indicate that gamification can be used to achieve other results beyond simply user motivation. There have yet been fairly few limitations as to what could be achieved through it; there is a plethora of game elements to utilize and even more scenarios to use them in. Although the implementation of gamification is often done on an intuitive understanding user of motivation theories of its reasoning have emerged. These theories are based on behavioural psychology and have received some recognition.

2.1.1. Motivation psychology relevant for gamification

Guiding end users effectively to do what is intended of them is not a straightforward task. Telling someone to do something is not often enough to get them actually do it. And even though they might do what they are told, there is no incentive for them to prolong their engagement or exert any extra effort into it. Motivation through gamification is one way to achieve better user interaction. In order to understand the effects of gamification it is important to understand the psychology of human

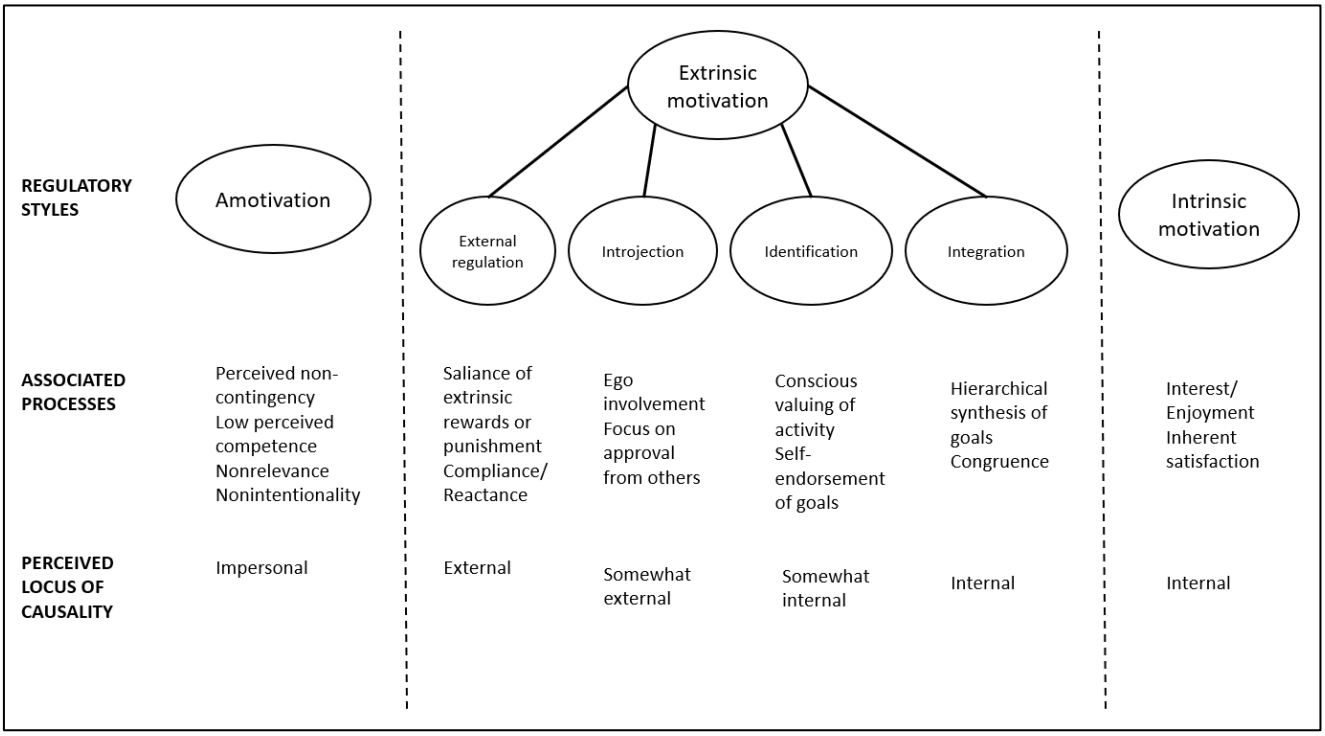

Figure 2: A taxonomy of human motivation. (Ryan and Deci) [21]

motivation. Concepts that one often come up are intrinsic and extrinsic motivation [21], self-determination theory [22] and flow theory [20].

Intrinsic motivation is motivation that derives from within, while extrinsic motivation derives from outer factors. In other words, intrinsic motivation is the desire to do something for the fun or challenge of it, whereas extrinsic motivation means that something is done as out of necessity or as a means to an end. For example, a child may play football because it brings them joy and because their parents tell them to strive to become a professional football player when they grow older. The joy of football in this case is an intrinsic motivator, while the pressure from the parents is extrinsic. As is apparent from this example, intrinsic and extrinsic motivation can both affect the end user while committing to a task. [21]

What is interesting when considering the differences between intrinsic and extrinsic motivation, is that intrinsic is typically more effective than extrinsic. People with authentic or self-imposed motivation tend to be more interested and excited about their tasks. This manifests as enhanced performance, persistence and creativity. [22]

Ryan and Deci [21][22] have done extensive work in the field of motivation psychology and especially self-determination theory (SDT) [22]. SDT is the theory that human nature repeatedly and consistently shows effort, commitment and agency when certain needs are met. These are, in essence, the manifestations of motivation and a high manifestation of these is a sign if exceptional intrinsic motivation. Ryan and Deci also

recognize the three major psychological needs that, when fulfilled or neglected, contribute essentially to the manifestation of these characteristics. These are competence, autonomy, and relatedness. Competence in this context means a person's feeling of competence and relatedness the level of social relations a person can share regarding the subject of the motivation. What is even more important is that two of these appear to be closely linked in terms of how they affect a person; a feeling of competence will not enhance intrinsic motivation without a sense of autonomy ("*an internal perceived focus of causality*"). [22]

Often the amount of intrinsic motivation is practically non-existent and external motivation needs to be applied for the end user to go through with the tasks. Purely extrinsic motivation like this is much less effective in comparison to intrinsic; although it may be enough for the task to be completed, it will most likely also manifest as disinterest, low effort, and other side-effects such as blaming others for negative outcomes. Ryan and Deci recognize that motivation often is not purely extrinsic or intrinsic. Extrinsic motivations can be integrated by a person, in which case they manifest more like intrinsic motivation (see Figure 2). This integration has four defined levels ranging from very external (external regulation in Figure 2) to very internal (integration). This is a subtheory of SDT called organismic integration theory (OIT) and it provides a great base for gamifications evaluation. [21]

In gamification both intrinsic and extrinsic motivation play a crucial role. Gamification is most often applied to tasks that are uninteresting, unengaging or monotonous; often even all these qualities. In such tasks, extrinsic motivation is the only reason for the end user to complete the task. It would be highly beneficial to integrate at least some of the external motivation into internal motivation. Very often gamification offers a valid solution to this.

"*A person in flow does not operate with a dualistic perspective: one is very aware of one's actions, but not of the awareness itself.*"

- Mihaly Csikszentmihalyi on merging action and awareness during flow [24]

Flow theory is a concept in positive psychology ("*a science of positive subjective experience, positive institutions and positive institutions*" [23]) that is also often brought up when gamification is discussed. The term was originally coined by Mihaly Csikszentmihalyi in 1975 who used it to refer to the mental state of a person fully immersed in an activity as flow state. [20]

Csikszentmihalyi identifies multiple elements of flow, the most notable of which is merging action and awareness. By this he means that a person in the flow is not actively reflecting on their actions even though they are aware of their actions. As an example of such a state, Csikszentmihalyi uses a tennis player who pays undivided attention to the ball and the opponent but pays no attention to the activity itself. The moment that awareness is split to reflect also on the activity "from outside" the flow is interrupted. As a direct result of this, the flow is difficult to maintain for any length of time. [24]

Centering of attention to a limited stimulus field is also recognized by Csikszentmihalyi as one of the five other elements of flow. He describes that during the flow the stimuli that would "intrude" on the task need to be kept out of attention. This also makes the merging of action and awareness possible during the flow state. Loss of ego (or loss of self-consciousness) is another element of the flow. This should not be interpreted as letting go of self-control. Quite the opposite, it means that the task involves the person with its demands for action to such an extent that "'selfish' consideration becomes irrelevant". [24]

A person in the flow is also in control of their action and of the environment. Csikszentmihalyi describes this control more as the condition of "not being worried by the possibility of lack of control" and remarks that the loss of ego intricately affects the state. Another element is that the task usually contains coherent and noncontradictory demands for action and a clear unambiguous feedback to a person's actions. Feedback is especially important and during the flow there is no doubt whether the action is "good" or "bad". Lastly, the nature of a task is autotelic if it induces the flow state. This means that the task appears to need no goals or rewards in itself; in other words, it is intrinsically motivated. [24]

Flow as a phenomenon is not too common an occurrence. Csikszentmihalyi recognized common restrictions to multiple situations where flow should be attainable and concluded that flow only occurs when persons face tasks that are within their ability to perform. He refines this into a flow state model (see Figure 3) and states that in order for flow to be achievable both the challenge level of the task and the skill level of the person should be on equal level. If a task is not challenging enough for the person, the person gets bored. If the person is not skilled enough for the task, the person gets anxious about the task. [20][24]

As a person is in the flow, in essence, completely engaged in whatever they in that moment are doing, the flow state can be viewed as the ultimate goal for

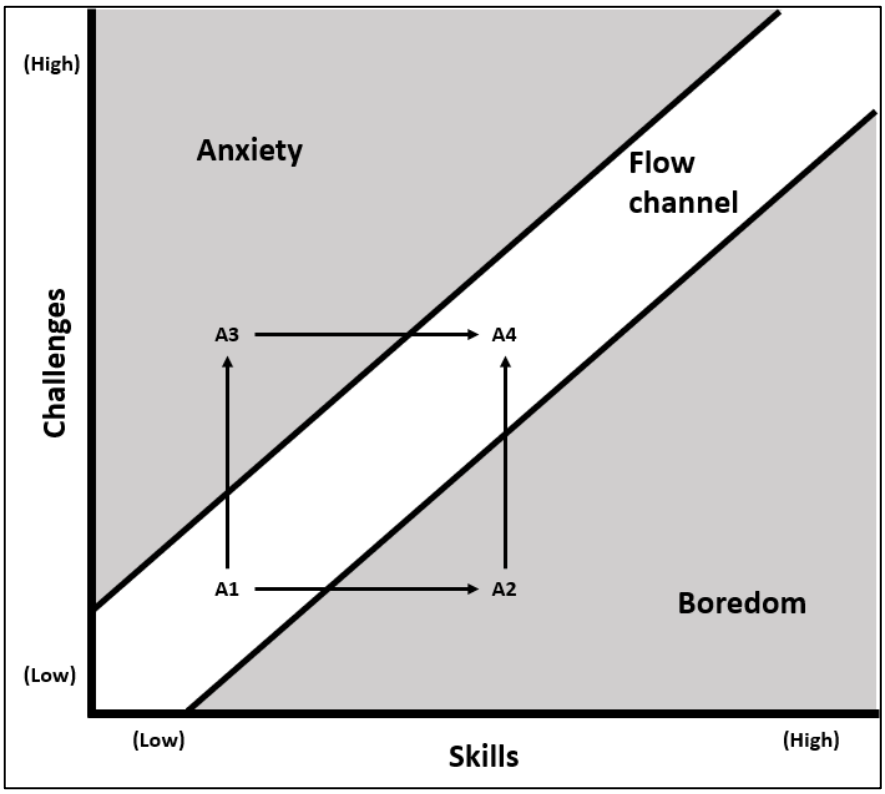

Figure 3: Diagram showing the stacking complexity of flow as the result of personal development. A1 is a low skilled person doing a low challenge task, A2 is a highly skilled person doing a low challenge task, A3 is a low skilled person doing a highly challenging task, and A4 is a highly skilled person doing a highly challenging task.[20]

gamification. Csikszentmihalyi gives an in-depth description of how flow functions and much of it should be considered in proper gamification design. Both SDT and flow handle subjects that are essential to gamification and its design and it would be highly beneficial to apply both theories while gamification is planned.

2.1.2. Other perceivable benefits of gamification

Meaningful gamification is a term developed to describe gamification that has achieved internalization, if the activity (as per SDT) does not depend upon external rewards [25]. Ruhi, however, refers to corporate scenarios where gamification is applied to integrate with existing business processes or information systems and used to help drive positive employee and organizational outcomes as meaningful gamification in an enterprise context [26]. This latter definition is more advantageous for the purposes of this study and so for the rest of this thesis “meaningful gamification” means meaningful gamification in an enterprise concept, as per Ruhi’s definition.

