



Gamification in Educational Software Development

Master Thesis Information Science

Author:

Achilleas Buisman

s0514020

mail@abuisman.nl

Supervisor:

Prof. dr. M.C.J.D. van Eekelen

marko@cs.ru.nl

Second reader:

Dr. E.M.G.M. Hubbers

e.hubbers@cs.ru.nl

September, 2014

Contents

1	Introduction	1
1.1	Problem statement	1
1.2	Research goal	4
1.3	Case study	5
2	Theoretical Background	6
2.1	Gamification	6
2.2	Serious-gaming and Game-based learning	6
2.3	Engagement	11
2.4	Motivation	12
2.5	Summary	15
2.6	Conceptual model	16
3	Methodology	17
3.1	Method of data collection	17
3.2	Data analysis	21
4	Analysis and results	22
4.1	Missing value analysis	22
4.2	Outlier analysis	22
4.3	Sample	23
4.4	Statistical assumptions	24
4.5	Reliability analysis	25
4.6	Descriptive statistics	25
4.7	MANOVA	26
4.8	MANCOVA	27
4.9	Post hoc comparisons between individual group differences	28
5	Discussion and conclusion	30
5.1	Theoretical implications	31
5.2	Practical implications	32
5.3	Limitations and future research	33
5.4	Conclusion	35
	Bibliography	36
6	Appendix	38
6.1	Survey	38

6.2	Points given for actions	40
-----	------------------------------------	----

List of Figures

1.1	Profile completeness on LinkedIn	3
1.2	HabitRPG game-like habit dashboard	3
1.3	Factory gamification	4
2.1	Levels of game design elements	10
2.2	Game mechanics and game dynamics	11
2.3	Gamification between game and play, whole and parts	12
2.4	Drives and organisational levers	14
2.5	Conceptual model	16
3.1	Gamified Redmine dashboard with comparison	19
3.2	Redmine leaderboard	20

List of Tables

4.1	'First course' at GiPHouse	23
4.2	Student study phases	23
4.3	Frequencies of 'Prior experience with Redmine'	24
4.4	Descriptive statistics on our dependent variables	25
4.5	Cross table of experimental groups on dependent variables	26
4.6	Statistics on our covariates	26
4.7	Descriptive statistics for manipulation checks	26
4.8	Experimental group size	27
4.9	ANCOVA results for 'Enjoyment in project'	28
4.10	ANCOVA results for 'Experience with redmine'	29
4.11	ANCOVA results for 'Group'	29

Introduction

IT studies are one of the few studies that enable students to work professionally in their field even before they graduate. The reason for this is that most IT studies involve learning how to program or design. These skills are sought after at many different levels of experience by businesses. Teaching professional coding is not just about teaching students how to program; students have to learn how to translate requirements into technical designs, they must communicate with clients and team members, they have to adjust requirements in order to meet deadlines, etc.

Most computer science educations teach students this by creating courses in which students have to practice all these things, often with real clients and projects. Students will work in teams to complete real products for their clients. We call this ‘Educational Software Development’ (ESD). The goal of ESD is to develop student’s necessary non-programming skills as to get them ready for work in professional teams. ESD courses are therefore often set-up as companies; there are clients, teams, management, in-house tools and methodologies, etc. There are professional responsibilities as well: logging hours, documenting code, progress reporting, etc. We can therefore say that ESD courses are companies and students are their employees.

1.1 Problem statement

ESD projects can often be challenging for students. They will have had some experience with programming, be it through personal projects or assignments in programming courses, but working for a client and, as part of a team is very different. There will be differences of opinions, conflicts, clients who change their minds, issues with third parties, non-disclosure agreements and many other unforeseen obstacles on the road to a finished product. At the same time, their employer, the ESD company, will need students to perform many, to them foreign, activities, such as documenting used systems for future teams and performing peer reviews. Not only are many of these things new to students, but they will also have to learn and apply them in a very short amount of time.

Many of the tools and methodologies that students will have to start using are, as is the case with companies, predefined by the ESD course’s creators: version control systems, project management tools, project methodologies, etc. This can be an advantage, as students will not have to make decisions about things they have little experience in. On the other hand, this can be a disadvantage, as students who do have

some experience, might feel reluctant to use the ESD's tools in favour of what they are familiar with themselves. This might happen very quickly when it does not concern something as specialised as version control, but something as common as a communication tool. Students will most likely already use something to communicate with each other. Now, the ESD course's tools and methodologies will have to compete with what students already know and use.

Besides motivating students, another challenge for ESD is individual supervision. Because all or most of the work is done in groups, it becomes very hard to spot when individuals are having problems with learning and applying their new skills. Not to mention grading when each deliverable has been worked on by several students. Individual problems going unnoticed makes it harder to help students develop the required skills, or making sure that they follow along as intended. Unnoticed, a student, following the path of least resistance, might, for example, not perform atomic commits with short clear commit messages, which will degrade both the quality of the work as well as the educational value of a project.

Students can be involved in the evaluation of group-colleagues, but students are often less inclined to judge each other openly and passing on critique to supervisors is often felt as betrayal. Efforts in ESD are made to counter this however; often peer-reviewing sessions are held and students are encouraged to solve issues amongst themselves because everybody benefits from receiving good feedback. The areas that are covered in these peer-reviewing sessions are subject to change and different questions could yield different results. Interpreting these results, however, requires time and the results' usefulness wavers. Peer reviews cannot be done very often due to the time they take to create and fill in. The results can usually only be used fairly late in the course by which time students might have formed less desirable habits already.

Increasing numbers of students involved in courses make individual supervision a growing problem. Face-to-face sessions with *every* student are made impossible due to the resources that this would require in terms of manpower. It is therefore essential to be able to use the resources that *are* available as efficiently as possible.

We should be able to leverage tools, methodologies and new developments in programming for bettering the ESD process. Automated systems might be able to take away some of the practical challenges that ESD faces. We should be able to apply existing methods of motivating students to learn new skills that they will need in their professional careers.

1.1.1 Gamification

In recent years, gamification has emerged as a means to engage and motivate users of (mostly) digital systems. Gamification is defined as the use of game design elements in non-game contexts (Deterding et al., 2011, p. 9).

We are probably all familiar with varying forms of gamification. Websites that feature profiles and show a progress bar to indicate how 'complete' your profile is, are an example (see Figure 1.1).

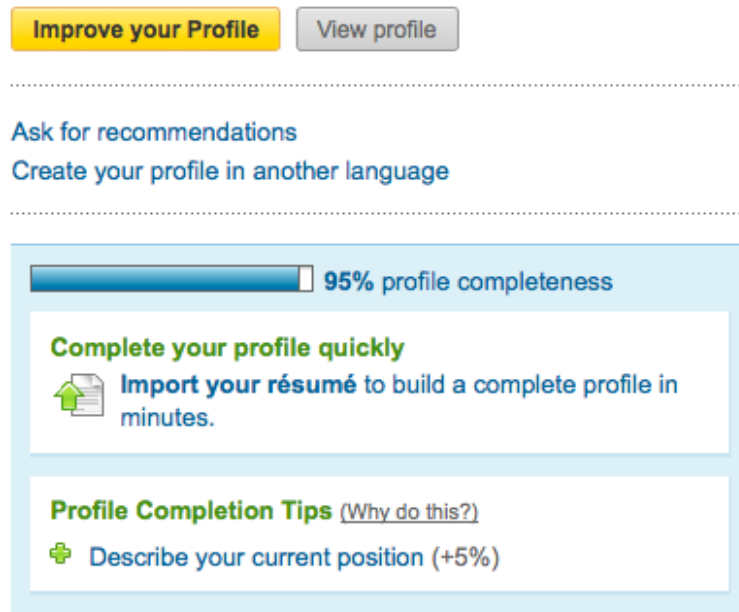


Figure 1.1: Profile completeness on LinkedIn (<http://linkedin.com>)

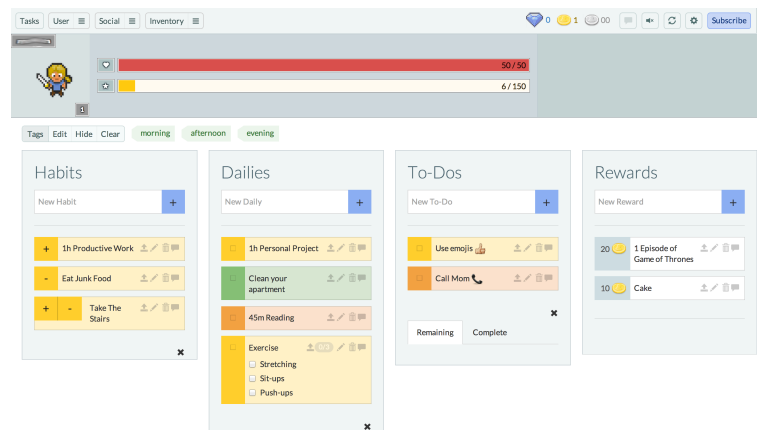


Figure 1.2: HabitRPG offers a game-like dashboard for forming good habits.

Applying gamification in a very immersive way will require additional facilities to be put in place. An example of this is HabitRPG ¹. HabitRPG is a very strongly gamified to-do list application where users have avatars, badges, power-ups, etc. that they can earn or change after completing certain tasks. Not performing tasks will cause them to lose ‘health’ and completing tasks will earn them experience points and coins to use on developing their character. Earning coins will only be relevant if users can then buy things with these coins (see Figure 1.2).

In her article, Van der Waals (2014) describes a gamified factory floor that, amongst

¹HabitRPG, a gamified to-do list <https://habitrpg.com>



Figure 1.3: Factory gamification (Van der Waals, 2014)

other things, presents workers with graphical representations of how much steel they have produced this day, week, month and year (see Figure 1.3). The effect has been that fewer errors are made in the production process, which means less money lost and higher throughput for the factory.

Examples like these and the research that has been done so far into applying gamification has shown that there is space for gamification in personal, educational and professional contexts. We wanted to investigate whether we could use it to improve the ESD process, in which education and professionalism meet.

