Universität Bremen*

The Gamification Inventory

An Instrument for the Qualitative Evaluation of Gamification and its Application to Learning Management Systems

Jan Broer

October 25, 2017

A thesis submitted for the degree of Doctor of Engineering (Dr.-Ing.) to the faculty of Mathematics and Computer Science at the University of Bremen.

1st supervisor: Prof. Dr. Andreas Breiter
2nd supervisor: Prof. Dr. Thomas Köhler

*EXCELLENT.

Abstract

Gamification has risen meteorically in popularity since the beginning of the decade, both in practitioner circles and among researchers. We show that empirical results of gamification's effects do not match the hype around it as studies have largely failed to prove any effects. We posit that a proper evaluation of gamification requires an understanding of how gamification can be expressed in real-world applications and employ Wittgensteinian family resemblances as a basis for such a definition. We have collected a set of properties that gamified applications can have through the analysis of goals and means of gamification mentioned in the literature and through an expert survey. We then used those results to create the Gamification Inventory, an instrument for the qualitative assessment of gamification in a given system. We have tested the instrument with a set of evaluators in the field of learning management systems (LMSs), informing both a refinement of the instrument and the preparation of an experiment with the intent of testing the effectiveness of common forms of gamification.

The analysis of these LMSs led to results very similar to what our analysis of previous empirical studies in gamification, and especially gamification in education, have shown: most gamification is concentrated on using points, badges, levels and leaderboards as game design elements. We argue that a large-scale, long-term experiment with a proper factorial design is needed to evaluate the effectiveness of gamification and have prepared such a study. Having identified points and badges as two major elements to be tested, we developed an extension to a competency grid add-on for the LMS Moodle that allows for a 2x2 factorial design of using points and badges. The system is designed for large-scale distribution among schools using the competency grid in Moodle, with minimal invasiveness in mind. We briefly discuss the challenges that come with such large-scale experiments, especially in German schools.

As a result, we present a new, tested, and refined instrument for the qualitative assessment of gamification in a given system, an overview over gamification as it is being used in the most popular LMSs, and an experimental setup to test the effectiveness of points and badges in schools, using custom add-ons to the competency grid for Moodle and to the corresponding mobile application.

Zusammenfassung

Seit Anfang des Jahrzehnts sind Bekanntheit und Nutzung von Gamification in Anwender- und Forscherkreisen stark angestiegen. Es wird aufgezeigt, dass empirische Studien zur Effektivität von Gamification diesen Hype bisher nicht unterstützen. Um diese messen zu können, ist ein Verständnis der zu Grunde liegenden Mechaniken notwendig. Auf den folgenden Seiten wird ein Ansatz beschrieben, in dem Familienähnlichkeiten nach Wittgenstein die Basis für die Definition und Analyse von Gamification bilden. Über eine Kombination von Literaturanalyse, mit einem Fokus auf die Ziele, die mit Gamification verfolgt werden, und die Mittel, mit denen Systeme gamifiziert werden, sowie von Expertenmeinungen konnte eine Liste von Eigenschaften generiert werden, die indikativ für die Mitgliedschaft in der Familie gamifizierter Anwendungen sind. Aus dieser Liste wurde das Gamification Inventory erstellt, ein qualitatives Instrument für die Evaluation von Gamification in einem System. Das Instrument wurde von Evaluatoren an Lernmanagementsystemen (LMS) getestet und damit sowohl das Inventory weiterentwickelt als auch eine Wissensbasis für ein Experimentaldesign geschaffen, mit dem die Effektivität von Gamification getestet werden kann. Die Ergebnisse der Analyse dieser LMS sind denen von existierenden Studien zum Thema sehr ähnlich: in den meisten Fällen ist Gamification auf Punkte, Abzeichen, Level, und Ranglisten limitiert.

Es wird argumentiert, dass groß angelegte Experimente über einen längeren Zeitraum mit faktoriellen Designs notwendig sind, um die Effektivität von Gamification zu messen; daraufhin wird das Design eines solchen Experiments beschrieben, mit Punkten und Abzeichen als experimentellen Faktoren. Das LMS Moodle mit einem speziellen Kompetenzraster wurde entsprechend erweitert. Das resultierende System ist so gestaltet, dass ein großflächiger Einsatz an verschiedenen Schulen möglich ist, die Moodle mit dem Kompetenzraster verwenden, ohne dabei in den Schulalltag einzugreifen. Kurz wird die Herausforderung diskutiert, groß angelegte Experimente durchzuführen, insbesondere an deutschen Schulen.