As stated earlier, gamification is often applied not because of the end user motivation and engagement, but because of other benefits that come with the heightened user interaction. Enterprise gamification can easily be divided into two different

segments based on the target of gamification: internal and external. Naturally, internal gamification in an enterprise context refers to gamification targeting to the companies' own employees, whereas external refers to gamification that targets to existing or potential customers. Both of these groups can be referred to as the end users or targets of gamification. So far, we have been using end user as the designated term and will continue to do so in this thesis. [27]

Even though the projects examined in this thesis focus on internal meaningful gamification, it is useful to take a brief overview on the external or customer-oriented aspect as well. Customer engagement is characterized by the repeated interactions of a customer with an organization through which emotional and psychological bonds with the organization are strengthened [17] [28]. This has been shown to boost customer loyalty [28]. Gamification has been shown to correlate between a positive effect on customer loyalty and even on positive buying behaviour which are indicative of better user engagement and can also be reasonably assumed to be caused by it [28]. External meaningful gamification is easily justified as an investment and can be, at least to some extent, measured by its effect on a company's profits.

Meaningful internal gamification in enterprise level seeks to improve employee engagement in the organization. Through this improvement of the organizations performance and better business outcomes can be achieved [29]. Employee engagement is emotional commitment to the organization and its goals [27]. This as a motivational construct is comparable to SDT and internalized motivation, where a person integrates extrinsic motivations as a part of their own behaviour.

Employee engagement has in itself an indirect impact on a variety of things. Especially strong effects on employee engagement have been found in relation to employee turnover, customer satisfaction-loyalty, and safety. Smaller but still notable correlation has also been observed between employee engagement and productivity, as well as profitability. The smaller correlation between the latter two to engagement, as opposed to the former three, could be the result of the latter being "more remote downstream variables" that are influenced by other variables besides employee engagement. All of the correlations were positive in their nature. Also, the correlations between these have been noticed to generalize across organizations, meaning that no matter the organization, similar correlation can be found between employee engagement and the above-mentioned aspects of an organization's business performance. [29]

When considering gamification from a business standpoint, it is interesting what results it can achieve on a single process. Gamification has been shown to improve the

amount of information a person can internalize [18], which would make it highly suitable for introductory and general training for enterprises.

Gamification's effect on the process speed has been studied only a little even though it would be extremely interesting from a business standpoint. Some research has been done on how "speed-gamification" affects other aspects of the performance by Jin et al. [30], but this hardly covers the subject in the way that would be beneficial in a corporate interest. This may be because when it comes to gamification, the aim is hardly ever to make the users to go through a process faster than before gamification was applied. Often the goal is to get better results from the process. This goal does not lend itself well to optimizing the process speed. Kaila et al. [18] note that they have utilized other automation in addition to gamification, automated code assessing for student exercises to be exact, that can significantly reduce the time spent on evaluation of student work, but this is not, strictly speaking, related to gamification. This should, however, be considered since direct and instantaneous feedback is an important aspect of gamification and in that sense automated evaluation is essential for the purposes of their gamified education project. Also, the end-user engagement not only can increase the effectiveness of the work done but has a chance of increasing the amount of time used working on the gamified task. As such gamification's effect on the process speed can be affected both negatively and positively by gamification which makes it less straightforward.

Both speed and effectiveness are variables that are relatively easy to measure in both projects this thesis focuses on. As such they will be taken into account during the conducted research.

2.1.3 Gamification design

In designing gamification for a task, there exists a common general procedure for coming up with the end gamification solution.
This path is as follows:

- preparation – the objectives of the gamification should be made clear (what should be achieved)
- analysis - the target task is analysed with the intention to identify necessary aspects to be taken into consideration like the knowledge of the users
- ideation - the point for coming up with different ideas for gamification design

- design – the implementation of the gamification design is decided on based on the ideas from the previous phase
- implementation – the design is implemented appropriately
- evaluation – corrections and improvements to the implementation are considered
- monitoring – the development continues based on the effect that gamification has [31]

What this model shows is that after understanding and analysing the desired goals and task at hand, there are four steps that can be looped to make gamification development: ideation, design, implementation and evaluation. After the design implementation is evaluated it can start over from the ideation state to improve on the now existing design. With a well-implemented pilot testing of the design even some aspects of the monitoring phase (mainly the end-user feedback) can also be implemented as a part of the development loop.

2.2. Evaluation of gamification benefits

As stated in the previous chapter, gamification has, in the short time it has been in the public eye, become an interesting research subject in multiple fields. As it has only been around for less than two decades and has properly surfaced into public consciousness less than a decade ago, both the content and the quality of its studies are reflected on with interest and tenacity. [3][5][10]

The previous research in the field of gamification has had a noticeable focus on educational applications [3][5]. To be fair, the field of education is, as stated above, an excellent target for applications of gamification as one of the consistent problems in education is student engagement and motivation [32].

To develop the evaluation of gamification benefits we will survey the previous work in the field of gamification research. Through this overview we will acquire a better understanding of how gamification is commonly measured and what shortcomings others might have recognized in their own work or in the works of others. Another point of interest is the level to which the study methods are applicable to our own research scenario.

2.2.1. Quantitative versus qualitative evaluation

In gamification studies, there are many ways to categorize the study itself. Two aspects of how gamification affects a task are the psychological and the behavioural outcomes of gamification. Behavioural refers to how gamification affects the end user's task execution. This can be (and usually is) studied through measurable qualities of the gamified task, such as task execution times and the quality of the results of the task. This means that those studies are usually quantitative by nature. Behavioural studies are relatively straightforward as the measured results can be evaluated and compared using statistical analysis. [3]

The psychological aspects of gamification refer to how the end users perceive the effects of gamification and usually include questionnaires or interviews. These studies are qualitative and their interpretation, while not always straightforward, have been extremely fruitful for gamification studies. [3]

There are many notable details that need to be brought up considering both ways of conducting gamification research that have been mentioned by Hamari et al. [3]. Often quantitative papers on gamification are descriptive and do not actually offer profound insight into how big of a difference gamification actually produced [3]. Descriptive statistics is fine as indicators of general performance and often adequate to demonstrate that the task is accomplishing its intended function either well or poorly. However, we consider that in there is interest in an enterprise setting to also differentiate the effects of gamification from the general effectiveness of the task as a whole.

Qualitative studies of gamification's effects on user motivation have commonly brought to attention both positive and negative user perceptions regarding gamification [3] and that more often than not these notions contradict one another. By this we mean that the same aspects of gamification liked by many are often also disliked by other end users [3]. Although this is an interesting result and provides fundamental insight into the nature and problems of gamification, it also means that qualitative tests might not always provide an unambiguous confirmation or denial for a study hypothesis. This is not to say that they are in academical sense without worth, far from it, but in an enterprise setting this ambiguity is not desirable as it makes the decision-making regarding gamification hard and unclear.

It is common for an academic research to measure both quantitative and qualitative (in general meaning simultaneously behavioural and psychological,

respectively) aspects of gamification. These studies are usually referred to as having mixed methods [3] [10]. This seems to us a highly effective tactic as it could help to define the relationship between the psychological and behavioural effects of gamification, either demonstrating underlaying correlations between the two factors or revealing contradictory dissonance between them.

Although we did chastise the usage of qualitative evaluation only namely because of the possibility of contradictory results that might produce unwanted corporate uncertainty and frustration, it is our firm believe that leaving it out altogether would be counterproductive as to the interest of the company. After all, we do recognize that even though the qualitative tests might produce mixed results, the quantitative tests might grant us with definitive results. In this manner, they might compensate for the lack of definitive answers by the qualitative tests. On the flipped scenario, if the quantitative studies produce, in a corporate mindset, unsatisfactory results, the qualitative studies might imply a clear shift in the employee motivation which might in turn compensate for the disappointment of the quantitative results. By unsatisfactory results we mean results that are neither positive or negative, thus providing no help for corporate decision making. From an enterprise standpoint, probably the most anxious result would be that neither quantitative or qualitative tests produce satisfactory results thus giving no clear guidance as to whether gamification was worth the investment it took or not and should the applied gamification be reversed or embraced.

2.2.2. Gamification evaluation in practice

Gamification has been studied on multiple different subject matters. By taking a look into how those studies were conducted we might come up with examples on how to implement gamification evaluation. Studies focusing on education gamification make up for a major portion of gamification research [3] [5] [18]. This is understandable, as education measures student performance constantly via exams, grading and other means like the percentage of students that have passed the course. This means that quantitative studies should expose the positive or negative effect of added gamification when comparing post-gamification results to their recorded pre-gamification counterparts. These studies most often focus on the impact gamification has on the student's grades, with the second and third most researched aspects being motivation and satisfaction [5]. It is not uncommon for these kinds of studies also to look into student interaction [5].

This might provide an interesting point, as increased interaction might imply better engagement.

In many other areas where gamification is applied, the end goal is to encourage end user participation and change behaviour [10]. Especially end user participation is a point of interest for marketing, online communities and social media platforms. In these settings, end user interaction and its development are relatively easy to monitor; the end user is not commonly an employee of a company, but first and foremost a consumer. The related applications have been built for gathering user data effectively precisely for that reason. However, when the focus shifts into employee gamification, things might not be so straightforward.

All in all, literature reviews provide plenty of criticism as to the current state of gamification research. Hamari et al. [3] states that gamification research tends to suffer from small sample sizes, where the amount of samples for testing is around 20, and time frames are very short. They speculate that in the latter case novelty might have skewed the data gathered from the experiment. Sometimes the performed experiments also lack proper control groups to compare the gathered gamification data to and many studies presented only descriptive statistics. Hamari et al. also observe that the reporting of results in gamification studies tend to lack clarity [3], which affects their general usefulness. These points are something that we need to address when describing our evaluation methods.

2.2.3. Evaluation methods for the research

In order to evaluate gamification benefits in a business environment it would be beneficial that the metrics being used would provide more than mere descriptive data. Although descriptive data might be interesting and even useful, it leaves the objective effects of gamification up to the interpretation of the researcher. If the interpretation cannot gain definitive results for either positive or negative effects of gamification, it can reasonably be assumed to have less informational value for any corporate decision making. Because of this, in an ideal setting we will need data to which compare our gamification results. If there exists previously collected data of the target task before the gamification process, that is also usable for comparing the gamified task and analysing the results of gamification

In this thesis, the research is carried out as a case-control study in which the control group will provide data that can directly be used to compare gamification

results. There were numerous other reasons for this, the first and foremost being that the gamification projects serving as the study targets were easy to convert into non-gamified versions and were thus well suited for it. For both tests a ready-defined pool of test subjects was also available: for the first test case they came from the customer and for the second they were in-house (by NordicEdu and its parent company). The customer of the first test case is a big enough organization for their employee count to facilitate the required test subjects. The decision was further affirmed by Hamari et al. [3] who criticized some experiments in their literature review as lacking control groups and relying solely on user evaluation. As stated, we consider the case-control study effective as its results can easily and immediately be interpreted based on comparisons to data from the control group.

Based on the theory, we have decided to gather mainly execution time and task result evaluations as data of user performance during testing. After testing the user also fills in a form with questions regarding their motivation toward the test, evaluation on its effectiveness and evaluation on the functionality of the task. The user evaluations of the effectiveness of the gamified solutions and the motivational stimuli they produce are the main point of interest for the qualitative part of this study. In the first test case (the one for the customer) there are also other questions that hold no meaning to the experiment itself. These are however useful to NordicEdu and the customer as they provide valuable information for the future development of the solution.

The research will thus be carried out as a mixed methods research. This seems sensible because, aside from the reasons stated above, it allows for the measurement of three different aspects of the gamification, namely its effect on the time spent by the end users on the target task, its effect on the results of the target task, and the motivational factors that affect the user during the testing. These are fitting metrics for evaluating gamification benefits in an enterprise scenario. The two former metrics are tangible in the sense that they are telling of direct benefits to the business of a company. The measurement of motivational factors affecting the end users is also a point of interest as improved employee engagement indirectly affects employee turnover and a multitude of other things.

In case of LOGE and its evaluation, there exists some overlap with the test participants. This is to say that a certain amount of the participants take part in the test twice through different discussions. In these cases, it has been made certain the groups are switched so that most of the participants that were in the control group are switched to the test group and vice versa (further explained in Chapter 3). Though this might

sound problematic, it provides us with an interesting case, where we have a small group that is has been through a non-gamified and a gamified discussion. It would be interesting to ask them to evaluate the differences, benefits, and disadvantages between the non-gamified and the gamified solution. This might provide interesting insight into the question: how well subjective comparison work does when evaluating gamification benefits.