1.2 Research goal

So far, gamification has been researched as a means to increase motivation amongst students in education, but little research has been done with applying it to more practical contexts such as ESD. Nevertheless, this context is very interesting and exploring this avenue might reveal new ways to motivate students.

We therefore set out to research: To what extent does the gamification of tools used in Educational Software Development lead to students using these tools more?

Understanding how to stimulate students to use the tools more is desirable as an increased use of the tools will lead to students understanding them quicker and through continuous feedback students could be guided towards the actions of importance to the course's designers.

Since gamification is often used to increase engagement and involvement and to motivate users in performing certain behaviour, we also wanted to examine: To what extent does the gamification of tools used in Educational Software Development affect the engagement, involvement and motivation of students?

There is no magic wand of gamification one can wave at something to have it effectively 'gamified'. Gamification can be applied in many different ways at varying levels of immersion as we will describe further in the theoretical background (chapter 2).

The goal of this research was to find a set of game elements that we could introduce to a project management tool in order to increase its use and to stimulate student's attitudes.

1.3 Case study

In order to research the questions stated above, we have made use of 'GiPHouse' as an illustrative case to present and analyse gamification as a means to improve ESD.

GiPHouse is a software development company run by students of the Radboud University in Nijmegen (NL), as part of different courses in Information Science and Computer Science studies. GiPHouse builds software for a variety of clients in different programming languages.

All of GiPHouse's projects are entered into Redmine, an open source project management tool. Redmine allows students to access issues, time tracking, documentation of projects and GiPHouse in general, etc. Redmine forms an important part of GiPHouse's continuity system from one semester to another. Since all employees leave at the end of a semester and are replaced at the beginning of the next, continuity is essential. Redmine's wiki and uploaded document features have served as a practical means to make this possible. Since Redmine is such an important part of the work at GiPHouse, the wish was to motivate students to use it as much as possible and to have them contribute in improving the information that is found within. We therefore applied gamification to Redmine in order to measure the effects on student activities, engagement, involvement and motivation. We will elaborate on this under 'Methodology' (chapter 3).

Theoretical Background

2.1 Gamification

Gamification is defined as “the use of game design elements in non-game contexts.” in order to increase engagement and to motivate certain behaviour (Deterding et al., 2011, p. 9). While the term is fairly young, mentioned first around 2008 and adopted as a widespread term in 2010, there has been a lot of interest for gamification in very recent years in both academia as well as commercial applications such as FourSquare¹, Gamification.co², Rick Gibson³ and Jane McGonigal⁴ (Cohen, 2011, Danforth, 2011, Deterding et al., 2011, Domínguez et al., 2013, Huotari & Hamari, 2012, Perryer et al., 2012, Simões et al., 2013).

Deterding et al. (2011) investigated common concepts in human-computer interaction to find out where the term ‘gamification’ comes from and how its origins relate to these similar or parallel concepts. They found that gamification has been used more and more to increase both user activity and retention in the fields of interaction design and digital marketing. Gamification has applications in many different fields from productivity, finance, health, education, to news and entertainment media sectors (Deterding et al., 2011, p.9).

Education especially has received much attention from both researchers and companies as a field to apply gamification in (q2l.org, schoooooools.com). Its motivational nature and its potential for increasing engagement are thought to have positive effects on students. Video games offer advantages that could be useful in a learning context: immediate feedback and information, collaboration or self-study, but not without risk in terms of content and assessments and focussing too much on measurable results (Domínguez et al., 2013).

2.2 Serious-gaming and Game-based learning

Despite the young age of gamification, many parallel terms have been used to describe the phenomenon, such as “‘productivity games’, ‘surveillance entertainment’, ‘funware’, ‘playful design’, ‘behavioral games’, ‘game layer’ or ‘applied gaming’

¹<https://foursquare.com/>

²<http://www.gamification.co>

³<http://www.develop-online.net/opinions/show-gamification-some-love/0117872>

⁴http://www.ted.com/talks/jane_mcgonigal_the_game_that_can_give_you_10_extra_years_of_life

(natronbaxter.com)” (Deterding et al., 2011, p. 9). These terms mostly revolve around the concept of ‘play’.

Two concepts that are often applied in education are ‘game-based learning’ and ‘serious-gaming’. Game-based learning refers to the use of games to support learning, whereas serious-games are specific games that have been developed with the intention of teaching (Deterding et al., 2011, Simões et al., 2013). An example of the use of game-based learning through serious-games in education is Quest to Learn (Q2L), a well-funded public charter school in New York City where textbooks have been practically replaced with game-based learning. Children defeat bad guys while learning maths, science, languages and other subjects, in virtual worlds built by researchers and educators based on what they know about how children learn (Cohen, 2011). Other examples can be found in the teaching of programming and IT concepts through games. There is a trend in instructional materials being created in the form of small serious-games (Perryer et al., 2012, Citing Winn); Rails for Zombies is a project by Code School, an organisation that creates interactive game-based courses for programmers, where players can earn badges and progress through stages by first learning and then solving programmatic exercises Rails for Zombies⁵. A more immersive example is the open source game CodeCombat where players control their characters in a full-fledged Role Playing Game (RPG) by writing code for every action their character performs⁶.

Neither game-based learning nor serious gaming have been used in our study. Instead we have focused on applying game elements to existing systems, as described further on.

2.2.1 Paidia vs. ludus

In his book ‘Man, Play, and Games’, Roger Caillois (2001) describes two opposite poles between which different games can be placed: the poles ‘paidia’ and ‘ludus’. “At one extreme an almost indivisible principle, common to diversion, turbulence, free improvisation, and carefree gaiety is dominant. It manifests a kind of uncontrolled fantasy that can be designated by the term *paidia*. At the opposite extreme, this frolicsome and impulsive exuberance is almost entirely absorbed or disciplined by a complementary, and in some respects inverse, tendency to its anarchic and capricious nature [...]. This latter principle is completely impractical, even though it requires an ever greater amount of effort, patience, skill, or ingenuity. I call this second component *ludus*” (Caillois, 2001, p.13). Caillois describes herein the difference between ‘play’ (*paidia*), a free form of enjoyment ungoverned by rules or systems, versus ‘games’ (*ludus*), which involve rules and the requirement to learn and apply these rules, often with competitive motives. This distinction between “playing” and “gaming” is supported in recent theoretical and empirical studies, where they are found to be

⁵<http://railsforzombies.org> by Code School - an interactive and gamified course on learning Ruby on Rails

⁶<http://codecombat.com> an interactive game that teaches programming

distinct modes or foci of behaviour and mind-set during the playing of video games (Caillois, 2001, Deterding et al., 2011, Domínguez et al., 2013, Perryer et al., 2012).

Applications of gamification have been criticised however to focus primarily on the rule based or quantifiable aspects of games (*ludus*) rather than taking into consideration the free-form exploratory nature of play (*paidia*) (Deterding et al., 2011, Perryer et al., 2012). However, Deterding et al. (2011) point out that, from a research perspective, this critique can be seen as a valuable observation of the fact that designing inspired by games allows for leaning more to one pole (*ludus*). In doing so we can gather empirical data, which has been given less emphasis in traditional human-computer interaction research (Deterding et al., 2011, p.11) (Perryer et al., 2012).

In light of this contrast, Deterding et al. (2011, p.11) suggest adopting the term ‘gamefulness’ (‘qualities of gaming’), attributing the term to Jane McGonigal, as a complement to ‘playfulness’ (‘behavioural qualities of playing’). Gamefulness is a coherent set of phenomena that is distinct from playfulness that has not received much focused attention so far (Deterding et al., 2011, p.11). They offer the following systematisation:

- Gamefulness (the experiential and behavioural quality),
- Gameful interaction (artefacts affording that quality), and
- Gameful design (designing for gamefulness, typically by using game design elements).

The distinction between gaming and playing within ‘gamified’ applications is not always as clear. Gamification may inspire playful behaviour and mind-sets just as video game players can switch between a focus of ‘gaming’ to ‘playing’ and *visa versa*. An example of this is the paper by Engqvist & Frommen (2008) called “The h-index and self-citations” in which they explain how scholars can, artificially, increase their h-index by citing their papers in other papers they write. To think about a scholar’s h-index in this way means that one displays the explorative and improvisational behaviour and mind-set of someone who is in a *modus* of playing, rather than gaming.

2.2.2 Game elements

Deterding et al. (2011, p.11) describe the use of game elements with the intention to create a gameful experience for a user as a requisite for being able to label something as ‘gamification’. They believe that this does not have to apply to digital environments only, though it often does (Deterding et al., 2011). A great example of gamification in the physical world is Volkswagen’s Fun Theory⁷ set of experiments; people were free to submit ideas for applying ‘fun’ to real world situations, in order to change people’s behaviour for the better. The winner was the ‘Speed Camera Lottery’ where all passing cars at an intersection were photographed by a speed camera. Those who obeyed the

⁷<http://www.thefuntheory.com/>

speed limit were automatically entered into a lottery to win the money earned by fining the people who drove over the limit. Over the course of three days, this resulted in a 22% drop in average speed, from 32 km/h to 25 km/h, between 24.857 cars.

What are game elements though? Deterding et al. (2011, p.11) define game elements as “elements that are characteristic to games – elements that are found in most (but not necessarily all) games, readily associated with games, and found to play a significant role in gameplay” (Deterding et al., 2011, p.12). Not all game elements will be found in all games however; some types of elements are more common in one game type than the other. For example, role-playing games often feature characters who receive experience points in certain skill areas whereas sports games often feature league leader boards and a count of wins and losses (Deterding et al., 2011, Simões et al., 2013). Both artefactual as well as social game elements need to be included. An example of an artefactual game element are badges one can earn, an example of social elements are leader boards where someone can compare their own performance to that of others (Deterding et al., 2011). Deterding et al. (2011, p.11) point out that these elements rather *afford* gameful interpretations, than that they *are* gameful. For example: the total distance driven by a truck driver is not gameful, but when it is compared to the distance driven by other drivers, in the form of a central leader board it becomes part of a gameful experience.