Als Ergebnisse der Arbeit werden ein neues, getestetes Instrument zur qualitativen Evaluierung von Gamification, eine Übersicht über den Einsatz von Gamification in LMS, und ein Experimentaldesign inklusive Software zur Messung der Effektivität vom Einsatz von Punkten und Abzeichen in Schulen präsentiert.

Acknowledgments

I would like to thank my supervisor Prof. Dr. Andreas Breiter for enabling me to follow my own research question and for guiding and supporting me in those years. Without his support and his contacts, this thesis would not have been possible.

I would further like to thank Michael Stütz and the other members of the Konzeptgruppe Moodle as well as Martin Kattner and GTN Solutions for giving me the opportunity to expand the competency grid employed in Baden-Württemberg with my own experimental features and supporting me whenever I had questions.

My colleagues at the Research Group Information Management (AGIM) and the Institute for Information Management (ifib) deserve all the thanks I can give for always being there for me when I needed them most.

Unfortunately, my thanks to the participating experts and evaluators have to remain anonymous. However, this research would not have been possible without their patience and active participation in forming the Gamification Inventory. To everyone who has participated: Thank you so much, you know who you are.

Finally, I would like to thank all my friends and family members who quite literally kept me alive all this time. Without you, there would never have been a thesis.

Contents

1 Introduction

The playing of games has mostly been seen as an activity of children in the past, with much focus in academic studies on training and learning (c.f. Grubbauer, 2011, p.21ff). Famously, Piaget (1975) studied the early development of children with a focus on play and Huizinga (2004) studied the element of play in culture, also focusing on the play of children and young animals. Caillois (1979, p.57), discussing the study of games, noted that “they have generally been regarded as simple and insignificant pastimes for children” but did not put a large focus on games among adults either. Playing adults were usually considered mostly in the realm of sport, with a focus on a “professionalization of attitude” (Blair, 1985). The past decades have seen a new form of games arrive, computer (or video) games,[1] and they have been treated similarly in the past. Fromme (2003) shows a focus on young people[2], discussing a move of video games into mainstream culture in the early nineties. Today, we see video games penetrate all age groups, with only 14% of gamers in the USA[3] being underage (Nielsen, 2015). Similarly, the move to a mainstream phenomenon is seen in studies that report that nearly half (49%) of American adults play video games (M. Duggan, 2015, p.2). Contrary to popular perception, this encompasses both men and women (M. Duggan, 2015, p.2). The two important trends to observe here are the move towards popular use of games (especially video games) on one hand and the move towards increased use among adults on the other.

Educators and researchers have picked up on the phenomenon of video games relatively early and have tried to incorporate them into learning, although usually with a focus on

[1] We will consider these two terms to be synonymous for the purpose of this research. We do not use them to distinguish between different gaming hardware.

[2] Note, that although he uses the term children in the title of the paper, Fromme speaks of young people in his text, already showing a gradual increase of the age of people who play video games.

[3] Other countries show similar trends (e.g. Germany Bundesverband Interaktive Unterhaltungssoftware, 2016), but the data available for the USA is most complete.

children. Ever since Abt (1987) wrote about the idea in his seminal work, researchers have examined various ways of using video games in education and there is no shortage of commercial products attempting the same.[4] Abt's work marks the beginning of a trend[5] of taking video games seriously. While the evidence for the usefulness of games for learning is still weak (c.f. Mayer, 2014), games are no longer seen merely as a pastime but as something that can be useful.

An interesting question that can arise when observing video gamers is that of their source of motivation. People who play computer games[6] often spend a very large portion of their time doing so. A particularly impressive figure that has often been quoted as a motivator for games research is that players of the massively multi-player online roleplaying game (MMORPG) World of Warcraft (WoW) had collectively spent about 5.9 million years playing the game (Hotz, 2012). While obviously correlated to the high number of users, this equals a per capita use of more than half a year full-time, which is about as much as three years of a full-time job. At the same time, these kinds of games often display many properties usually attributed to work, rather than play,[7] culminating in EVE Online, a massively multiplayer space game that is arguably more about spreadsheets than spaceships. And yet, even though these games can often resemble work, players are drawn to them without any physical compensation, they are even willing to pay monthly fees to play them.[8] Bogost (2007, p.ix) refers to the unique persuasive power of video games, based not in their content but in what he termed *procedural rhetoric*, highlighting that games are different from traditional media in their focus on processes and models.

The serious games movement discussed above (and the researchers that came before it) looks at how children (both human and animal) learn through play and tries to use video

[4]We discuss the various forms of serious games further in subsection 2.2.1.