2.2.3. Recognized limitations of the study

The time frame for the research has been decided to be roughly four weeks for both projects. For the scope of this thesis it is not profitable for us prolong the experiment for any lengthy amount of time. Four weeks has been deemed adequate for enough volunteers to participate and considering the gamification targets, which will be further elaborated on in Chapter 3, this should not be a problem. Although it is a valid concern that too short of a time frame for the experiment may lead to distorted or skewed results due to novelty being a major factor in user experience, this should not be a problem for either of our two test cases. This will also be further elaborated on in Chapter 3.

3. The gamified solutions and how to evaluate them

In this chapter, we present both test projects and their gamification solution. In the case of the unnamed customer project we also go through the development process of said gamification project. We describe the research plan to measure the effect of gamification for both solutions and hypothesize on the effectiveness of the evaluation. We also look at how the research went and what problems we faced during it.

"*Gamification requires holistic information system design; taking into account not only the stellar technical aspects but also the manifold and multidimensional aspects of user psychology and engagement*."

- Morschheuser et al. [31]

3.1. Customer project

3.1.1. Requirements and constraints for the project

As stated in Chapter 1, the customer ordered two gamification projects from NordicEdu. In both projects the objective was to develop a gamified approach to the training of customer service employees of the customer. The first project and the evaluation target for this thesis had to do with the customer's renewed e-commerce system. They wanted a gamified solution for the purpose of getting their sales personnel and customer service employees acquainted with its new properties and practices that were related to working with it. The latter also included the conventions and regulations that apply to the handling of orders that come through the new system.

In this project, the customer knew exactly what the information content of the new solution should be. They also had clear guidelines on the general appearance of the solution due to their established brand. We will not elaborate further on either of these due to the anonymity of the customer. The customer had no fully set model for a gamified solution, so NordicEdu had relative freedom to design the way the solution should present the information and gamify the training.

The customer wished for the solution to function in their own learning management system (LMS), as a SCORM package [33]. SCORM comes from the words Sharable Content Object Reference Model and is a standardized way to define a packet for LMS environments like Moodle. This posed no noticeable restrictions to the

gamified solution, although the development team had to define what data needed to be stored by the package and at which points of the training they should be communicated to the LMS. The customer had no other requirements for the stored data except for the point in which the LMS should consider the training passed. This was defined during development.

The project itself was technically straightforward. The solution was developed as a simple web page (HTML, CSS, Javascript), which was then packed into a SCORM package. For testing purposes, a simple and ready-made Moodle environment was used to check that the SCORM package functioned as it was supposed to.

The biggest single constraint for the project was time. The renewed e-commerce system was under development during the beginning of the project and the training was scheduled to be completed around the same time as the new system would be made public. This left NordicEdu with approximately two months to develop the gamified training solution.

3.1.2. The development of the project

The NordicEdu's development team for the target gamification project consisted of four people (although a few more people took part in the development during the process): two software developers, a graphics and user experience designer (UX), and a project manager, who was also in charge of the customer communication. The whole development team took part in the gamification ideation and the design of the final solution.

The solution the team decided to create is essentially a simple multiple-choice quiz about practices concerning the new e-commerce system, its uses and functionalities, as well as how it should be utilized in a customer service setting. The questions are divided into different categories based on what aspects of the system they focus on. These categories are presented so that once the training has begun, all categories are visible to the trainee but only one category is available for answering. The other categories open one after another to be answered by the trainee as the trainee goes through the categories and answers the questions in them.

After additional refinement, the solution became such that in the beginning the trainee is taken through a brief starting info that tells them how the gamified training proceeds. After that they are taken to a main view, where they can start any category of questions that they have unlocked. Unlocking the categories works as described in the

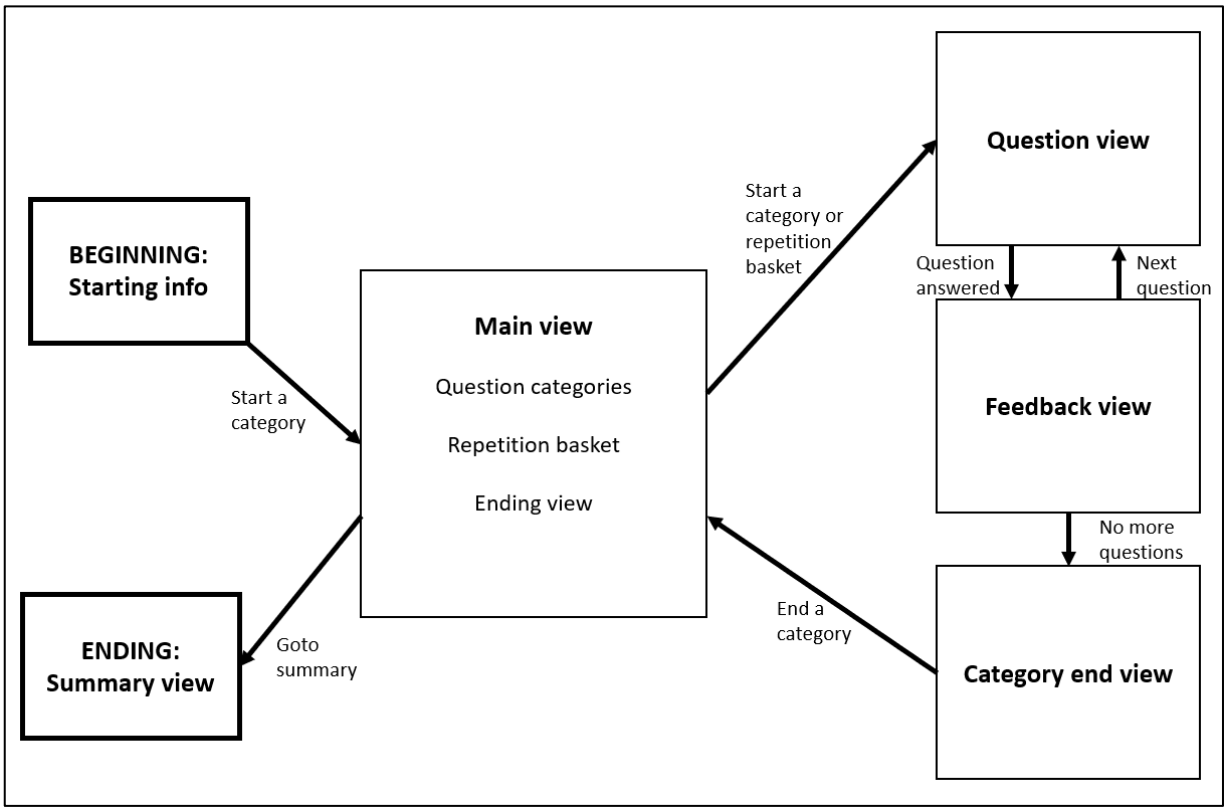

Figure 4 Flow chart of the gamified solution

previous paragraph. Once a category is entered, the trainee is shown multiple-choice questions that they then answer. In this view, they can also choose to use a hint before answering the question. This has no negative consequence on the advancement of the training. After every answer, they are shown whether the answer they gave was right or wrong in a feedback view. At the same time, they also receive additional information and, if they answered wrong, the information on what the correct answer would have been. At the end of the category, the trainee is shown how well they did in a category end view. This is shown as a score of how many questions they got right in that category out of all of the questions in the category. They are also rewarded with a humorous picture that reflects how well they did. Once all the categories have been passed, the trainee enters a summary view that shows them certain statistics of their training process and also awards them with a celebratory title and a humorous picture. This also marks the training as passed in the customer's LMS. The flow chart for the gamified training process is further visualized in Figure 4.

Once a category is completed (i.e., every question of the category is answered) the next category opens for the trainee to complete. The next category opens up no matter how many questions the trainee answered correctly in the previous category, even if they got all of the questions wrong. However, the category is not considered passed until every one of its questions has been answered correctly. The state of the

category is visualized by an intuitive colour-coded circle beside the category's name in the main view. If the circle is red, the category has had a very low percentage of correct answers while a yellow circle means that the trainee got relatively close to passing that category. Once the circle turns green, it means that all questions in that category have been answered correctly and the category is passed. This completion visualization system became referred to as the "traffic lights" by the development team.

The requirement for the category to be passed came from the customer. This was done to make sure that all of the necessary information had been internalised from the questions by the trainee. However, the development team recognized that if the end users were forced to answer all questions in a category in order to retry one that they got wrong, it would result in extreme amounts of frustration and decreased motivation. In order to avoid the negative effect this needless repeat would have on the trainees, the development team decided to use a "final gauntlet" type element, they named as "the repetition basket".

The repetition basket functions in a way that as the end user goes through the categories answering questions, the solution keeps track of all the questions that the user gets wrong. A basket button was added to the main view showing the amount of questions that are being tracked. Every time a question receives a wrong answer, it will be added to the tracking list. If a trainee retries a category and gets a question right the second time, the question will be removed from the repetition basket. Once all categories have been answered, the end user may enter the repetition basket. In the repetition basket all tracked questions will be asked from the end user, and the questions that they get right will be removed from the list. Once all categories have been completed and the repetition basket is empty, the final summary view is made available. This way the training ensures that the object of getting every question right is fulfilled before the training can be completed by entering the summary view.

During the development process the development team organized a piloting period with the customer that lasted approximately ten work days. In the beginning of this session, the development team presented an alpha version of the gamified solution to a piloting group selected by the customer. They tested it and reported criticism and general development suggestions to the development team. The handling of the feedback was done through a formulated Google Forms questionnaire and an open Google Docs document. When around two thirds of the piloting had been done, the development team released a new version of the solution which contained improvements based on the feedback to the piloting group. At the same time separate

Google Forms questionnaire and Google Docs document were opened in order to gather separate data from the improved version and of the piloting process itself.

Based on the feedback the piloting process can be evaluated to having had a huge impact on the product itself. Both times the questionnaires contained the same questions regarding the solution itself and the second form had additional questions regarding the piloting process. Both forms had the same number of participants at the time of summarising of the piloting results and, upon closer inspection, the two datasets revealed great improvement on the reception of the whole gamified solution in general. Especially both the clarity and mood of the solution improved drastically during the piloting process. The three categories that showed the least amount of improvement (from none at all to very little) were the hints, the humorous pictures and the starting info. The one category that received unanimous praise during both rounds was the repetition basket. In the end, most of the participants in the piloting group felt that the piloting process had been effective at least to some degree and that it had given them a chance to contribute and impact the development of the solution in a meaningful way.

After the piloting process, the gamified solution experienced very few major changes. Most of the changes relating to it were either cosmetic, slight alterations to the questions and their content, or fixing errors in the solution's functionality. The solution itself was delivered to the customer for publication in their LMS in January 2019.

3.2. LOGE

LOGE's initial development has been ready and the first versions of it have been available for customer use for quite some time before the research for this thesis was started. During that time it has been consistently updated and developed on by NordicEdu in association with Talent Vectia. [34]

LOGE's essential functionality that it is a tool to organize, structure and steer team discussion. The aim is to make business meetings and other similar discussions more effective through gamification. There are three main features that are important for us to present before we go into our research plan:

- how a discussion session is created in the admin view, so that we are aware of our limitations while constructing a session for our research
- the discussion session and its functionality, as this is the focus of our research considering LOGE

- the feedback functionality, as it is a valid option for collecting feedback from participant in our test

3.2.1. The admin view – creating a discussion and observing feedback in LOGE

LOGE's admin view is a comprehensive tool for controlling the organization-specific content of LOGE. Through it we can create LOGE discussions, which are the most essential functionality of the software. The relevant terms in LOGE while creating a discussion session are theme, question set, and season. In order to have a discussion in LOGE one first needs to create a question set with a theme. After this, a season with one or more scheduled discussions sessions is created. During creation, the season is also assigned a theme. The theme determines, which question set is used for the discussions. From this set a structure for the discussion will be formed.