There are, of course, countless specific game elements, but they can usually be grouped in a few categories (Deterding et al., 2011, Perryer et al., 2012, Simões et al., 2013, Van der Waals, 2014):

- Badges
- Leader boards
- Immediate feedback
- 3D or highly graphic environments
- Avatars
- Narrative context
- Ranks and levels
- Marketplaces
- Competition
- Teams
- Communication systems (chat, voice, etc)
- Time pressure

The designer may apply these elements in different ways and with different sub-goals within gamification. Many authors have therefore made the distinction between game elements and game design (elements). Game design may also be referred to as game mechanics and is the application of game elements in order to provide a gameful experience. Applying game elements as they are, is not, per se, going to create a good gameful experience and might even have adverse effects. For example, as mentioned earlier, Van der Waals (2014), describes the gamification of a steel production factory

Level	Description	Example
<i>Game interface design patterns</i>	Common, successful interaction design components and design solutions for a known problem in a context, including prototypical implementations	Badge, leaderboard, level
<i>Game design patterns and mechanics</i>	Commonly reoccurring parts of the design of a game that concern gameplay	Time constraint, limited resources, turns
<i>Game design principles and heuristics</i>	Evaluative guidelines to approach a design problem or analyze a given design solution	Enduring play, clear goals, variety of game styles
<i>Game models</i>	Conceptual models of the components of games or game experience	MDA; challenge, fantasy, curiosity; game design atoms; CEGE
<i>Game design methods</i>	Game design-specific practices and processes	Playtesting, playcentric design, value conscious game design

Figure 2.1: Levels of game design elements (Deterding et al., 2011, p.12)

of the company Wuppermann Staal Nederland B.V. in Moerdijk, the Netherlands; each shift is run by a different team and each team sees their current progress, number of mistakes and their overall progress (week, month, year). The idea is to create an incentive to perform well, in terms of production quantity as well as quality. If the steel factory were to apply leader boards to their gamified production hall, teams could start to purposely hinder each other in order to ‘win’. Instead, the company has chosen to set team targets expressed in volumes in order to have the teams compete with the targets rather than each other (Van der Waals, 2014).

Designers must be mindful of what they want to achieve and which elements are applicable to their context. Several authors have defined some main areas of sub goals of gamification and what mechanisms support these. Lee & Hammer refer to the ‘cognitive area’, ‘emotional area’ and ‘social area’ (Lee et al., 2011). We will explain these in the following paragraph (section 2.4).

Deterding et al. (2011, p.11) have divided game design elements in levels ranging from the concrete to abstract as can be seen in Figure 2.1. This hierarchy can be used to select which game design elements to use when designing a gameful experience. We have chosen to work with the highest level ‘Game interface design patterns’ in this research, because this would not require changing the way in which the course is taught.

Simões et al. (2013, p.11) have adapted a list of game mechanics and dynamics from Bunchball⁸ that shows game mechanics (elements) with the dynamics that they influence or are part of as seen in Figure 2.2. Points, for example, belong to the game dynamic of rewarding players and leaderboards allow for a competitive dimension. We

⁸<http://www.bunchball.com/sites/default/files/downloads/gamification101.pdf>

Game elements	
Game mechanics	Game dynamics
Points	Reward
Levels	Status
Trophies, badges, achievements	Achievement
Virtual goods	Self expression
Leaderboards	Competition
Virtual gifts	Altruism

Figure 2.2: Game mechanics and game dynamics (adapted from Bunchball) (Simões et al., 2013, p.348)

used both of these in our gamification.

2.2.3 Summary

Deterding et al. (2011, p.13) summarize gamification as:

- the use (rather than the extension) of
- design (rather than game-based technology or other game- related practices)
- elements (rather than full-fledged games)
- characteristic for games (rather than play or playfulness)
- in non-game contexts (regardless of specific usage intentions, contexts, or media of implementation).

The distinction between play and gaming is important to make as, in a work or educational context, it will be the difference between consciously applying elements of games or create a playful ('fun') environment. Furthermore, an important distinction is the distinction between serious games, which are full-fledged games and gamification, which is the applying of game elements. To illustrate the position of gamification in terms of play versus gaming and partial and whole systems, Deterding et al. (2011, p.11) created the matrix shown in Figure 2.3. Using games or game technology is not necessarily gamification and in the end the intentional designing for a gameful experience is what gamification is.

2.3 Engagement

Work engagement is defined as a positive work-related state that is fulfilling and characterized by vigour, dedication and absorption and seen as the opposite of a burnout (Schaufeli et al., 2002, p.74).

In our study, the relevance of work engagement is obvious, as gamification has as a goal to increase engagement. The context of educational software development is an interesting one, as it is a mixture of a professional environment, a software development company, on one side and an educational environment, part of a course,

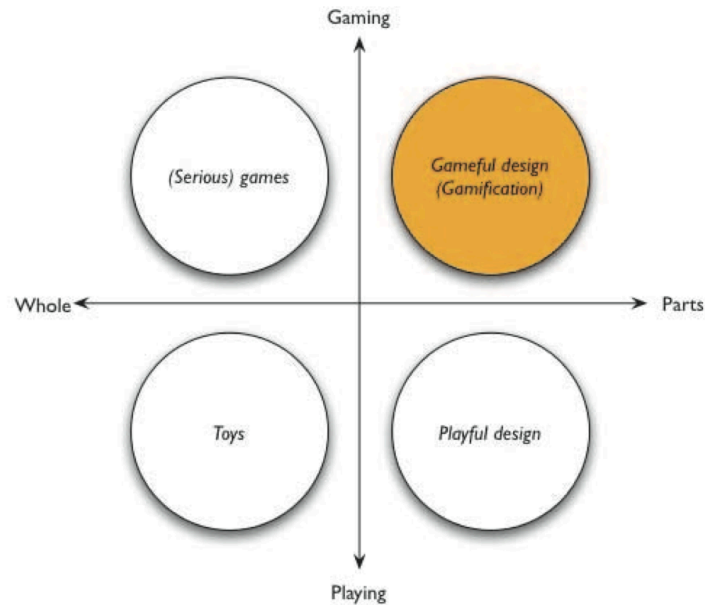


Figure 2.3: Gamification between game and play, whole and parts (Deterding et al., 2011, p.13)

on the other. Some parts of these types of courses are part of their education and other as part of their professional development.

So far, gamification has often been researched as a means to increase motivation amongst students in education, but little research has been done in applying motivation of employees in the workplace (Perryer et al., 2012).

Engagement and motivation are concepts that are closely related as they are both descriptive of a person's frame of mind and emotional state.

2.4 Motivation

Gamification is often used as a tool to increase student or employee motivation. When someone is motivated he is energized or activated to perform certain tasks or behave in particular ways (Ryan & Deci, 2000a). There are a few key concepts of motivation that are especially relevant when considering the use of gamification. Gamification evolves around using game design elements in non-game contexts to increase engagement and motivation in users. Which elements and dynamics should be chosen largely depends on the goals that we have in terms of what type of behaviour we want to motivate.

Lee et al. (2011) mention three areas of motivation that play part in games or gameful experiences, namely the 'cognitive area', 'emotional area' and 'social area'.

The cognitive area of motivation is the part of the player that wants to understand and learn how things work. There are cycles within games that teach a player its rules and make him behave a certain way. Mechanisms that can be used for this are visual

representations, story telling and structured or hierarchical tasks. In an educational context, it would mean not just telling facts, but showing students why something is relevant and what is next in the tree of knowledge. A subject such as maths lends itself very well for this, while it would be fairly hard to apply it to reading and understanding literature. In a broader context, it refers to accommodating to someone's need to understand and master concepts and tasks (Cohen, 2011, Lee et al., 2011).

The emotional area of motivation refers to the concept of rewarding and penalizing for wanted and unwanted behaviour. When designing for creating an emotional experience, the goal is to make users involved with the success or failure of either completing a task properly, or not being able to do so. Elements involved in this are reward systems, player penalties, levels, badges and trophies, to name a few. The challenge herein lies in balancing the difficulty. Since one needs to feel happy for being able to complete a challenge, it must not be too easy. In order to keep users motivated to try again it must not be too difficult. Ideally even, the difficulty of tasks or the amount of reward players receive should be adapted to their skill level (Deterding et al., 2011, Domínguez et al., 2013, Lee et al., 2011, Simões et al., 2013).

A concept that is characteristic for games, but not for education or work is the low-risk role of failure. Making mistakes is part of learning how a game works and learning how to become better at it. In educational contexts, we have examinations, and at work our performance is monitored. In both, failure is not often accepted as a necessary part of learning. Gamification can play a role in this (Lee et al., 2011). Examples of this are online practice exams prospective drivers can take in preparation of theory traffic exams⁹. Traditionally, one would learn the traffic rules and take the exam to prove their grasp of the rules. Interactive practice exams that one can take as often as he likes however, add the element of trial and error that is so characteristic to games. This often works better and is more engaging for students as Domínguez et al. (2013) Simões et al. (2013) have also shown. This same concept could be applied to other parts of education and work through gamification (Lee et al., 2011).

The social area of motivation revolves around identity and the ability to communicate with others or compare one's progress with that of others. When playing, players take on certain roles. These roles allow them to behave in a different way from their day-to-day identity (Lee et al., 2011). This is a very powerful element of video games. In non-game contexts it becomes interesting to have people take on and accept roles that are relevant to the situation: leader, academic, caretaker. When people not only perform tasks that are required because of external motivators, but because they accept them as parts of an identity, they are much more involved. Game elements that lend themselves to implementing this area are avatars or characters, interface customizations, virtual currencies and markets (exchangeable awards), leader boards, communication features (Deterding et al., 2011, Domínguez et al., 2013, Lee et al., 2011).

Lawrence's Four-Drive theory of motivation expands the field of motivation further

⁹<http://www.theorieles.nl>

Drive	Primary Lever	Actions
The need to acquire	Reward System	<ul style="list-style-type: none"> • Differentiate performance • Tie rewards to performance
The need to defend	Performance-Management and resource-allocation process	<ul style="list-style-type: none"> • Build trust • Increase transparency
The need to bond	Culture	<ul style="list-style-type: none"> • Foster mutual reliance • Value collaboration
The need to comprehend	Job Design	<ul style="list-style-type: none"> • Design meaningful jobs • Design challenging jobs

Figure 2.4: Drives and organisational levers (Perryer et al., 2012)

(Lawrence, 2011). It explains where our motivation can come from. This provides us with insights in how we can steer it and why we do things. He claims that humans have four basic drives (Lawrence, 2011):

- “The drive to acquire - to get what we need or value from food, shelter, and offspring to promotion, expertise, or excitement.”
- “The drive to defend - to protect what we need or value, including our company’s market share and reputation.”
- “The drive to bond - to form long-term, trusting, caring relationships that may provide but are not limited to mutual benefit. These include relationships with co-workers, customers, suppliers, and investors.”
- “The drive to comprehend - to make sense of the world and ourselves, which includes forecasting, inventing, and problem solving.”

Perryer et al. (2012) have created a table in which they have mapped these four drives to areas of organisational ‘levers’ and related actions as seen in Figure 2.4.

A very basic, but important distinction in the field of motivation, that is especially relevant to applying gamification, is the differentiation between intrinsic and extrinsic motivation (Perryer et al., 2012). Intrinsic motivation refers to doing something out of finding it interesting or enjoyable, while extrinsic motivation is doing something for which there is an external outcome that is unrelated to the task itself; promotion, points, rewards, etc. (Perryer et al., 2012, Ryan & Deci, 2000a). Because of their extraneous nature in relation to the tasks that we perform, extrinsic motivators are only affective for as long as they are present. Once someone receives their promotion, they will stop putting in the extra work; once a raise is earned, employee standards are adjusted to the new standard and motivation drops (Perryer et al., 2012). Intrinsic motivators, however, are more durable and therefore preferable in most scenarios. An example is job satisfaction rather than salary; someone will continue to be motivated at

their job when their reward for performing well is their own satisfaction in doing it (Perryer et al., 2012).

When designing for gameful experiences, it is therefore important to consider the type of motivation that needs to be applied and how it will affect users. Levels or badges are extrinsic motivators when applied on their own, but when they are added to a profile or character and can be seen by others they may become part of the identity of users of the gamified experience. Having badges and other tokens of performance displayed as part of a character's identity can give someone a more lasting feeling of achievement, and when social status is derived from these, it can even create the intrinsic motivation to keep hitting targets or getting rewards (Domínguez et al., 2013, Perryer et al., 2012, Simões et al., 2013, Van der Waals, 2014).

Adapting job designs to our basic drives is a way in which to ensure that we do our jobs better and are more motivated to perform. Gamification can be used complementary to this, but again is not without risk. A risk that is involved in applying gamification incorrectly is creating the wrong motivation (Lee et al., 2011, Van der Waals, 2014). This applies to any form of tool to influence motivation. Especially extrinsic motivators like pay rises, bonuses, etc. can become perverse incentives in any field, education or other wise, of which the Rabobank Libor Rate Fraud¹⁰ is a clear example. Care is therefore required when introducing motivational instruments to any context.

2.5 Summary

From the outset, gamification seems like a tool with potential to increase student engagement and apply stimuli for motivating students to perform desirable behaviour in digital systems. Which elements of gamification can we apply in the context of ESD however?

Because the goal is not to teach programming to students in ESD courses, but rather to teach them how to work professionally, we think it would be most appropriate to use a low level of gamification with minimal impact on the general functions of existing tools. A more impactful type of gamification would involve creating new activities for employees. An example of this is the earlier mentioned interactive programming game CodeCombat¹¹ or the adaptation of an existing e-learning environment to include interactive tutorials and assignments as researched by Simões et al. (2013). A lower impact method of gamifying an existing system is rewarding users for actions with points and providing them with ways to view their own points and those of others.

Since an ESD course has a limited time frame in which students are employees for the course, there is a lesser need for using intrinsic motivators such as job satisfaction over extrinsic motivators, such as rewards in form of points or grades. In one's career however, we believe it would be best to be extrinsically motivated to do things a

¹⁰<http://dealbook.nytimes.com/2014/01/13/three-former-rabobank-traders-charged-in-libor-case/>

¹¹<http://codecombat.com> an interactive game that teaches programming

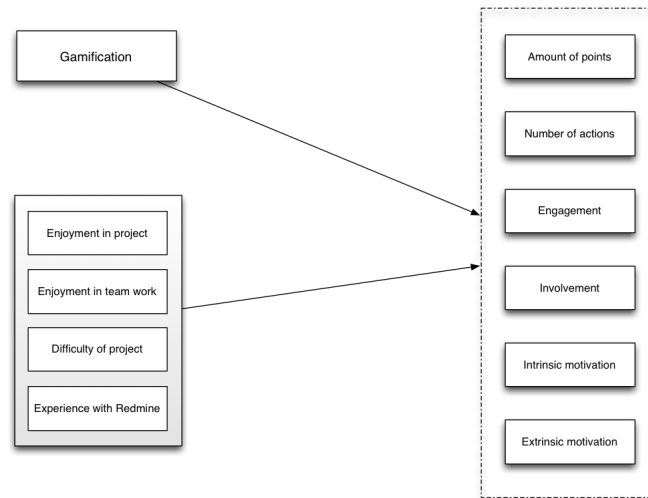


Figure 2.5: Conceptual model

certain way. With ESD however, we believe it is important for students to do things as intended by the course's creators; they will be introduced to this way of working and a chosen set of tools, really use them for a limited period of time and are then free to work in whichever way they want after the course. Through extrinsic motivators, such as points received for actions, students will quickly learn what behaviour is desirable and be motivated to perform that behaviour, the short length of the course possibly negating the issue of short term effects of extrinsic motivators.

2.6 Conceptual model

In order to summarize the presented theory we propose the conceptual model as seen in Figure 2.5.

Methodology

In the following chapter, we will describe the methods used for collecting and analysing our data.

3.1 Method of data collection

We used two types of data collection, namely an experiment where we gathered statistics on performed actions and points and a survey at the end of the semester where we used validated scales to measure the students' engagement, involvement and motivation. Finally, we collected all the students' grades and added them, anonymously, to the data set.

3.1.1 Experiment

In an experiment, a researcher assesses the effect of independent variables, by manipulating them, on dependent variables. The manipulated independent variable is referred to as the 'treatment' or 'factor' (Hair et al., 2010, p. 241). The subjects are divided in experimental groups that each receive different treatments in order to analyse the effect of the treatment on the dependent variables. Dependent variables are then measured after exposing the experimental groups to the treatment. The mean difference between the groups will then be analysed to see what the effect of the manipulation has been. The researcher must then apply data analysis techniques appropriate to his model to be able draw valid conclusions (Field, 2009, Hair et al., 2010).

Often however, these conclusions cannot be drawn on just the values for the dependent variables. There could be other factors that influenced the results as measured in the experiment. As an example, a study on whether people who swim regularly are less prone to having heart attacks, must also take into consideration whether subjects have other habits that might influence their health, such as smoking or eating a lot of vegetables. Not everything will be relevant to the study however, so researchers must take care into selecting an appropriate set of these 'control variables' for their research (Hair et al., 2010).

Applied gamification

As part of our experiment, we applied game elements to an existing open-source project management tool, Redmine, and applied this in the context of our case study 'GiPHouse'.

As described in the theoretical background of this study (chapter 2), gamification involves a broad spectrum of concepts and specific elements that can be used in gamifying (digital) systems. Domínguez et al. (2013) describe a high-intensity gamification of an e-learning system, where activities have been created that are part of the new gamified experience. This involves wizards, interactive tutorials, homework and specific actions for teachers to support these. We created a lower intensity system where no extra processes were created for teachers; a 'set it, and forget it' experience. Other examples of low-intensity gamifications can be found in Van der Waals (2014) and Engqvist & Frommen (2008).

Redmine allows users to perform various actions such as creating wiki pages, logging hours and keeping track of news, and students received points for performing these actions. Our system gave out these points only once a day for some actions (for recurring actions such as logging in), or just once all together for others (for atomic action such as closing a specific issue). However, all actions were logged whether they gave a student points or not. We made use of three experimental groups. Group 1 could see their points and in which fields they earned them, compare their scores to those of others and see their position compared to others on a leaderboard. Group 2 could only see their points and in which fields they earned them, but could not compare themselves to others. Group 3 did not have any gamification applied to their screens, although they received points for their actions in the background, they could not see any part of the gamification at all. In order to not influence the results, we did not inform students that the experiment was taking place. The points a student had, were constantly displayed next to their name in a sidebar. There was also a score dashboard where students could see their point history, see a timeline of their points and see a profile that showed them where they received their points in. Students in the group that could compare their points to those of others, additionally saw a leader board where they can see their ranking and they can compare their timeline and profile to the mean of all students.

Technology

Our case study ESD course, GiPHouse, uses Redmine for their project management. Redmine is an open-source tool built in Ruby on Rails, which has many features including wikis, issue tracking, time tracking, and an announcement system.

Redmine is expandable through a plug-in system. We used this system to add our gamification elements through a custom plug-in which we named 'Scoreme'. The choice of elements used in this gamification is described in the next section. In order to enable further development of our system, we have open sourced this plug-in and

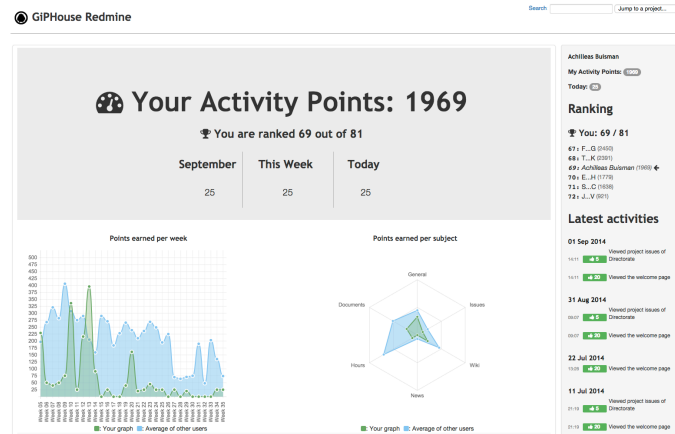


Figure 3.1: Gamified Redmine dashboard with comparison

added it to Redmine’s plug-in listings.