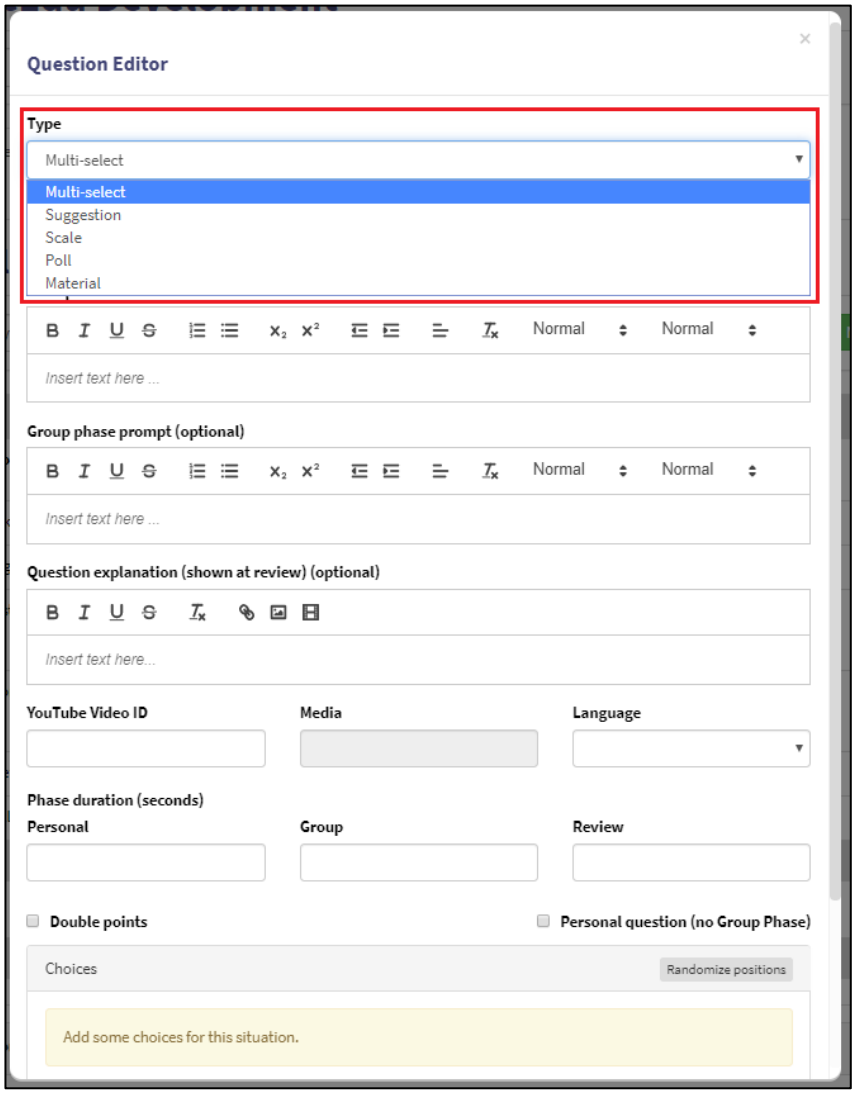

Figure 5 The question editor for Loge admin view. The different question types are in the dropdown marked with red.

The question sets for discussions in LOGE can contain multiple different question types to support different types of needs during discussions. The question types support ideation, decision making and evaluation in a varied way, depending on how they are formatted (see Figure 5) and a time limit can be set for all of them as the organizer of the discussion sees fit. A short description for each question type is as follows:

- **Multi-select**: a multiple-choice question with predefined options. Each option can have predefined feedback for the participants and a score value. During a discussion each participant is also offered a chance to write a short reasoning for their answer.
- **Scale**: either a simple slider question or a two-dimensional one (visually resembling a scatter plot). Chance for multiple options that can be predefined or polled from the participants.
- **Poll**: a multiple-choice poll with predefined options, with a fill-in "other choice" as an extra option. The way the poll results are shown can also be defined as either bar, pie, or circle.
- **Suggestion**: a suggestion with an open answer field for comments. Also usable as a question with a field for a written answer.
- **Material**: A way to show audio-visual material (eg. a video) as a part of the discussion.

In LOGE it is possible to also predefine other materials for the theme. More interestingly, self-evaluation questions can also be set for discussion participants to answer after they are done with the group discussion. The evaluation questions can be one of two types: rating or open field/written. The rating questions pose a question with the answer being a value between one and five.

A season in LOGE is a simple way to define a set schedule when discussions for a certain theme are held. Discussions in a season are adequately customisable. A maximum amount for both questions used from the set and the maximum participants for participants of a discussion can be adjusted, times when discussions are available can be chosen, description for season can be written, and time limits for the discussion parts can be set. The season ends automatically after the last scheduled discussion has been held or time for it has passed. This is further elaborated in Figure 6.

As implicitly stated, the feedback received from not only the self-evaluation questions, but also from the discussions themselves. We can also view statistics of the

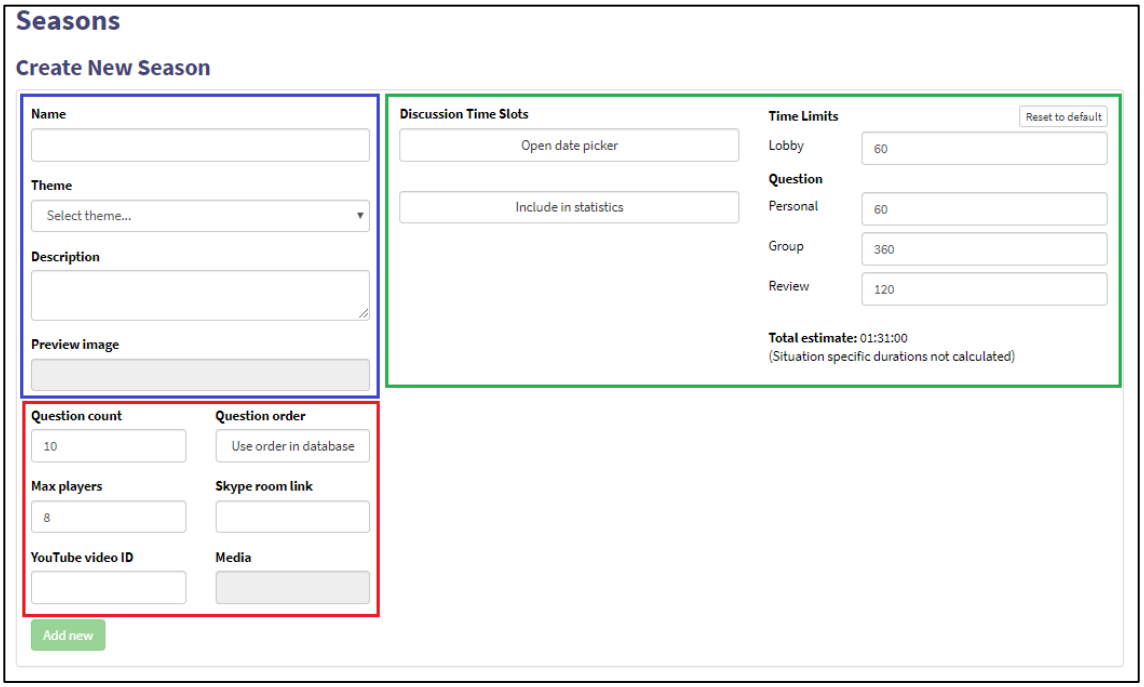

Figure 6 Loge Season creation options. In the blue are highlighted descriptive fields and theme selection, in the green time related fields, and in the red are fields affecting the discussion format

discussions and gather the discussion time and what points the participants got from the discussion.

3.2.2. The gamified discussions in LOGE

The discussions in LOGE are guided by the questions in their theme. The length of the discussion depends on how many questions from the theme the discussion should cover and the amount of time (divided to its phases) each question is given during creation. Each question has three phases: a personal phase, a group phase, and a review phase. During the whole discussion the participants are at the same chat room with both audio chat on and text chat open in the sidebar. The progress of the discussion is shown in a progress bar that is constantly visible. At the beginning of each discussion the group chooses its name by polling for suggestions from participants and then having them vote for the best name.

The handling of a single question phase by phase is a straightforward process. During the personal phase each participant is posed a question defined in the theme. The participant answers said question and has, in most cases, a chance to give a short explanation for why they answered the way they did. After that the discussion moves on to the group phase of the question, during which all the personal answers are discussed and during which a randomly selected chairman compiles a collective answer for the question based on the personal phase answers and the groups discussion. At the end of a

question comes the review phase, during which the answer of the chairman is revealed to the group, and the answer can be discussed along with any other discussion brought up during the question that needs to be addressed. Examples of these phases for an example scale question are shown in Figures 7, 8 and 9.

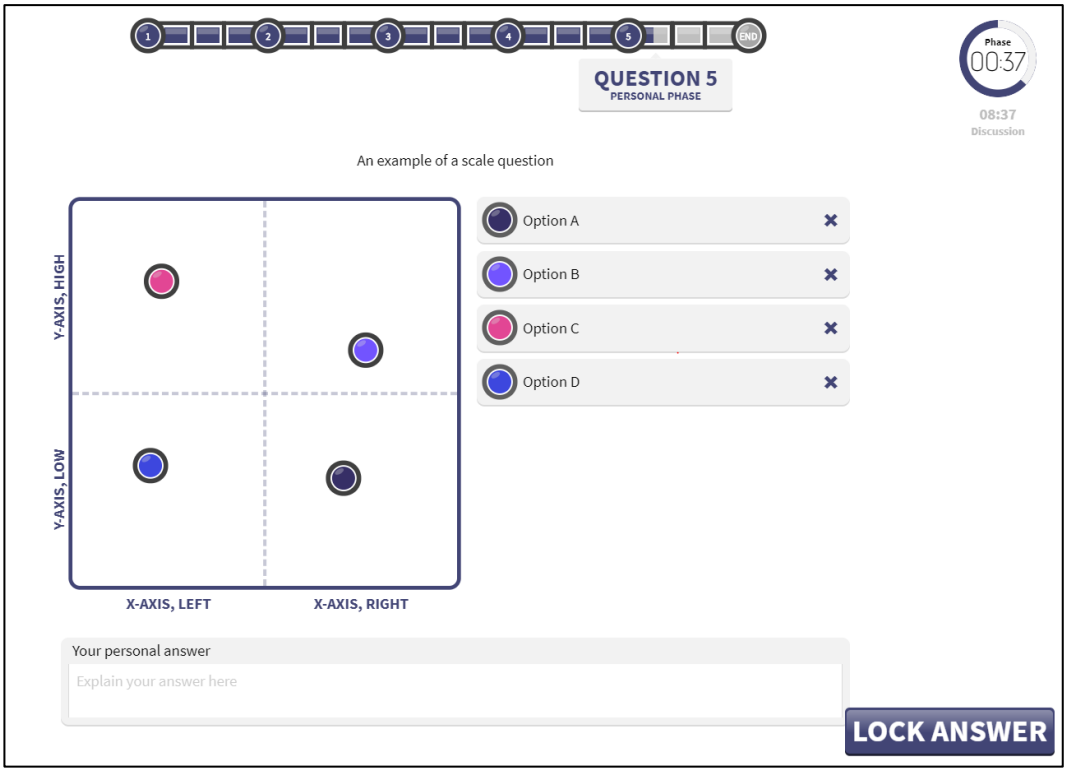

Figure 7 The personal phase of an example scale question answered. The balls representing the different options have been dragged to the matrix by the participant, who has decided to give no freeform personal answer.

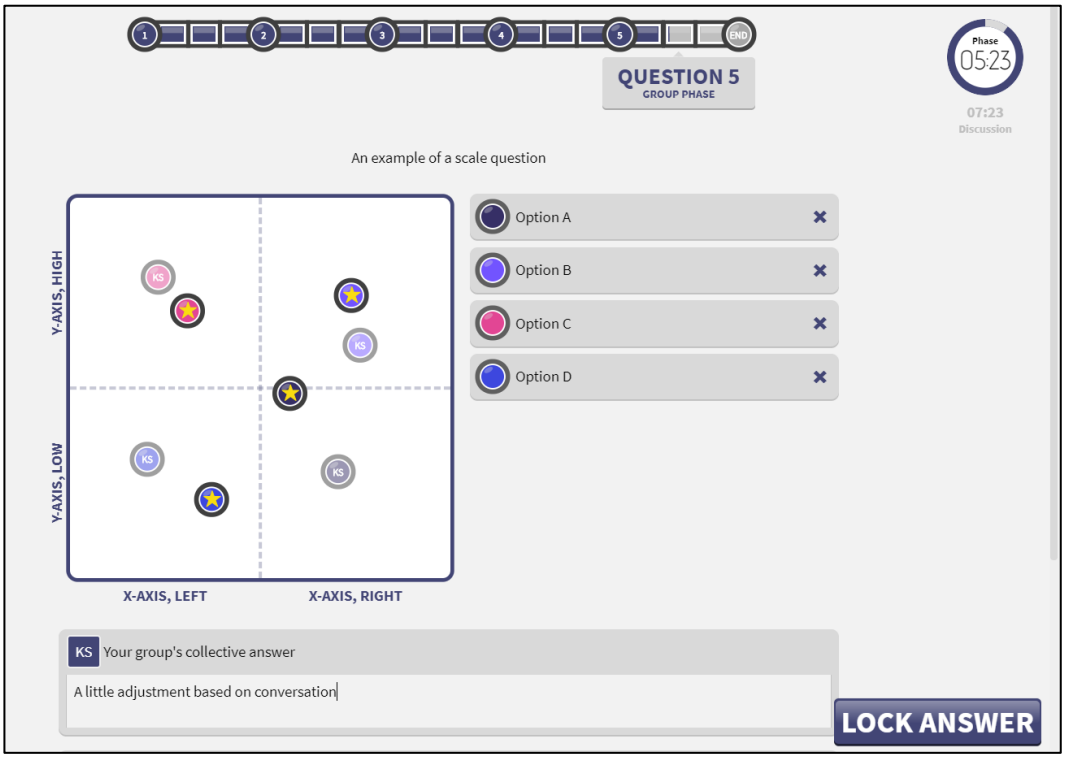

Figure 8 The group phase after the phase in Figure 7. The chairman has adjusted the positions of the groups final answers and described the answer further in the text box.

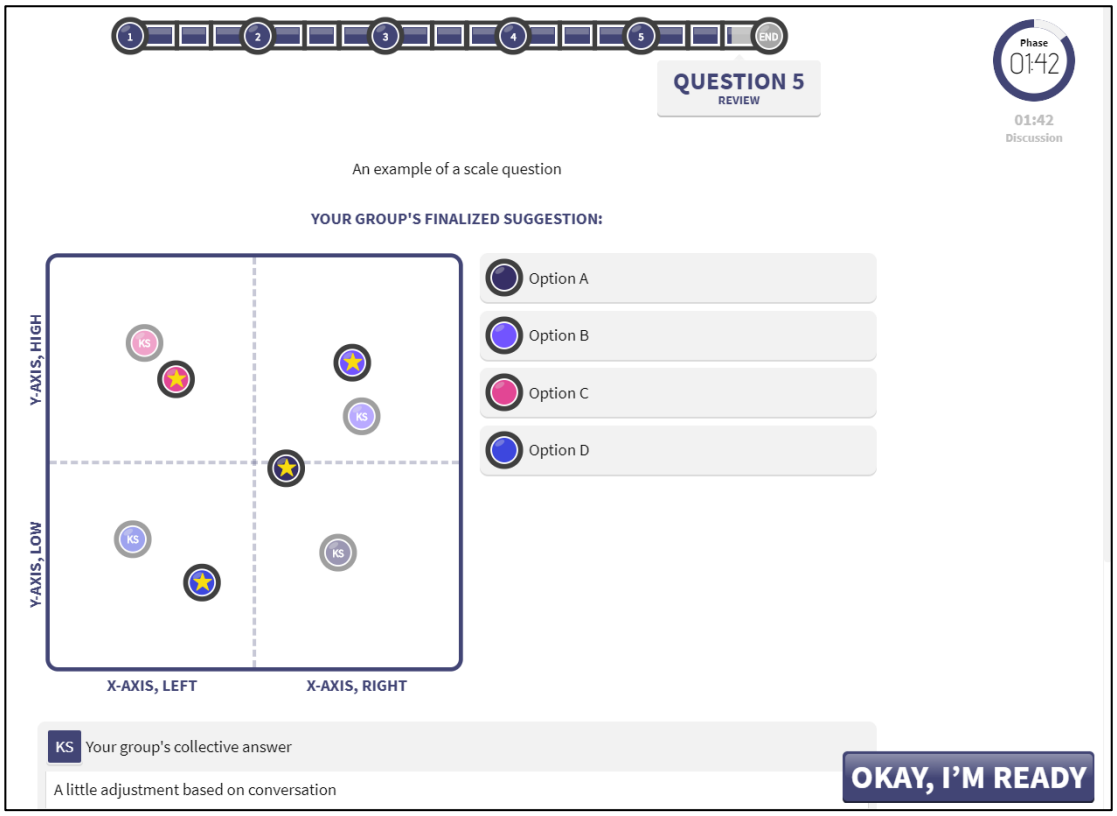

Figure 9 The review phase after figure 8. The answer of the group is revealed, and it is discussed.

After all of the questions have been properly addressed the discussion ends and a score card for the discussion is shown to the participant. Most question types have a set amount of 25 points that the participants get by answering. In the case of multiple-choice questions, points can be addressed to individual options according to the organizer's liking.

After the participant has had enough of the score board, they can move on to the self-evaluation portion of the discussion. If the organizer has defined questions for this part, they are shown along with two default questions that are always shown even if no other questions have been defined. These two questions are an open field question with a prompt for open feedback and a rating question which asks the participant how they would evaluate their LOGE experience. The scoreboard and self-evaluation following a discussion are illustrated in Figures 10 and 11.

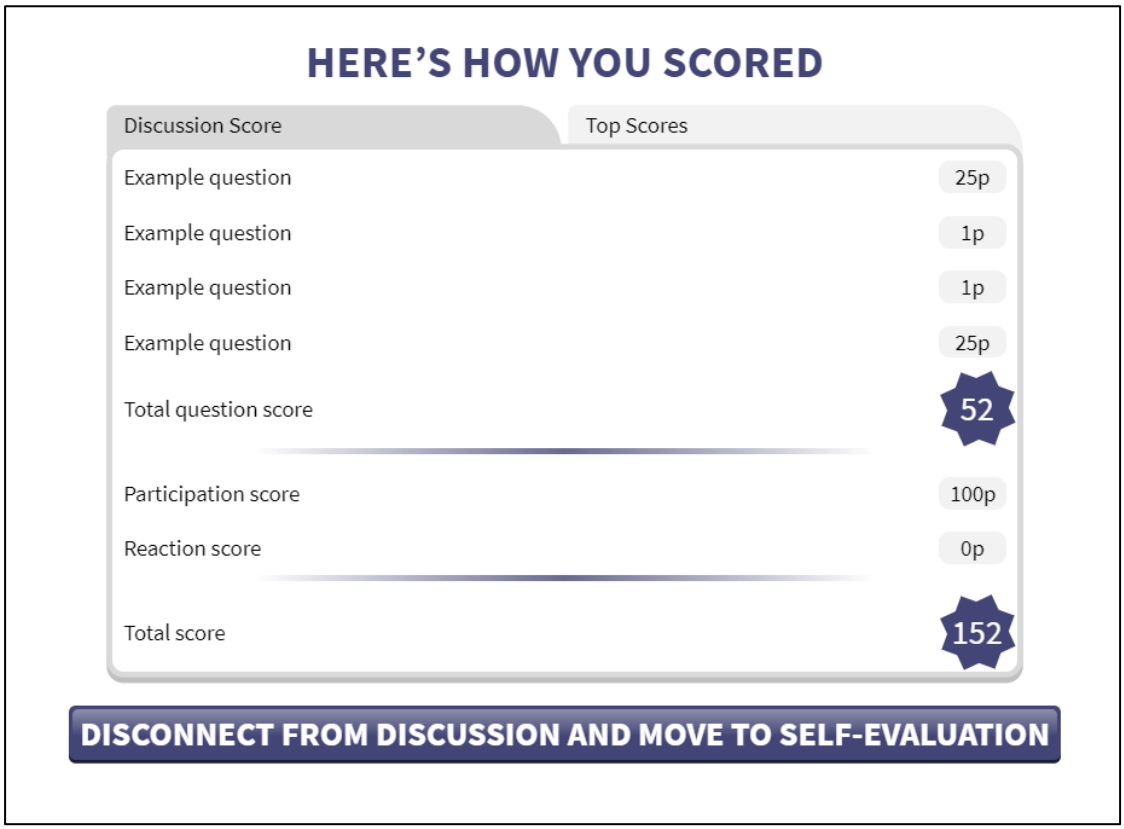

Figure 10 An example of a personal Loge score board after a discussion.

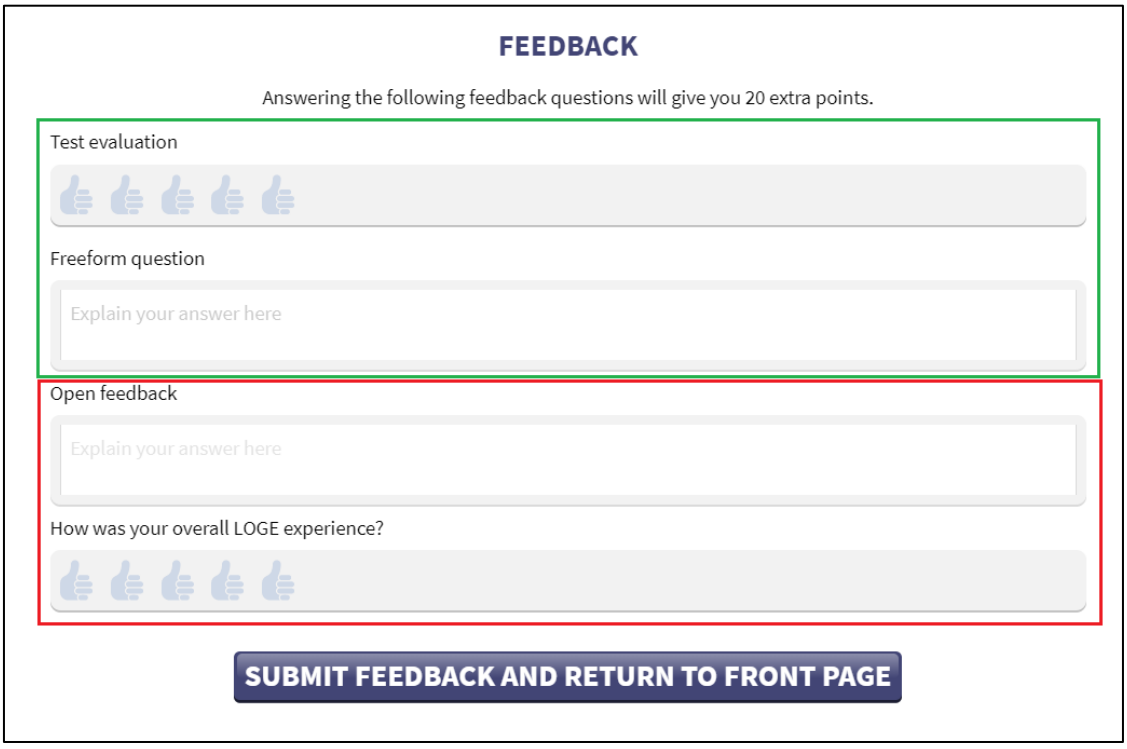

Figure 11 An example of a feedback questionnaire after a Loge discussion. The organizers questions are highlighted in green, the default questions are highlighted in red.

After a discussion in LOGE has ended, its results can be downloaded from the admin view as a CSV-file, which contains the data for all the interactions during a

session. In addition to this, the suggestion, scale and poll questions have a way to view them from the admin view, which makes the results much simpler to interpret. The feedback given for each discussion can also be examined in a separate view. However, if a participant has chosen to give no feedback, the view shows nothing to examine.

3.3. Research plan

"*As the research on gamification progresses, care should be taken to ensure that future results are more comparable.*"

- Hamari et al. [3]

As stated in Chapter 2, the research for both cases will be carried out as case-control studies. However, because of the differing nature of the test subjects, it is easier to present the research plan for both of them separately.