Elements

In line with the nature of a student’s course, results being expressed in grades, we have aimed our gamification towards a point system as an extrinsic motivator for students to perform certain actions. This is a common pattern in many of today’s popular websites, like the h-index as described by (Engqvist & Frommen, 2008) and the popular Questions and Answers website [stackoverflow¹](http://stackoverflow.com/) which uses points as a basis for its community building.

The main elements from games that we used in our manipulation are points and leader boards.

Points Students received points for certain actions on the system. These actions were chosen for the specific case study of GiPHouse, based on what the educators found important for either the successful execution of projects or the continuation of information within GiPHouse after each semester. The amount of points received for each action has been determined based on how wanted these behaviours are for GiPHouse. Points as a reward system affect emotional motivation as described by Lee et al. (2011).

Leader boards Students were able to see how many points they have and how their amount of points positioned them against other students (depending on the experimental group they were assigned to). We decided to anonymise the leader board so that students could only see anonymised versions of the two people above and below themselves. Additionally, students could see in which areas of Redmine they scored points, and compare this to the mean profile of all other students combined in

¹<http://stackoverflow.com/>

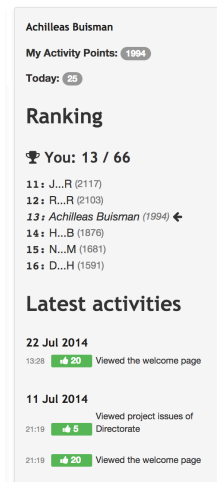


Figure 3.2: Redmine leaderboard

the dashboard. The leader boards inspire competition and trigger on a social level (Lee et al., 2011, Perryer et al., 2012). A screenshot of the leader board can be seen in Figure 3.2.

3.1.2 Survey

In addition to our experiment, we conducted a survey in which we used validated scales for determining engagement, motivation and involvement amongst all students of the course. In the following section we will discuss the scales we used to measure our variables. In our survey we differentiated between dependent variables and extraneous variables.

3.1.3 Dependent variables

Gamification is used as a means to increase user engagement. We therefore, firstly, wanted to measure the effect of gamification on student engagement. In order to measure engagement we used the validated scale of Schaufeli et al. (2002) that contained 17 items. We reduced the scale to nine items, based on content validity and item loading, in order to adapt the scale to our context and the needs of our study (see Appendix 6.1).

The short-term nature of (ESD) courses at universities often involves having to motivate students for the duration of the course rather than for a longer period of time. Short-term motivation could be realised through extrinsic motivators. We were interested in exploring whether the gamification of tools had an effect on the nature of motivation with students - intrinsic or extrinsic. We used Vallerand et al. (1992)'s Academic Motivation Scale to measure intrinsic and extrinsic motivation. Again, we selected items based on their applicability and our needs (see Appendix 6.1).

Finally, we wanted to know if gamification has had an effect on the level of involvement with students towards the course's organisation. In order to measure involvement we used Zaichkowsky's scale (Zaichkowsky, 1994) of which we used all items (see Appendix 6.1).

3.1.4 Independent (extraneous) variables

As stated above, there could have been other factors that influenced the results of our study. We theorised which factors these could be and added items for these to the survey. An example of one of the items for our control variables was "To what extent did you enjoy working on your project?" Our control variables were answered on a 5-point Likert scale (see Appendix 6.1 for all items).

3.2 Data analysis

We have collected different types of data: points were given for each action on Redmine and each action was logged, students completed a questionnaire wherein we measured their engagement, involvement and motivation and finally grades were collected for each student and added to the data set. Collecting and joining these data was done through scripts that also anonymised the information.

Before interpreting the collected information we had to analyse whether the gamification we applied is responsible for possible differences in our dependent variables, or if some other factors were responsible. In order to do this, we used analysis of variance. Because our main independent variable, a subject's experimental group, is categorical and because we have multiple metric dependant variables, we used a Multivariate Analysis of Variance (MANOVA). Additionally, since we also had multiple metric independent variables that we used as covariates, we have also made use of Multivariate Analysis of Covariance (MANCOVA) to analyse our data (Malhotra & Birks, 2007, p. 562). We also performed a post hoc analysis to discover the specific differences between experiment groups (Hair et al., 2010). For our analyses, we used the data analysis tool SPSS.

Analysis and results

In this chapter, we will present the results of the analysis of the data collected during our experiment. We have combined three series of data, namely:

- (Game) Points and action logs collected in Redmine
- Survey responses
- Grades given at the end of the course

4.1 Missing value analysis

Before conducting our main analyses, we checked our data set for missing data. Initially, there were 66 students who took part in the course. There were no missing values on individual variables. However, despite the mandatory nature of the survey, there were seven students who did not fill in or complete the survey. As this leads to missing values on several dependent variables, these respondents were removed from the data set.

In total, after our missing values analysis, we had 59 observations in our data set.

4.2 Outlier analysis

Outliers are observations that, compared to the majority of observations in the data set, have a distinctly different value for a certain variable (univariate outliers) or a distinctive set of values for a combination of variables (multivariate outliers). Usually, the difference entails an unusually high or low value. Having outliers is not a bad thing per se; one must evaluate the nature of outliers and judge whether they fit the nature of the collected data (Hair et al., 2010, p.64).

Yet, since outliers can have a significant effect on the results of our data analysis, we must find them and determine whether they truly represent a segment of our population, or that they represent erroneous data. Hair et al. (2010) state that outliers should be retained in order to maintain generalizability, unless it can be proven that they are truly abnormal (Hair et al., 2010, p.67).

4.2.1 Univariate outlier analysis

Univariate outliers have unusual high or low values for certain variables. In order to detect outliers, we converted the values of all metric variables to their standard scores.

For small data sets, Hair et al. (2010) state a rule of thumb that outliers are those observations that show a standard score above 2.5 or below -2.5 (Hair et al., 2010, p.66).

We had two outliers in the survey where further analysis revealed a suspicious answer pattern and so these observations were removed. There were five other outliers in the ‘Amount of points’ received and ‘Grade’ yet we retained these for the sake of generalizability (Hair et al., 2010, p.67). Retaining these outliers will have consequences for our statistical assumptions (e.g. non-normality). These are described further below in the section statistical assumptions.

4.2.2 Multivariate outliers

Because we had a fair amount of variables, we did a multivariate outlier analysis to determine if there were any sets of variables with a combination of values that might qualify as an outlier. For this, we calculated the Mahalanobis D^2 and its associated probability. The resulting probabilities for our observations all were at least 0.001 which means that our data set did not contain any multivariate outliers (Hair et al., 2010, p.66).

4.3 Sample

After removing missing values and univariate outliers, our final data set had 57 observations. These were all students who followed the GiPHouse course between January and June 2014 as part of either their Bachelor or Master courses for the studies of Artificial Intelligence, Information Science and Computer Science at the Radboud University in Nijmegen. As can be seen in (Table 4.2), most respondents were enrolled in a bachelor study (84.2%), while 15.8% were master students. For 82.5% it was the first course at GiPHouse, while the other 17.5% already had worked for GiPHouse before. With respect to Redmine, 80.7% of the respondents stated that they had no experience with Redmine prior to the course. 3.5% had little experience, 8.8% had some experience and 7% of the students stated to have a lot of experience with the program.

Table 4.1: ‘First course’ at GiPHouse

First GiPHouse Course	Percentage
Yes	82.5%
No	17.5%

Table 4.2: Student study phases

Study phase	Percentage
Bachelor	84.2%
Master	15.8%

Table 4.3: Frequencies of 'Prior experience with Redmine'

Prior experience with Redmine	Percentage
None	80.7%
Little	3.5%
Some	8.8%
A lot	7.0%

4.4 Statistical assumptions

For statistically valid results in our main analysis, we should ideally meet certain statistical assumptions for the test to be as accurate as possible. Our dependent variables should be normally distributed, the observations should be independent from each other and there should be homoscedasticity across the groups (Hair et al., 2010, p.479).

4.4.1 Normality

We tested for normality using the absolute values of skewness and kurtosis where we used a threshold value of ± 3 and we saw a leptokurtic (peaked) distribution on the variables 'Amount of points', 'Number of actions' and 'Grade' (Hair et al., 2010).

A non-normal distribution of our dependent variables, especially with a smaller sample size, is problematic for the results of our MAN(C)OVA. We therefore had to remedy the non-normal distribution of the variables. In addition to this, remedying non-normal distributions also helps in meeting the rest of our statistical assumptions (Hair et al., 2010, p. 72,74).

We applied a log transformation to remedy the non-normal distribution. This transformation was only effective in normalizing the variables 'Amount of point' and 'Number of actions', but it worsened the distribution of the variable 'Grade'. We therefore chose to exclude this variable from our further analyses, as we could not assure valid test results for it (Field, 2009).

4.4.2 Homogeneity of covariance matrices

Next, we needed to test the homogeneity of the covariance matrices. In this assumption, we tested whether the three categories of our independent variable 'group' display an equal amount of variance on the different dependent variables. Based on a non-significant Box's M test result, we can assume that the variances of the dependent variables are stable across the different experimental groups (Hair et al., 2010, p.480).

4.4.3 Independence of observations

As stated by Hair et al. (2010), the most basic but also the most important assumption underlying MANOVA concerns the independence of observations. For our statistical test results to be valid, the responses in one group are not allowed to be dependent on

responses in any other experimental group. Although there are no statistical tests to prove the independence of observations, we ensured it as much as possible by the random assignment of the experimental manipulations (Hair et al., 2010, p.458).

4.5 Reliability analysis

Although we made use of validated scales, we wanted to test their reliability. We therefore checked whether they measure what they were intended to measure and whether all items were consistent in their values (Hair et al., 2010, p.93).

We used Cronbach's alpha as the reliability coefficient for our scales and we examined the effects of removing individual items from a particular scale to see whether the deletion of any item could improve the scale. For the Cronach's alpha, we used the generally accepted lower limit of 0.70 (Hair et al., 2010).