3.3.1. Customer project

For a case-control study to become possible, we will need to find a way to create a version of the training that has all the same information content as the gamified solution, but none of the gamification aspects. Then we need to device a way to test

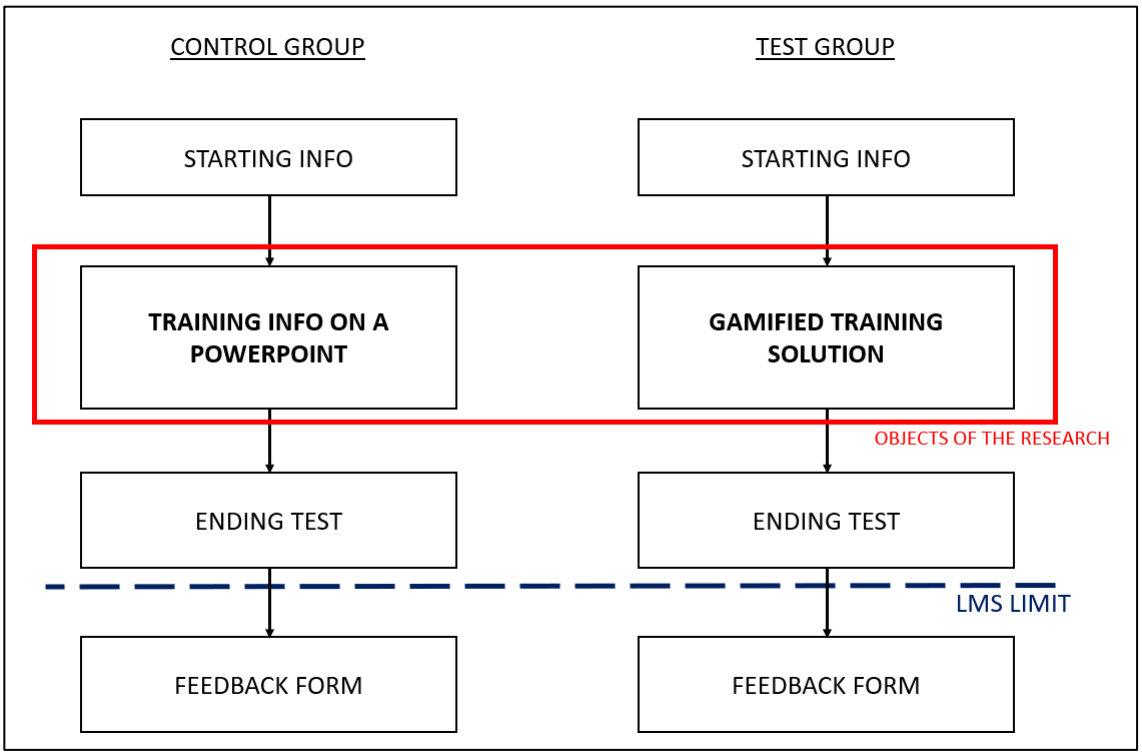

Figure 12 The flow of the study for both the test and the control group. The customer's LMS limit is marked by the dashed line and the objects of the research by the red rectangle

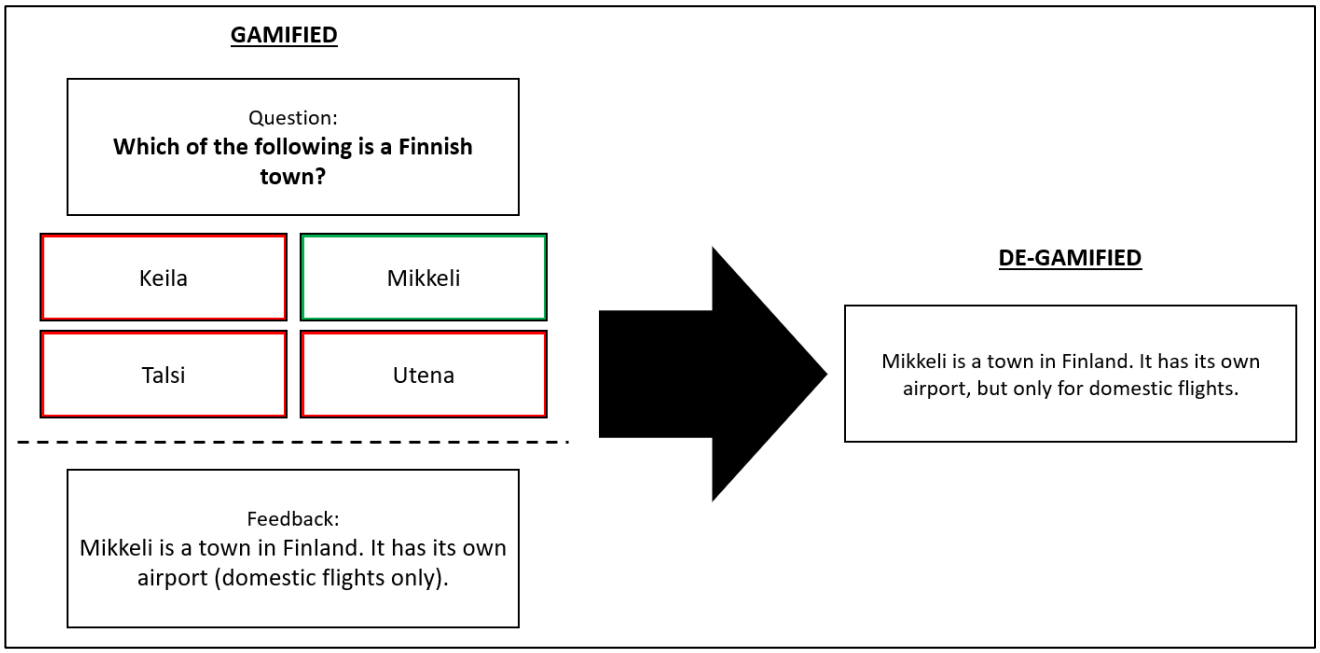

Figure 13 An example of the formatting of the training information. The correct answer highlighted in green.

both the test subjects and the control group in a way that allows us to easily compare the results produced by both groups. Lastly, the testing environment (aside from the aspect of gamification) should change as little as possible between the two groups.

The latter of the three points proved to be the simplest of the four to answer. In order to assure minimal change in setting for both test and control group, we decided to utilize the SCORM package and the LMS of the customer. As the original gamified solution was itself a SCORM package, it lent itself well for modification. By doing the test through the LMS of the customer we not only had a ready-made and easy to use way to have volunteers participate in our study, but also a set format for how the study would be structured. A SCORM package can be divided into multiple different parts. As the development team had created the training solution as the only part of its package, new parts could easily be added without disrupting the functionality of the original solution. We added two more parts to the SCORM package to frame our study: an introductory start and a test to the end of the package. In order to gather the information we need, we decided to use Google Forms. Once the end of the SCORM package is reached, the end user is instructed to press a button that will take them to a Google Forms questionnaire. The data required for the quantitative research is filled in automatically and the participant answers the questions for the qualitative gamification evaluation. The flow design for this research and the SCORM structure is further elaborated in Figure 12.

The ending test for both the control and the test group are to differ from each other in no respect, as they are crucial to measure differences for the quantitative

research. The starting information parts should similarly resemble one another as closely as possible, although some deviation might be necessary in order to avoid falsities in the instructions. In the same vein, the feedback form for the test group will contain all the same questions as the form of the control group. However, the form of the test group will also contain additional questions that are specific to the gamified solution developed by NordicEdu. The qualitative questions for both Google forms are presented in Appendix A.

In order to create a valid version of the training material that was not gamified we had to decide on what information in the gamified solution was meaningful for the purposes of the training. In the end, the meaningful data conveyed in the gamified solution came from three different elements in the solution: the questions, the answers and the feedbacks after the questions. To neutralize all gamification all of the mentioned parts were read through while at the same time copying the questions, the correct answer options and the feedback extra information to a separate file. The copied data was once again read through and formatted as a structured informative text (see Figure 13). This text is then divided into otherwise plain presentation slides with titles matching the gamified question categories. The slides were saved as a PDF file, which made the implementation very simple. PDF files as a part of a SCORM package work just like an ordinary PDF, which means that the slides did not need any special implementation in the LMS.

For the ending test we decided to use a similar question format as in the gamified solution. However, we came up with a completely new set of questions that could in theory be answered correctly by a participant whether they had gone through the gamified training quiz or the de-gamified training slides. The ending test comprises 20 questions out of which the first 5 are binary true-or-false questions and the 15 following are similar multiple-choice questions as most of the questions in the gamified training. While the gamified training contained some questions with multiple correct answers, in the ending test every question has a solitary correct answer. We defined a pass limit of 16 correct answers for the ending test. Although the test does not provide any sort of feedback after every feedback, at the end the result is shown as a number of how many questions out of 20 the participant got right. If the pass limit of 16 is surpassed a confirmatory thumbs-up symbol is shown, whereas a thumbs-down tells the participant that they could have done better. At the same time the participant is given the options to re-do the test or continue on to the Google Forms questionnaire. The test can be taken as many times as the participant likes.

Once the participant decides to move on from the test and into the questionnaire the ending test automatically fills in the data for the quantitative parts of the form. For the quantitative research, we decided on a few key factors that interested us in our study:

- starting time of the training
- starting time of the ending test (the first time it is entered)
- time of the first passed attempt of the ending test (if it is passed)
- ending time of the training (the time Google Forms is entered)
- amount of correct answers for the first attempt of the ending test
- amount of correct answers for the first passed attempt of the test (if an attempt is passed)
- the highest amount of correct answers gotten from an ending test attempt
- amount of times the ending test was begun
- amount of times the ending test was completed
- number of the attempt that the ending test was passed for the first time
- number of the best attempt at the ending test

From this data we should be able to notice some correlation between the effects of gamification on the training. The factors should also be sufficient enough so that even a lack of correlation should tell us something about the effect the gamification has had on the training.

Technically, the study for the customer was not difficult to facilitate. As the ending test used many of the same assets as the gamified solution and the starting information was modelled after its starting information view as well, the bulk of the work could focus on other things. Most of the work went into refactoring the gamified training into a non-gamified slide show and devising the way that the quantitative factors were stored and delivered to us. This too borrowed many technical solutions from the gamified solution which lessened the amount of work it required. Altogether, the research materials took about a week's worth of effort from us.

3.3.2. LOGE

With LOGE we went differently about creating the test material. As LOGE is a platform where the discussions can be freely defined, the question became more about what subjects we should discuss. From NordicEdu we got two subjects, one for the technical personnel of the company and one for everyone in the company. These discussions are

from here on referred to as the technical discussion and the all-encompassing discussion. From NordicEdu's parent company we also got one subject for a test case. This will be referred to as the parent company discussion.

When it came to creating test cases for all the discussions, the procedure was straightforward. First, we defined the general objectives for the discussion. After this the objectives were defined into the form of a coherent and more controlled LOGE discussion. The participants of each discussion were then divided into two groups: one that would use LOGE (i.e. the test group) and one that would have a traditional freeform discussion (i.e. the control group). For the freeform discussion a chairman was assigned who understood the objectives and time limits of the conversation. This was done in order to ensure a formal discussion with purpose. After both discussions the participants were asked to answer to a questionnaire about the discussion and how well they thought it played out. The control group filled the questionnaire via Google Forms, whereas the test group utilized the internal evaluation system in LOGE. These evaluation questions are further elaborated in Appendix B.

With LOGE the only quantitative metric we have chosen to measure is the amount of time taken by the discussions. However, in addition to the evaluation questions posed to all participants individually, we also ask the company facilitating the test (either NordicEdu or its parent company) to evaluate the usefulness of the results of the discussions and decide on which between the control group and the test group produced more useful results for the company. Though this cannot be considered quantitative, we consider this a straightforward and valuable metric when evaluating the effectiveness of gamification in a single LOGE discussion.

As stated in Chapter 2, we have some overlap in our participants for NordicEdu discussions. This should however pose no problem, as the discussions are separate and individual. We have also made sure that the groups have been assembled in a way that no participant is twice in a discussion that uses LOGE. This has been done to mitigate any effect the overlap would have on the results. It might be worth noting that one participant has been assigned to two discussion groups of which neither utilizes LOGE. This is because of his status as LOGE's current main developer, which might affect his experience drastically. He hasn't been completely excluded from the discussions because the discussions hold value to NordicEdu outside of this experiment, as does his participation in them.

Because of the overlap we have a chance to examine how well subjective comparison does in evaluating gamification. As we have participants that have taken

part in both a non-gamified discussion and a LOGE discussion, we will ask them to compare those discussions and evaluate how effective and helpful they consider LOGE's gamification based on their experiences. The evaluation will be done through a Google Forms questionnaire. The questions in the questionnaire are further elaborated in Appendix C.

For LOGE testing we have mapped out three subjects each with separate groups for those that use LOGE and those that do not. The descriptive numbers of each of these discussions for LOGE are presented in Table 1.

DISCUSSION	Technical discussion	All-encompassing	Parent Company
AMOUNT OF PARTICIPANTS	7	10	6
AMOUNT OF QUESTIONS	9	16	12
OF WHICH PERSONAL OR DESCRIPTIVE	2	3	0
AVG. PERSONAL PHASE TIME LIMIT	03:27	04:28	04:55
AVG. GROUP PHASE TIME LIMIT	08:34	09:18	09:50
PREVIOUS LOGE EXPERIENCE	FEW	SOME	NONE

Table 1: Descriptive attributes of LOGE discussions.

DISCUSSION	Technical discussion	All-encompassing	Parent Company
AMOUNT OF PARTICIPANTS	7	10	7
LEVEL OF GUIDANCE	BROAD	BROAD	DETAILED

Table 2: Descriptive attributes of non-LOGE discussions.

In Table 1 a personal phase only-question is a question that does not have a group phase. Of these 5 questions, 3 were used to set discussion context, 1 to offer a five-minute coffee break, and 1 to map out the personal interest of discussion members to further participate in development of the discussion subject. In the Loge discussions of the parent company the participants had no previous experience with LOGE, whereas in both NordicEdu's discussions at least few of the participants had tried LOGE before. In every LOGE group however, at least half of the participants have never used the system before. The group that had no previous experience with the platform were provided guidance regarding how to use LOGE with minimal interference to the actual discussion.