For the variable 'Engagement', we had a good reliability with a Cronbach's alpha of 0.88 and removing items would not yield a better figure. Our scale for the variable 'Involvement' had an alpha of 0.91, also without possibilities of improving the value by removing items. The Cronbach's alphas for Intrinsic and Extrinsic motivation respectively were 0.84 and 0.79, again with no room for improvement by removing items.

Since validation using Cronbach's alpha determined that there was no room for improving the used scales by adjusting which items we selected to analyse them, we left all measured variables intact.

4.6 Descriptive statistics

Table 4.4 presents the descriptive statistics of our dependent variables. The variables 'Engagement', 'Involvement', 'Intrinsic motivation' and 'Extrinsic motivation' were measured on a 7-point Likert scale.

Table 4.4: Descriptive statistics on our dependent variables

	Minimum	Maximum	Mean
Amount of points	1791	15473	4716.54
Number of actions	154	3084	703.65
Grade	5.0	9.0	7.3246
Engagement	1.67	6.11	4.1676
Involvement	2.0	7.0	4.6053
Intrinsic motivation	2.0	7.0	4.4772
Extrinsic motivation	2.17	6.83	4.2573

Table 4.5 shows the values for the variables crossed on the experimental group that students were in. What draws attention is the fact that students who were gamified and could compare their points to those of others have performed more actions on Redmine and scored more points. Counter-intuitively we see that the second group,

‘Gamified without comparison’ displays higher levels for the other dependent variables such as ‘Engagement’ and ‘Involvement’.

Table 4.5: Cross table of experimental groups on dependent variables

Group	Eng.	Inv.	Mot.int	Mot.ext	#Points	#Actions
Gamified with comparison	3.95	4.46	4.20	4.10	6042	962
Gamified	4.58	4.97	4.95	4.56	4079	581
No gamification	3.98	4.38	4.29	4.11	3916	546

An interesting statistic, as shown in Table 4.6, is that hardly any student (mean: 1.4) had prior experience with Redmine. This is most likely due to most students being in their Bachelor phase and that most students have had no previous GiPHouse courses. It also seems that the difficulty of projects is rated fairly low (mean: 2.4) considering that most students have little to very little experience in working on professional projects.

Table 4.6: Statistics on our covariates

	Minimum	Maximum	Mean
Enjoy working on projects [1-5]	2	5	3.75
Enjoy working with team [1-5]	3	5	4.35
Difficulty of project [1-5]	1.0	4.0	2.40
Prior experience with Redmine [1-4]	1	4	1.42
Opinion on Redmine [1-7]	1	6	4.28

In our survey, we asked the students whether they had noticed the game elements added to Redmine. Of the 39 respondents who were in the gamified groups, four respondents indicated that they did not see the game elements we implemented. Strangely, of the 18 respondents who did not have any gamification applied to their interface, two did notice the gamified system. Table 4.7 describes the distribution of the answers to our manipulation check questions.

Table 4.7: Descriptive statistics for manipulation checks

	Minimum	Maximum	Mean
Compared own points to those of others	1	5	2.33
Know which activities earn points	1	5	2.97
Know how many points activities are worth	1	5	2.28
Performed actions in order to gain points	1	5	1.86
Enjoyed receiving points on Redmine	1	5	2.97

4.7 MANOVA

In Table 4.8 we can see the distribution of the students over the experimental groups. Due to practical constraints, namely the number of students enrolled in the GiPHouse course, we have a relatively small experimental sample.

As shown in Table 4.5, there are differences between the groups concerning the means of the various dependent variables, so we wanted to analyse whether these

Table 4.8: Experimental group size

Experimental Group	Size	Percentage
Gamified with comparison	20	35.1%
Gamified	19	33.3%
No gamification	18	31.6%

differences are significant and thus test whether the experimental manipulation had an effect on the dependent variables. In order to do this, we applied a MANOVA as described in the chapter Methodology under Data Analysis (section 3.2).

The MANOVA we performed, showed that the overall effect of gamification on our dependent variables is not significant (Pillai's Trace: $F[12, 100] = 1.47, p=0.150$). However, when examining the separate ANOVA's, we can see a significant effect of the gamification on the 'Amount of points' that students received in Redmine ($F[2, 54] = 5.03, p=0.01$) and the 'Number of actions' they performed in Redmine ($F[2, 54] = 5.16, p=0.01$). We can also see a significant effect of the gamification on the student's engagement ($F[2, 54]=3.73, p=0.03$). However, there was no significant effect of the gamification on 'Involvement', 'Intrinsic motivation' or 'Extrinsic motivation'. We must keep in mind however that the MANOVA does not take covariates into account. We will look at these in the following analyses (section 4.8).

4.8 MANCOVA

We wanted to be able to measure whether other factors might have had a significant impact on our dependent variables. To this purpose, we measured covariates (control variables) in our survey. To test whether these covariates have influenced our dependent variables significantly, we applied a MANCOVA in which we added covariates to calculate their effect (Hair et al., 2010, p.440).

Although we had measured five control variables in our survey, we could not use all of these in our analysis. Hair et al. (2010) state that researchers should limit the number of included covariates as much as possible as they can make the statistical tests very sensitive. Thus, before running the MANCOVA, we needed to select viable covariates for the analysis. The authors present a rule of thumb for calculating the maximum number of covariates: (Hair et al., 2010, p.257).

$$\text{"Maximum number of covariates} = (0.10 \times \text{Sample size}) - (\text{Number of groups} - 1)\text{"}$$

In our research, this means the maximum number of covariates is three.

Effective covariates correlate highly with the dependent variables, but not with the independent variable (Hair et al., 2010, p.456). To analyse the correlations between the variables, we used Pearson's correlation test (Field, 2009, p.178) which showed that 'Enjoyment in the project' relates significantly with 'Engagement', 'Involvement', 'Intrinsic motivation' and 'Extrinsic motivation' ($r=0.57, p=0.00, r=0.68, p=0.00$,

$r=0.67$ $p=0.00$, $r=0.51$ $p=0.00$ respectively). Additionally, 'Enjoyment in the project' had a non-significant correlation with the variable 'Group' ($r=0.57$ $p=0.67$). Our two other dependent variables 'Amount of points' and 'Number of actions' showed a strong correlation with the control variable 'Experience with Redmine' ($r=0.45$ $p=0.00$ and $r=0.50$ $p=0.00$ respectively). We therefore chose 'Enjoyment in the project' and 'Experience with Redmine' as the covariates of our MANCOVA.

The MANCOVA showed a strong significant effect of both covariates on the set of dependent variables (Pillai's trace $F[6, 47]=7.50$, $p=0.00$) for 'Enjoyment in the project' and for 'Experience with Redmine' ($F[6, 47]=4.53$, $p=0.001$). Also, after entering the control variables into our analysis, the variable 'Group' no longer had a significant effect on our dependent variables ($F[12, 96]=1.44$, $p=0.161$).

When we looked at the individual ANCOVA's that were performed as part of the MANCOVA, we could see the individual effects of the variables we used while taking into account the effect of the other independent variables used in the analysis. Here we could see that the effect of 'Enjoyment in the project' was strongly significant for 'Engagement', 'Involvement', 'Intrinsic motivation' and 'Extrinsic motivation' as seen in (Table 4.9). Furthermore, we learned that 'Enjoyment in the project' accounted for all of the significance of 'Group' on 'Engagement' that first was found by the MANOVA. In the ANCOVA for 'Group', we could see no significant effect on 'Engagement' any more ($F[2, 52]=1.25$, $p=0.30$).

Table 4.9: ANCOVA results for 'Enjoyment in project'

Dependent Variable	F	p
Amount of points	0.014	0.905
Number of actions	0.003	0.956
Engagement	20.672	0.000
Involvement	37.056	0.000
Motivation Intrinsic	34.512	0.000
Motivation Extrinsic	14.534	0.000

We found that the student's experience with Redmine has a significant influence on the 'Amount of points' and 'Number of actions' performed in Redmine (both $p=0.00$) as can be seen in Table 4.10. However, the variable has no significant influence on the other dependent variables. Although 'Experience with Redmine' had a significant effect, and accounted for some of the variation, the applied 'Group' still had a significant effect on both the 'Amount of points' and the 'Number of actions' as presented in (Table 4.11). Nevertheless, it had no significant effect on the other dependent variables.

4.9 Post hoc comparisons between individual group differences

We were also interested to discover which experimental manipulations exactly had which effect on the dependent variables. In the MANOVA, we statistically tested

Table 4.10: ANCOVA results for ‘Experience with redmine’

Dependent Variable	F	p
Amount of points	19.19	0.000
Number of actions	24.98	0.000
Engagement	0.846	0.362
Involvement	0.029	0.865
Motivation Intrinsic	0.158	0.693
Motivation Extrinsic	0.451	0.505

Table 4.11: ANCOVA results for ‘Group’

Dependent Variable	F	p
Amount of points	7.43	0.001
Number of actions	8.28	0.001
Engagement	1.25	0.296
Involvement	0.500	0.609
Motivation Intrinsic	0.713	0.495
Motivation Extrinsic	0.154	0.857

whether there is any significant difference between the three experimental groups at all. However, we could not see where exactly the difference lies. We saw, for example, a significant effect of the variable ‘Group’ on ‘Amount of points’, but we could not see which groups differed from each other significantly. To analyse this, we performed a post-hoc test for the dependent variables that our independent variable ‘Group’ had significant effect on in the MANOVA, namely, ‘Amount of points’ and ‘Number of actions’. Since we could assume homogeneity of the covariance matrices, and since our experimental groups contain a fairly equal amount of respondents, we performed a post hoc comparison based on Tukey’s honestly significant difference (HSD) method (Hair et al., 2010).

The results of the post hoc comparison showed that there is a significant difference for ‘Amount of points’ between group ‘Gamified with comparison’ and the other two groups: ‘Gamified without comparison’ and ‘No gamification’ ($p=0.025$ and $p=0.021$ respectively). There is however no significant difference between the groups ‘Gamified without comparison’ and ‘No gamification’ ($p=0.993$). We could therefore say that our experimental manipulation caused the bigger ‘Amount of points’ scored by students in the ‘Gamified with comparison’ group when compared to both other groups, as can be seen in Table 4.5.