Table 2 describes the precious few attributes the non-LOGE discussions can be attributed beforehand. The level of guidance refers to the description the groups or their chairmen were given for the discussions. In both NordicEdu's discussions the chairmen were given a general description of the subject and major topics they should try to cover

during their discussion (the same topics that were the basis of the LOGE questions for the test groups). However, the parent company felt it more appropriate to give the control group the same questions as the test group but without the gamified structure of LOGE. This makes the parent company's case is a different, but all together valid test case.

3.4. Hypotheses for the research

Before our research begins, we would do well to form hypotheses of the results and the viability of the research itself based on existing theories. In this thesis, we have presented plentiful amounts of theory regarding to our study and so have good grounds to form hypotheses for it.

3.4.1. The customer project

Firstly, we should look at the resources that are at our disposal and factors which contribute to the participation for our study. The test users for the research are arranged by the customer and participate voluntarily. We have laid the absolute minimum number of participants to 20, 10 for both the test and the control groups. This number is low and will affect poorly on the impact of the study should the amount of participants end up remaining this low. The customer is, as mentioned, a large enterprise that has multiple outlets all over Finland and the size of the participant pool we aim for is 40 (20 participants per group). The size of the customer company keeps us hopeful that we could receive more participants than we are aiming for. They also have already in place a functioning LMS that allows every potential participant easy access to attend our study. The participation should take between 20 to 60 minutes time from start to finish, which should not pose a problem for user participation. Our hypothesis of the amount of user participation is that we will reach between 30-40 participants (15-20 per test group).

We perceive no reason that the results of the case-control group test should contradict the previous research considering gamification. This is a very straightforward test of which there have been multiple similar ones before. Thus, our hypothesis is that gamification has a positive effect on the employee training. We can add a conditional hypothesis that should the gamification not have a positive effect on the training, it at least will not have a negative effect on it. Our thesis focuses on the development of the evaluation methods themselves and so this hypothesis is not essential to it. It is noted

here purely for added interest and to demonstrate our pre-disposition in our approach to the test.

When evaluating the research of the customer project, we have recognized multiple factors that affect the credibility of the test results we might receive. These include:

- Group/sample size – 40 participants should offer us an adequate sample amount. If the amount drops, it has a negative effect on the credibility of the results.
- Implementation of the ending test – the ending test has been modelled after the question parts of the gamified training. It might have an unintended positive effect on the test group thus having an effect on the comparability of the groups. However, this effect has been evaluated to be small enough for the purposes of this thesis.
- Participation situation – the employees take part on the test during their work day, which might cause them to rush through it. In ordinary circumstances they can properly take time to go through the training. This affects both the test and the control group and does not thus affect the comparability of the groups.
- Participant pool – the participants are all current employees of the customer. This might have a positive effect on the results of the test. Again, this affects both the test and the control group and does not thus affect the comparability of the groups.

All in all, the recognized elements do not have major meaningful effect on the credibility of the test results. The biggest concern we have is about the sample size, which could in an extreme case drop low (close to 20) and so affect the study. However, since our assumption is that we will have close to 40 participants, our hypothesis is that the test results will be credible.

When considering the relevance of the test results we are talking about their relevance for the customer. The customer is an enterprise and will be mainly interested to see either a change in the performance or in the attitude of the employees, or the effectiveness of the training. All of these aspects are directly evaluated by the test, either by its quantitative part or its qualitative part. Our hypothesis is, therefore, that the results of the evaluation are relevant to the customer. Our secondary hypothesis also is that the results of the test are academically relevant but also less impactful. This reflects

what was stated previously about the straightforwardness of the test and our focus on the gamification evaluation methods used in it rather than its results.

3.4.2. LOGE

With LOGE our test case differs vastly from the customer project. To evaluate the effectiveness of gamification on the results of the discussions we have three pairs of test discussions. The metrics for this evaluation are the time taken by the discussion (quantitative) and evaluation on the usefulness of the results performed by the facilitating company (qualitative). Since we do not have a bigger sample pool for either of these metrics to form any academically conclusive judgment, we must view our results based on them as only indicative results. However, this does not mean that we could not evaluate the used study methods themselves based on these metrics.

For the evaluation of the effect of gamification on user experience our sample pool is much more sizeable, as each participant provides a sample for it. Based on our test plan our control groups will total at 24 participants and our test groups at 23. This is more than a reasonable group size for our test scale and should give our results a certain amount of credibility. However, should something drastic happen and our participation count drop drastically, we maintain the same absolute minimum sample size of 10 as with the customer project. Due to them being sessions that are actively organized by the collaborating companies, our hypothesis is that we will receive at least 20 participants per group.

As with the case of the customer project we have the same hypothesis that gamification will have a positive effect on the discussions, or that it at least will not have a negative effect on them. As with the customer project, this hypothesis is based on previous similar studies.

Identified factors affecting the credibility of the test results for the three LOGE cases include:

- Group/sample size for discussions – six discussions divided between testing and control groups is too low to produce any academically definitive results, though from a business standpoint they might be enough to affect judgement.
- Group/sample size for participants – as we are fairly confident that we will have at least 40 participants altogether, our sample size on this front is more

than enough to create credibility in our results. The event in which the amount dropped vastly is extremely unlikely.

- Participant pool – the participants for the discussions are not all random and instead they have a certain profile. For the technical discussions there is expected to be only one female participant. Also, in the parent company's discussion, the participants are relatively young, most of them being under 25 years of age. However, the tests are arranged in such a way that as there is no notable difference between two discussions in a test case discussion pair. Thus, the participant pools should not have a notable impact on our results.
- Different levels of guidance – as brought up in Table 2, the non-LOGE discussion of the parent company has a more detailed level of guidance which more closely resembles its corresponding LOGE discussion. This might have an effect on the flow of the discussion in comparison to the other two non-LOGE discussions and also on its comparison to its LOGE counterpart. This will be taken into account during the evaluation of the results.
- Overlap in discussion participants – in the NordicEdu discussions, the participation pool of the all-encompassing discussion contains the participation pool of the technical discussion. However, as the two discussions do not share any common subject matter, this should not be an issue. In addition, the effect of overlap is further minimized by having most of the participants in the technical discussion be on the "inverted" group in the all-encompassing discussion (i.e. those previously in LOGE will do the non-LOGE and vice versa).

The recognized elements, as with the customer project, do not have major meaningful effect on the credibility of the test results. Since we are almost certain to have a sufficient amount of participants, our most major worry is that the participant overlap has an effect on our results. As precautions have been taken to circumvent this, our hypothesis is that the test results will be credible.

With LOGE the one most interested in our test results is NordicEdu, who is also the main developer of the product. They are interested to see the effects their product had in direct comparison to an non-gamified discussion about exactly the same subject. If the results were that LOGE improves on discussions, this can provide valuable marketing material. Even if the results were to be the opposite they may be used to

improve on LOGE's current design. In both cases the tests (as they are planned) should produce clear indications to how well and to what extent LOGE has implemented successful gamification mechanics. Our hypothesis is that the results will be relevant to NordicEdu. As for our secondary hypothesis about the academical relevance of our test results, we believe these tests to be relevant but, as with the customer project, less impactful.

3.4.3. Hypotheses concerning the evaluation methods themselves

Arguably our most interesting hypothesis concerns the evaluation methods themselves, namely their practicality and usability for business purposes. We do recognize that the case-control study requires for us to receive double the amount of participants for it to be effective, which poses a major problem when trying to implement it in a commercial setting. The study, although effective, requires for either a high participation percent (in practice that participation is mandatory for a number of people) or a huge number of potential participants. In the project of the unnamed customer we would not have considered it in any way a viable evaluation method, were it not for the large size of the customer company. This is also alleviated by the already in-place LMS of the customer that enables participation practically anywhere. In the case of LOGE, we are counting heavily on the internal drive among both NordicEdu and its parent company to produce a readied pool of participants for the test cases of LOGE. The chosen key factors for the quantitative test and the questions for the qualitative test are relatively common in such cases and were easy to implement. This can be generalized to similar situations, or to situations where the incentive for participation is high enough to allow for the required sample size (in other words: participation).

Our main hypotheses for this thesis are as follows:

H1. The evaluation methods are effective and produce valuable results from a business standpoint.
H2. The evaluation methods are practical and easy to implement in a business setting.

3.5. Conducting the research

After we had begun our research on the customer project we soon discovered a bug in our test package that forced us to reject three samples. This was extremely unfortunate and discouraging, as participations for the research had proven scarce. The bug was

fixed in a timely manner before it could affect any more samples and the test was continued according to plan. After the research for the customer project had been going on for a few weeks it soon became clear that one month was not going to be long enough for us to reach even the minimum of 20 participants. This led us to prolong our research. After a total of three months we were still unable to attain the absolute minimum of 10 valid participants per test group. We even collecting a separate supplementary sample pool from non-retail workers of the customer with the idea that the usability of these separate samples and their validity for our research would be closely examined before taking them as a part of our results. This too was however to no avail as altogether, counting in the supplementary participations, after three months of collecting data, we had following sample sizes:

CONTROL GROUP: 8
TEST GROUP: 4
REJECTED: 3

After failing to reach the minimum participation level in this time, we terminated the test.

It should be noted, that during our research, we frequently fell out of contact with our contact person from the customer company. Despite our numerous attempts, we were unable to receive any response from them, sometimes for weeks at a time. Though the actual effect that this one-sided communication had on the research can only be speculated on, it seemed to play a part on the failure of the test.

In the case of LOGE our research went almost completely as planned. Not all participants were able to attend every discussion fully, but we managed to gather large enough sample sizes for us to present credible results. With the parent company's LOGE discussion we also had few minor technical mishaps, which affected the length of the discussion. This will be taken into account, when the results of the tests are presented.

4. Results

4.1. Project results

- Present the results of gamification evaluation

4.2. Relevance of results

- Evaluate the credibility of the results
- Evaluate the relevance of the results
    - x …for the project
    - x …for the user
    - x …for the customer
    - x …for the gamifier

4.3. Observations about used methods and applied theories

- Were there any noticeable problems or special conditions for applying the evaluation methods

5. Analysis

5.1. Effectiveness of used methods and applied theories

5.2. How to improve gamification evaluation methods

6. Summary

1. What current gamification theory and metrics are most suitable to evaluate commercial gamification?

2. How can the evaluation of gamification benefits best be implemented in a commercial environment?

References

[1] Jouni Smed, Mika Luimula & Werner Ravyse. 2017. Course material for course TKO_3111: Gamification.

[2] Dirk Basten. 2017. *Gamification*. IEEE Software Magazine.

[3] Juho Hamari, Jonna Koivisto & Harri Sarsa. 2014. *Does Gamification Work? — A Literature Review of Empirical Studies on Gamification*. Hawaii International Conference on System Science.

[4] Google Trends. 2019. Trend for "gamification" search term in the world between January 1st 2009-March 3rd 2019. <https://trends.google.com/trends/explore?date=2009-01-01%202019-03-03&q=gamification> Visited: 3.3.2019

[5] Ana Carolina Tomé Klock, Aline Nunes Ogawa, Isabela Gasparini & Marcelo Soares Pimenta. 2018. *Does gamification matter? A systematic mapping about the evaluation of gamification in educational environments*. Annual ACM Symposium on Applied Computing.

[6] Gameffective. 2019. Sales slogan from the front page of the company's web page. <https://www.gameffective.com/> Visited: 7.3.2019

[7] Michelle da Silva. 2017. *Why Retailers Should Consider Gamification to Boost Engagement*. Shopify blogs <https://www.shopify.com/retail/why-retailers-should-consider-gamification-to-boost-engagement> Visited: 7.3.2019

[8] NordicEdu Oy. Company description (in Finnish). <https://nordicedu.com/yhteys> Visited: 11.3.2019

[9] Lassi Puolakka, marketer/game and gamification designer at NordicEdu. 11.3.2019. Personal interview.

[10] Katie Seaborn & Deborah I. Fels. 2015. *Gamification in theory and action: A survey*. International Journal of Human-Computer Studies.

[11] Sebastian Deterding, Rilla Khaled, Lennart E. Nacke & Dan Dixon. 2011. *Gamification: Toward a Definition*. CHI 2011. ACM, Vancouver

[12] Juho Hamari, Lauri Keronen & Kati Ahla. 2015. *Why Do People Play Games? A Review on Adoption and Use.* Hawaii International Conference on System Sciences.

[13] Darina Dicheva, Keith Irwin & Christo Dichev. 2018. *Motivational Factors in Educational Gamification*. IEEE International Conference on Advanced Learning Technologies.

[14] Amir Matallaoui. 2018. *Towards More Effective Gamification: Does Deploying Semiotics Help Design Better Perceivable Badges?* International Conference on Computer and Technology Applications.

[15] Michael Sailer, Jan Ulrich Hence, Sarah Katharina Mayr & Heinz Mandl. 2017. *How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction*. Computers in Human Behavior, no. 69.

[16] LinkedIn. <https://www.linkedin.com/> Visited: 18.3.2019.

[17] Bhupesh Raj & Deepak Gupta. 2018. *Factors Influencing Consumer Responses to Marketing Gamification*. International Conference on Advances in Computing, Communications and Informatics (ICACCI).

[18] E. Kaila, M.-J. Laakso, T. Rajala & E. Kurvinen. 2018. *A Model for Gamifying Programming Education: University-Level Programming Course Quantified*. International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)

[19] Ming-Hui Wen. 2017. *Applying Gamification and Social Network Techniques to Promote Health Activities*. IEEE International Conference on Applied System Innovation IEEE-ICASI.

[20] Mihaly Csikszentmihalyi.1990. *Flow: The Psychology of Optimal Experience*. HarperCollins Publishers.

[21] Richard Ryan & Edward Deci. 2000. *Intrinsic and Extrinsic Motivations: Classic Definitions and new directions*. Contemporary Educational Psychology 25.

[22] Richard Ryan & Edward Deci. 2000. *Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being*. American Psychologist.

[23] Martin E. P. Seligman & Mihaly Csikszentmihalyi. 2000. *Positive Psychology: An Introduction*. American Psychologist.

[24] Mihaly Csikszentmihalyi. 2014. *Flow and the Foundations of Positive Psychology: The Collected Works of Mihaly Csikszentmihalyi*. Springer Dordrecht Heidelberg New York London.

[25] Scott Nicholson. 2012. *A User-Centered Theoretical Framework for Meaningful Gamification*. Games+Learning+Society 8.0, Madison, WI.

[26] Umar Ruhi. 2015. *Towards a Descriptive Framework for Meaningful Enterprise Gamification*. Technology Innovation Management Review, Volume 5, Issue 8.

[27] Abdul Aziz, Ana Mushtag & Misbah Anwar. 2017. *Usage of Gamification in Enterprise: A Review*. International Conference on Communication, Computing and Digital Systems (C-CODE).

[28] Paul Harrigan, Uwana Evers, Morgan Miles & Timothy Daly. 2017. *Customer engagement with tourism social media brands*. Tourism Management 59.

[29] James Harter, Frank Schmidt & Theodore Hayes. 2002. *Business-Unit-Level Relationship Between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis*. Journal of Applied Psychology.

[30] Yong Jin, Xin Yang, Raula Kula, Eunjong Choi, Katsuro Inoue & Hajimu Iida. 2015. *Quick Trigger on Stack Overflow: A Study of Gamification-influenced Member Tendencies*. Working Conference on Mining Software Repositories.

[31] Benedikt Morschheuser, Karl Werder, Juho Hamari & Julian Abe. 2017. *How to gamify? A method for designing gamification*. Hawaii International Conference on System Sciences.

[32] Leah Taylor & Jim Parsons. 2011. *Improving Student Engagement*. Current Issues in Education, vol 14.

[33] Description of SCORM. <https://www.scorm.com>. Visited: 1.4.2019.

[34] Topi Salonen, software and game developer, and lead developer of LOGE at NordicEdu. 2.9.2019. Personal interview.

[35] NordicEdu Oy. LOGE description. <https://nordicedu.com/portfolio/loge?lang=en> Visited: 18.8.2019

Appendix A: The qualitative questions in the customer test

Below are the questions from the qualitative part of the testing form. The ones marked with * are mandatory questions. In the end there are questions that are specific to the test group and only visible to them.

- How would you evaluate your starting knowledge before you took the training?*
    - I knew the subject of the training well.
    - I knew some things of the subject matter and the practices concerning it.
    - I knew practically nothing of the subject matter and the practices concerning it.
- Did you think that you could benefit from this training before taking it?*
    - Yes
    - No.
- What was your opinion on the training?*
    - Good and educational
    - It was OK
    - Bad
    - Other: _____
- Did you find the training useful?*
    - Yes
    - Only parts of it
    - No
- How well do you think that you learned the contents of the training?*
    - I fully learned and internalized the content and the operating models from the training.
    - I learned and internalized the content and the operating models from the training really well.
    - I learned and internalized a lot of the content and the operating models from the training, but a lot of it I also did not learn.
    - I did not learn much from the training.
    - I learned nothing from the training.

- How well did you understand the reasoning behind the subject matter of the training?*

    - I understood all or almost all of the reasoning behind the subjects of the training.
    - I understood most of the reasoning behind the subjects of the training.
    - I understood a lot of the reasoning, but often I also did not understand it.
    - I understood very little of the reasons behind the subjects.
    - I understood none of the reasoning behind the subjects of the training.

- How well do you think you will remember the subjects of the training in the future?*

    - I think I will remember all about the subjects of the training in the future.
    - I think I will remember most about the subjects of the training in the future.
    - I think I will remember a lot about the subjects in the future, but also forget a lot about them.
    - I think I will not remember much about the subjects of the training in the future.
    - I think I will not remember anything about the subjects of the training in the future.

- Which adjectives would you use to best describe the training? (choose all that you would use)

    - Clear
    - Educational
    - Motivational
    - Challenging
    - Fun/nice
    - Compact/tight
    - Pleasant to look at
    - Other: _____

- What did not work in the training in your opinion? (freeform text answer)

- Did you face any bugs or errors during the training?*

    - Yes
    - No

- What bugs/errors did you face? (Answer only, if you answered yes to the previous question) (freeform text answer)

The training material

- How easy did you find the internalization of knowledge from the training?*
    - The internalization was very easy.
    - The internalization was relatively easy.
    - The internalization was not easy, but not hard either.
    - The internalization was relatively hard or required some effort.
    - The internalization was very hard or required a lot of effort.

- How well was the training material divided into categories?*
    - The division was good, clear and logical.
    - The division was somewhat clear and logical.
    - The division was not clear, but neither was it unclear.
    - The division was somewhat unclear and illogical.
    - The division was bad, unclear and illogical.
    - I did not pay any attention to the categories.

- How hard were the subjects of the training material in your opinion?*
    - Too hard or they focused on irrelevant details.
    - The focus was in the right things, but some of the subjects were too hard or focused on the wrong details.
    - The subjects of the training were well balanced and not too hard or easy.
    - The focus was in the right things, but some of the subjects were too easy or self-explanatory/obvious.
    - Too easy or self-explanatory/obvious.

- How did you feel about the length of the training material?*
    - The training material was too long.
    - The training material was just the right length.
    - The training material was too short.

- How much time did you use with the training material?*
    - Too long, the content could have been shorter or some of it could have been left out.
    - Too long, I could have been quicker.
    - Just the right amount, considering the length of the training material.
    - Too little, I rushed through it.
    - Too little, the material was insufficient or unclear.

- Were the instructions for the training clear?*
    - Yes
    - No

- How did going through the training material feel? (choose all that you would use)

    - Interesting
    - Fun
    - Boring
    - Appropriate
    - Inappropriate
    - Quick
    - Slow
    - Rewarding
    - Frustrating
    - I did not feel any noticeable emotions

- Did you use the customer's website as a tool during your training?*

    - Yes
    - No

- Feedback for the training material. What was good? What was bad? What would you want more of? What would you want less of? (freeform text answer)

The ending test

- How challenging did you find the ending test?*

    - Too hard
    - Appropriately challenging
    - Too easy

- Did the ending test test relevant subjects in the training material?*

    - The ending test focused into irrelevant details.
    - The ending test focused on relevant subjects in the training.
    - The ending test focused on too self-explanatory/obvious subjects.

- Where the questions in the ending test specific enough?*

    - The questions focused on too minute details in the material.
    - The questions focused on the material with appropriate specificity.
    - The questions focused on too vague notions in the material.

- How did you find the length of the ending test?*

    - Too long
    - Appropriate
    - Too short

- How did you find the ending test as a part of the training? (choose all that you would use)

    - Rewarding
    - Educational
    - Revising
    - Anxious
    - Frustrating
    - Repetitive
    - Discouraging
    - Useless

- Did you use the customer's website as during the ending test?*

    - Yes
    - No

- Feedback for the ending test. What was good? What was bad? What would you want more of? What would you want less of? (freeform text answer)

Questions about the gamified training (only for the test group)

- What did the training most resemble?*

    - A quiz
    - A training material
    - An official form
    - An exam
    - Other: _____

- What did you like about the training?* (choose all that you would use)

    - Clarity
    - Ease-of-use
    - Presentation
    - Flow
    - The humorous images
    - The celebratory titles
    - The questions
    - The repetition basket
    - UI
    - None of the above

- What did you not like about the gamification? (freeform text answer)

- What did you think of the repetition basket?*
    - Using the repetition basket was difficult
    - It was useful
    - It made it too easy to complete the training
    - I did not use the repetition basket
- What did you think about the humorous pictures?*
    - They were good and funny
    - They were OK
    - They were boring and bad
    - They were inappropriate or offensive
- What did you think of the celebratory titles?*
    - They are a good and fun addition
    - They were OK
    - They were boring or unnecessary
- What did you like about the UI of the training? (choose all that you would use)
    - Clarity
    - Colour choices
    - Background images
    - The customer's brand
    - The humorous pictures
    - The icons [of the UI]
    - The fonts
    - The responsiveness
    - Other: _____
- What did you not like about the UI of the training? (freeform text answer)
- How were the font sizes in your opinion?*
    - They were good
    - They were too small
    - They were too big
    - They were at times too small and at times too big
- What did you think of the clarity of the questions?*
    - The questions were clear and the answers could be found sensibly
    - The questions were clear, but the answers could not be found in a logical way
    - The questions were unclear, but the answers could be found in a logical way
    - The questions were unclear and the answers could not be found in a logical way

- What did you think of the hints?*

    - They were useful and clear
    - They were too easy
    - They did not help enough
    - I did not notice that the training had hints

- Do you think that gamification brought added value to the training?*

    - Yes, a lot
    - Yes, some
    - Yes, but only a little
    - No
    - I cannot say

- Did you enjoy more the training because of gamification?*

    - Yes
    - No
    - I enjoyed the training less because of the gamification
    - I cannot say

- Did the gamification help you internalize the subjects of the training?*

    - Yes, a lot
    - Yes, to some degree
    - Yes, but only a little
    - No
    - I cannot say

- Feedback for the gamification. What was good? What was bad? What would you want more of? What would you want less of? (freeform text answer)

Appendix B: The qualitative questions in the LOGE test

Below are the evaluation questions for LOGE test cases. The ones marked with * are only posed to the participants that use LOGE.

- What was the discussion about?
- How would you say the discussion went? (1= very poorly, 5= extremely well)
- How useful would you describe the discussion? (1= completely useless, 5= very useful)
- Would you describe the discussion as having value to the company? (1= the discussion was worthless to the company, 5= the discussion was very valuable to the company)
- How well would you say the objectives of the conversation were achieved? (1= not at all or just scratched the surface, 5= they were completely achieved)
- Which of the following adjectives best describes the discussion (feel free to use multiple): fluent, participating, clear, clever, enlightening, dull, inappropriate, slow, exciting. You may also use your own adjectives.
- What did not work well or was bad about the discussion?
- Which of the following did the discussion most remind you of: an exam, a panel, small talk, a meeting, a jury, none of these. Feel free to provide an explanation to your answer, if it feels appropriate.
- How well would you say the questions guided the discussion? (1= extremely poorly, 5= really well)*
- How well was the discussion divided into different categories? (1= the division was illogical and had a negative effect on the discussion, 5= the division was logical and made the discussion more fluent)
- Which of the following words would you use to describe the difficulty of the subjects that were under discussion: plain/obvious, straightforward, not easy or challenging, hard, insurmountable, of varying difficulty.
- How long was the conversation in terms of its content? (1= much too short, 3= of an appropriate amount, 5= much too long)
- How long was the conversation in terms of the time it was given? (1= much too short, 3= of an appropriate amount, 5= much too long)
- As which of the following would you describe the time restrictions of the discussion (feel free to use multiple): oppressive, hurrying, slow, exciting, challenging, none of these.*
- Were the instructions clear for the discussion? (Yes, No)

- Wich of the following roles best describes the part you had in the discussion: leader/instigator, participant, ideator, commentator, moderator, listener, bystander.

- Freeform feedback for the discussion.

Appendix C: The questions for subjective gamification evaluation

Below are the questions from the qualitative part of the testing form.