Furthermore, we could again see a significant difference between the group ‘Gamified with comparison’ and both other groups ‘Gamified without comparison’ and ‘No gamification’ ($p=0.018$ and $p=0.025$) for the variable ‘Number of actions’. As before, there was no significant difference between the groups ‘Gamified without comparison’ and ‘No gamification’ ($p=0.996$). Here as well, we can say that ‘Gamification with comparison’ has had a positive effect on the ‘Number of actions’ that students performed.

Discussion and conclusion

In this study, we researched to what extent applying gamification to a project management tool would affect student activity, engagement, involvement and motivation. The reason for doing this is that in ESD we can see that students have to learn many new skills in order to complete projects with real clients successfully. Part of these new activities is the use of tools and methodologies that they are not yet familiar with and that have been chosen for them by an organisation, in this case their teachers. We wanted to research to what extent gamifying existing systems would benefit the process of adopting new behaviour.

Gamification was defined as ‘the use of game design elements in non-game contexts’ (Deterding et al., 2011, p. 9) in order to increase engagement and motivate certain behaviour. Examples of this are the use of badges, avatars, leader boards, elements of time pressure and many other elements found in games. Gamification can be applied at different scales and at different levels and goals (Danforth, 2011, Deterding et al., 2011, Domínguez et al., 2013, Perryer et al., 2012, Simões et al., 2013).

In our research, we have applied a relatively subtle gamification to the project management tool used by students. We had three experimental groups: a control group that had no gamification, a group that received points for various actions within the system and could only see their own points, and finally a group that received these points and could additionally compare their points to those of others. Examples of activities that a student received points for are the creation of a wiki page or the logging of hours.

The gamification functionality we built, did not only give points for certain activities, but it also kept track of all of the activities of students on the system. These numbers (points and number of activities) were (anonymously) combined with student grades. Additionally, we measured self-reported engagement, involvement and intrinsic and extrinsic motivation through a survey held at the end of the semester.

Our results showed that self-reported student engagement, involvement and motivation were not significantly affected by the applied gamification. This means that students who were exposed to the gamification were no more enthusiastic about their work, they were not more bursting with energy or happier while working than the students who were not gamified.

The amount of points and number of performed actions *were* influenced significantly however. Students who received points and could compare their score to others had significantly more points and performed more activities than those who were not

gamified. This means that adding a point system with comparison feature to the project management tool used by students increased their use of the system. At the same time, the fact that there was no significant difference between the students who just received points, but could not compare themselves to others and both other groups means that adding the comparison feature made the crucial difference. It is strange, however, that motivation was not influenced significantly as suggested by Lee et al. (2011), who suggest that social motivation is triggered by being able to compare oneself to others.

Our analysis also showed that previous experience with Redmine was responsible for a significantly higher number of actions and points. This might mean that students who have previously worked with Redmine are more comfortable with performing the tasks that are important to our case study's way of working. Based on our analysis, we also know that there was a significant effect from the students' experience with Redmine if the course they followed during the experiment, was their first GiPHouse course.

Moreover, we saw a significant effect of enjoyment in the project on engagement, involvement and motivation. This means that when students are enjoying their project, they will lose themselves in their work and they will feel that there is meaning and purpose in it. We also saw that enjoyment in working with the team had a significant effect on the enjoyment in the project. This seems logical, but it is important to note.

In our survey, we asked the students who noticed the gamification (35 out of 39 gamified students) a few questions about their experience with the gamification elements. Students generally enjoyed receiving points on Redmine and indicated to know which activities would earn them points. Students have indicated to not have performed tasks in order to gain points, which could mean that they performed the tasks as a necessary part of their work.

There was an open question in the survey that asked students for any remarks they had regarding the gamification system in Redmine. Most comments indicated that students were unclear as to the purpose of the system. When taking into consideration that students were not informed about the system at all this is an expected reaction. Some students were critical of the amount of points that were given for certain actions and presented alternatives. This shows that these students spent some amount of time analysing the point system and coming up with ways in which to improve it.

5.1 Theoretical implications

The gamification that we applied to the system, had a significant effect on its use by students. This means that gamification can be an effective tool to stimulate the actual use of systems within an ESD context. The fact that it was the group who could compare their points with those of others that had a significant effect on the number of actions, suggests that there is a certain need for social dynamics when it comes to gamification. Just receiving points apparently, was not enough. This means that when aiming for stimulating the use of such tools, system designers should use socially stimulating game elements like leader boards and comparisons of profiles as suggested

by the literature (Deterding et al., 2011, Lee et al., 2011, Perryer et al., 2012).

As far as we are aware, there has been no research in applying gamification to full sized software development projects within an educational context. We have shown that applying gamification in this context can yield positive, significant results. Results that are closely in line with what is important in an educational context, namely, adhering to the concept of ‘practice makes perfect’. Stimulating users to use presented tools more will make them more experienced in using them. We see this as especially useful for activities that are very new to students, since this is where they will learn the quickest, but will most likely need the most encouragement.

We could see that the enjoyment of the project lead to an increase in engagement, involvement and motivation amongst students. The literature on gamification often mentions fun as an important element of gamification (Deterding et al., 2011, Domínguez et al., 2013, Perryer et al., 2012, Ryan & Deci, 2000b, Simões et al., 2013, Wang & Sun, 2011). Our result with respect to enjoyment having a positive effect on engagement is in line with the observations made in the literature. In turn, we could see that enjoyment in working with the team had a significant effect on the enjoyment in the project.

Gamification is often used as a means to increase engagement (Deterding et al., 2011), yet our results did not confirm this effect. Nevertheless, we do not believe that this invalidates the effect that gamification *can* have on engagement. The application of gamification in this research has been very minimal and mostly aimed at extrinsic motivators (points) (Perryer et al., 2012, Ryan & Deci, 2000a). Extrinsic motivators often only have a short-term effect that weakens over time (Perryer et al., 2012). We measured self-reported engagement, involvement and motivation through the survey at the end of the semester. It might be that a short-term effect of the gamification had passed by then, which could explain why we could not find a significant effect in our data.

5.2 Practical implications

How much students enjoy their projects has a significant effect on their engagement, involvement and motivation. This means that if we want students to be engaged, we will have to provide them with enjoyable projects. This is easier said than done however. ESD courses are very dependent on the projects they attract. A project will have to meet certain requirements in terms of size, difficulty and sometimes domain and clients will have to be prepared to wait longer for their product and accept uncertainties in terms of delivered quality. We propose that ESD courses take into account how much fun a project is when deciding which projects to take on. Students already have to learn new skills and apply them directly; them enjoying the work will only help in their understanding of the purpose of these new skills.

Furthermore, enjoyment in the team showed a significant role in the enjoyment of the project itself. Teams can be formed in different ways: according to experience with

a project's programming language, based on equal division of disciplines (in the case of multi-disciplinary teams), and many others. We think it might be beneficial for ESD courses to take actions to determine whether teams fit together well. It might be a good idea to let students form their own teams rather than assigning them to teams (as was the case with our case study). Alternatively, students could perform personality surveys that will allow personality based matching such as suggested by Belbin (1993). Also, a recommendation we can make is to research whether team building exercises will increase the students' enjoyment in their projects. This could be a potentially valuable addition to ESD courses.

Though we did not see a significant effect of our gamification on the attitude of students, we did see a significant difference in the amount of actions they performed on Redmine. The fact that we can encourage the use of systems through gamification means that we can steer students into positive behaviour. As we stated before, this would be especially useful for new types of skills that students need to develop, such as making many atomic commits in version control systems, creating branches rather than pushing to master, writing automated tests, etc. Gamification to stimulate these wanted behaviours can be applied without requiring extra work from teachers. This means that students will be actively stimulated to perform work in a way that teachers would like to see, without requiring individual supervision.

5.3 Limitations and future research

Due to practical limitations in terms of the size of the experimental groups, we had a small data set to analyse. Ideally, we would have had more students take part in the course, which was not possible for us. The effect of this smaller sample size is that we have less external validity. Also, there might be factors that would have shown significant effects if the sample size had been bigger. It would therefore be good for this research to be repeated with larger experimental groups.

We feel that the amount of points students received for actions could have been balanced better. Some small actions, such as logging in, gave students a large amount of points while other activities that seem more important reward fewer points. An example is that logging in earns you 20 points whereas creating a wiki page only earns you 60. Students indicated the same in the survey and also stated that it was possible to artificially generate points, by, for instance, downloading specific files from the wiki without truly needing them. Creating a more balanced point distribution would benefit the system as a whole.

Our study used extrinsic motivators. These might have had short-term effects on our dependent variables that were measured at the end of our study through our survey (Perryer et al., 2012). A long-term experiment with multiple measures of engagement, involvement and motivation would give insight in how these are influenced over time, over the course of a project or multiple projects.

Apart from mending these limitations, we have several recommendations for future

research that could add to academic knowledge around gamifying ESD courses and bettering the tools and techniques we can use in ESD.

We have found no research on the subject of ESD specifically with regards to manipulating engagement and motivation. We do believe, however, that professional software development within an educational context is a subject that merits exploration as a separate concept. In ESD, students have to learn new skills and work in a different state of mind than they do in traditional courses. Next to having to develop new skills, working professionally in a course is different from a career: it is short-term, incentives are different and teams are formed differently, to name a few contrasts. This means that ESD course designers will face unique scenarios. A question that might require answering is whether training in the tools used could improve the ESD courses. We saw a significant effect of prior experience with the project management system on the amount that the system was used. Workshops and training early in the semester for specific tools or methodologies might also have a positive effect on their use within teams.

We measured engagement, involvement and motivation through *self-reported* variables in a survey. This might yield different results than other means of measuring these concepts. It might be interesting to further our research by measuring our dependent variables by using ‘implicit measures’, which might give a more honest measure of these concepts (Fazio & Olson, 2003). This could yield interesting insights into how student’s attitude and motivation can be manipulated through gamification.

Another possible improvement is the use of extrinsic motivators like badges and levels in conjunction with other, intrinsically motivating, game elements that relate more closely to someone’s identity. An example of this are characters and avatars from which users will derive social status, thus becoming part of a more intrinsically motivating experience (Perryer et al., 2012). Using more intrinsically related game elements might yield different results than our research has and might be beneficial to ESD in different ways than our experiment has been. There are many unexplored combinations that can still be made between game elements and elements of ESD courses.

Gamification does not only apply to digital systems, although Deterding et al. (2011) mention, that this is the most common application. Challenges in classes, prizes or story driven game experiences might all be made part of the execution of ESD courses. The effects of which might prove valuable in improving student behaviour and attitude. An example would be to hold quizzes in lectures where teams would need to answer questions about the projects of other teams. This might increase the overall involvement of team members with the course as a whole.

We divided whole project groups, rather than individuals, randomly between the experimental groups. We did this to prevent team members from having different looking dashboards, which might have been confusing or worked aversely on their attitude towards the point system. It might be useful to analyse whether applying the different levels of gamification (compare, just points, none) heterogeneously within

project teams will have different effects.

The leader board system we applied, showed anonymised entries where only the first and last letters of someone's name were visible. The comparison feature did have a significant effect however. We therefore suggest a study with full names and more comparison features as a possible way to gain more from gamification in this context.

We used a narrow set of data in our research, namely, activities and points from Redmine through a self built gamification plugin and a survey. There are, however, more pieces of information that could be gathered in the duration of a course. For example, version control commits could be counted and analysed for size and commit message length. Additionally, committed code could be automatically analysed for style and tested through automated tests. These automated analyses could then be translated into, for example, points and achievements. Think of a 'Fixed broken Unit Test' badge or a 'Git rebase to fix code style'.

We used positive feedback; students only received positive amounts of points for certain wanted actions. Negative feedback, however, also has an important place in gamification and can be used to adjust unwanted behaviour (Perryer et al., 2012, p. 377). Players will quickly learn what they cannot do in a game when they, for example, lose health or virtual money. This could be applied to project management as well. An example could be that a user receives negative points or loses 'health' when they neglect to look at the company news announcements for a week or they forget to log hours. We can see this pattern of using negative feedback in, for example, HabitRPG¹ where not performing daily tasks results in a loss in 'health'. We suggest research into how negative feedback can be used in gamifying ESD tools.

5.4 Conclusion

Adding a point based gamification with comparison feature to the project management tool of an ESD course proved to have significant effect on the amount of use by students. Engagement, involvement and motivation were not significantly affected by the gamification, but by the amount of enjoyment students experienced in the project. We are positive about the results and about the possibilities for further exploring the options for applying gamification in ESD courses in order to guide students into using the tools and methodologies that are presented to them during the courses. This case study may also serve as an example for teachers to apply gamification in their own ESD courses.

¹HabitRPG - <https://habitrpg.com/>

Bibliography

- Belbin, R. M. (1993). *Team Roles at Work*. Oxford: Elsevier Ltd, first edit edition.
- Caillois, R. (2001). *Man, Play, and games*. Urbana and Chicago: University of Illinois Press.
- Cohen, A. M. (2011). The Gamification of Education. *The Futurist*, 45(5), 16--17.
- Danforth, L. (2011). Gamification in Libraries. *Library Journal*, 136(3), 84--84.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, number 10.1145/2181037.2181040 (pp. 9--15).
- Domínguez, A., Saenz-de Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63(1), 380--392.
- Engqvist, L. & Frommen, J. G. (2008). The h-index and self-citations. *Trends in ecology & evolution*, 23(5), 250--252.
- Fazio, R. H. & Olson, M. a. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual review of psychology*, 54(1), 297--327.
- Field, A. (2009). *Discovering Statistics Using SPSS*. London, UK: SAGE Publications Ltd, third edit edition.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective*. Upper Saddle River, NJ: Pearson Education, 7th edition.
- Huotari, K. & Hamari, J. (2012). Defining gamification: A service marketing perspective. *Proceedings of the 16th Iternational Academic MindTrek Conference*.
- Lawrence, P. R. (2011). What leaders need to know about human evolution and decision making from instinct to decision making. *Leader to Leader*, 2011(60), 12--16.
- Lee, J. J., College, T., Ph, D., Hammer, E., & Interdisciplinary, M. (2011). Gamification in education: What, how, why bother? *Academic Exchange Quarterly*, 15(2), 1--5.

- Malhotra, N. K. & Birks, D. F. (2007). *Marketing research - an applied approach*. Essex: Pearson Education Limited, third edit edition.
- Perryer, C., Scott-ladd, B., & Leighton, C. (2012). Gamification: Implications for workplace intrinsic motivation in the 21. "*Asian Forum on Business Education Journal*", 5(3), 371--381.
- Ryan, R. & Deci, E. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54--67.
- Ryan, R. M. & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, 55(1), 68--78.
- Schaufeli, W. B. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, 66(4), 701--716.
- Schaufeli, W. B., Salanova, M., Bakker, A. B., & Alez-rom, V. G. (2002). The measurement of engagement and burnout: A two sample confirmatory factor. *Journal of Happiness Studies*, 3(1), 71--92.
- Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, 29(2), 345--353.
- Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The academic motivation scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement*, 52(4), 1003--1017.
- Van der Waals, K. (2014). Gamification: Spelend Rijk. *Vrij Nederland*, 12(1), 44--49.
- Wang, H. & Sun, C. (2011). Game reward systems: Gaming experiences and social meanings. *Think Design Play: The fifth international conference of the Digital Research Association (DIGRA)*, January(1), 1--15.
- Zaichkowsky, J. L. (1994). The personal involvement inventory: Reduction, revision, and application to advertising. *Journal of Advertising*, 23(4), 59--70.

Appendix

6.1 Survey

6.1.1 Engagement

Adapted from Schaufeli (2006)

The following statements are about how you feel while working for GiPHouse. Please read each statement carefully and decide if you ever feel this way about working for GiPHouse. If you have never had this feeling, select 'never'. If you have had this feeling, indicate how often you felt it.

1. While working for GiPHouse, I feel bursting with energy.
2. I find the work that I do for GiPHouse full of meaning and purpose.
3. Time flies when I am working for GiPHouse.
4. I am enthusiastic about my work at GiPHouse.
5. My work for GiPHouse inspires me.
6. I feel happy when I am working intensely for GiPHouse
7. I am proud of the work that I do for GiPHouse.
8. To me, my job at GiPHouse is challenging.
9. I get carried away when I am working for GiPHouse.

Answers are given on a 7-point Likert scale:

Never	Almost never	Rarely	Sometimes	Often	Very often	Always
0	1	2	3	4	5	6

6.1.2 Involvement

Adapted from Zaichkowsky (1994) p. 70

To me Redmine(GiPHouse/My project) is:

1. Important ----- Unimportant

2. Boring ----- Interesting
3. Relevant ----- Irrelevant
4. Exciting ----- Unexciting
5. Means nothing ----- Means a lot to me
6. Appealing ----- Unappealing
7. Fascinating ----- Mundane
8. Worthless ----- Valuable
9. Involving ----- Uninvolving
10. Not needed ----- Needed

6.1.3 Motivation

Adapted from Vallerand et al. (1992)

Please indicate to what extent the following statements apply to your work for GiPHouse.

(Intrinsic)

1. While working for GiPHouse, I generally experience pleasure and satisfaction.
2. While working for GiPHouse, I experience pleasure in learning new things.
3. While working for GiPHouse, I feel satisfied in achieving personal goals.
4. While working for GiPHouse, I enjoy communicating my own ideas to others.
5. While working for GiPHouse, I feel satisfaction in being able to excel in my studies.

(Extrinsic)

1. Working for GiPHouse will help me get a more prestigious job eventually.
2. Working for GiPHouse helps me prepare for my career path later on.
3. Working for GiPHouse helps me feel important.
4. Working for GiPHouse helps me make a better choice regarding my career orientation.
5. Working for GiPHouse helps me show myself that I am an intelligent person.
6. Working for GiPHouse helps to improve my competence as a worker.

The questions were presented in a randomised order to the list above. Answers are given on a 7-point Likert scale:

Does not apply at all		Applies moderately		Applies Completely		
1	2 3	4	5 6	7		

6.1.4 Control variables

1. To what extent did you enjoy working on your project? (1--5)
2. To what extent did you enjoy working with your team members? (1--5)
3. How do you grade the difficulty of your team's project? (1--5)
4. Did you have experience with Redmine prior to this course? (1--5)
5. What is your opinion on Redmine in general? (1--7)
6. Do you have any remarks on using Redmine? (open)

General

1. Is this your first GipHouse course? (Yes/No)
2. Are you currently doing your Bachelor or Master? (Bachelor/Master)

6.1.5 Manipulation check

This semester we have included a point system 'Activity Points' that has given randomly selected students points for certain actions on Redmine. Please answer the following questions regarding this feature:

Q: Have you noticed the Redmine GipHouse point system?

- Yes:
 - Have you compared your points to those of others?
 - To what extent do you know which activities earn you points? (1--5)
 - To what extent do you know how many points you get for activities? (1--5)
 - To what extent did you perform certain actions on Redmine in order to increase your points? (1--5)
 - To what extent have you enjoyed receiving points on Redmine? (1--5)
 - Do you have any remarks with regards to the activity points on Redmine? (open)
- No

6.1.6 Closing

Q: Do you have any general remarks with regards to this questionnaire or the [subject]? (open)

6.2 Points given for actions

Activity	Points earned
Viewing the welcome page	20
Logging in	20
Viewing your project's main page	8
Looking at the list of issues	5
Creating an issue	10
Closing an issue	25
Updating an issue	20
Downloading a document	12
Creating a wiki page	60
Updating a wiki page	30
Reading a wiki page	10
Logging hours	30
Looking at the news page	10
Read new news on the first day	40
Read new news on the second day	35
Read new news on the third day	30
Read new news on the fourth day	25
Read new news on the fifth day	20
Read new news on the sixth day	15
Read new news on the seventh day	10