[5]You could argue that this trend began earlier with Huinzinga, Caillois, or Piaget. Abt differs from these authors in two ways: for one, video games were of no concern for the earlier authors; for another, the work following Abt is less focused on understanding play as an element of culture and childhood learning but as a tool for various purposes.

[6]We will henceforth refer to people who play computer games as gamers, even though only few actually identify with the term (M. Duggan, 2015, p.2).

[7]Some games were even described as "obligation, tedium, and more like a second job than entertainment" (Yee, 2006)

[8]Recently, many online games have moved to a so-called free-to-play model. Such games are free to acquire and play but usually include additional features that players can pay for (c.f. Alha et al., 2014).

games for a similar purpose. Looking at the willingness of large parts of the population to spend their time playing video games suggests another aspect however: What is it that draws people to these games and can we use it elsewhere? This is the core idea of the relatively new concept of gamification. Deterding et al. (2011) define it as “the use of game design elements in a non-game context”, meaning that one takes something — a game design element — out of games and puts it to use elsewhere. The hypothesis being that there are individual elements of games or combinations thereof that make games so attractive to their players and that these could be employed beneficially outside a game context, e.g. to increase learner motivation or medication adherence. This is the basic concept at the core at our research.

As our analysis of various definitions of gamification in section 2.2 and of the empirical work in chapter 3 shows, implementations of gamification have mostly relied on very simple game elements in the past, with a focus on extrinsic rewards. Users are given points, badges, and levels, or are shown their standing on a leaderboard, usually in attempt to increase their usage of a given system. Much as in games for learning, empirical evidence for this approach actually being effective is scarce. We have re-reviewed major meta studies as well as additional literature not covered therein, both for gamification as such and for its application to educational contexts in chapter 3 and pointed out the lack of empirical evidence for success in (Broer, 2014). Therein we refer to Gartner's popular hype cycle for emerging technologies (Gartner, 2013) and point out that gamification has likely not yet reached its plateau of productivity but is rather in its trough of disillusionment.[9] One reason for the lack of measured success may be the focus on very few game elements that are not even core to most game experiences (see e.g. Barik et al., 2016), another could lie in the experimental setup of most studies.

Figure 1.1 illustrates the hype about gamification. It shows that a Google Scholar search for publications including the term gamification yields an ever increasing amount of hits for consecutive years, showing approximately quadratic growth and only now tapering off slightly for 2016. While these numbers clearly do not reflect the numbers of actual, peer

[9] The Hype Cycle postulates that emerging technologies follow a similar curve in the attention paid to them over time: A period of overinflated expectations, followed by a sobering period of dissapointment and an eventual rise to medium levels of attention when the technology matures.

reviewed publications on the topic, they illustrate the trend nicely. Similarly, we have seen a number of non-academic books published on the topic, (e.g. K. Duggan & Shoup, 2013; Zichermann & Cunningham, 2011; Zichermann & Linder, 2013; Werbach & Hunter, 2015; Kapp, 2012), and seen industry conferences become very successful. The Gamification World conference, for example, claims to have more than 1000 attendees (Gamification World S.L, 2016). At the same time, gamification projects often failed to meet expectations, with Gartner predicting in 2012 that by 2014, 80 percent of gamified applications would fail (Gartner Inc., 2012).

One observation one can make is that there is a rather large gap between practitioners and researchers in gamification, beginning with the definition of the term and the success of the concept. Whereas practitioners seem to prefer a loose, encompassing definition, researchers tend towards hard and fast rules (see section 2.2). Neither side seems to be able to agree on a definition, however. This is very similar to game studies, where no consensus for the term game can be found. Every attempt at a synthesized definition (e.g. Juul, 2003) leads to another definition to choose from. We discuss these definitions in section 2.1, both because an understanding of games is integral to an understanding of gamification and because their history provides us with valuable lessons for our attempt at defining gamification. We conclude our discussion of game definitions with an introduction to Arjoranta's idea of using Wittgensteinian family resemblances (c.f. Arjoranta, 2014) instead of a nominal definition. Arjoranta embraces the idea that games are not all the same, but rather part of a family in which members share certain properties but where these properties do not have to be common to all members of the family. This idea lies at the core of this thesis: we use the concept of family resemblances in gamification as the foundation for analysis of existing gamification — both in its implementation and in its effect. We posit that an understanding of the properties of gamification is essential for both effective research and effective implementation of gamification and provide a step towards such an understanding with our work.

The second big disagreement, the success of gamification, can for example be seen in Zichermann's claims that gamification has been proven to work (Zichermann, 2014); a claim

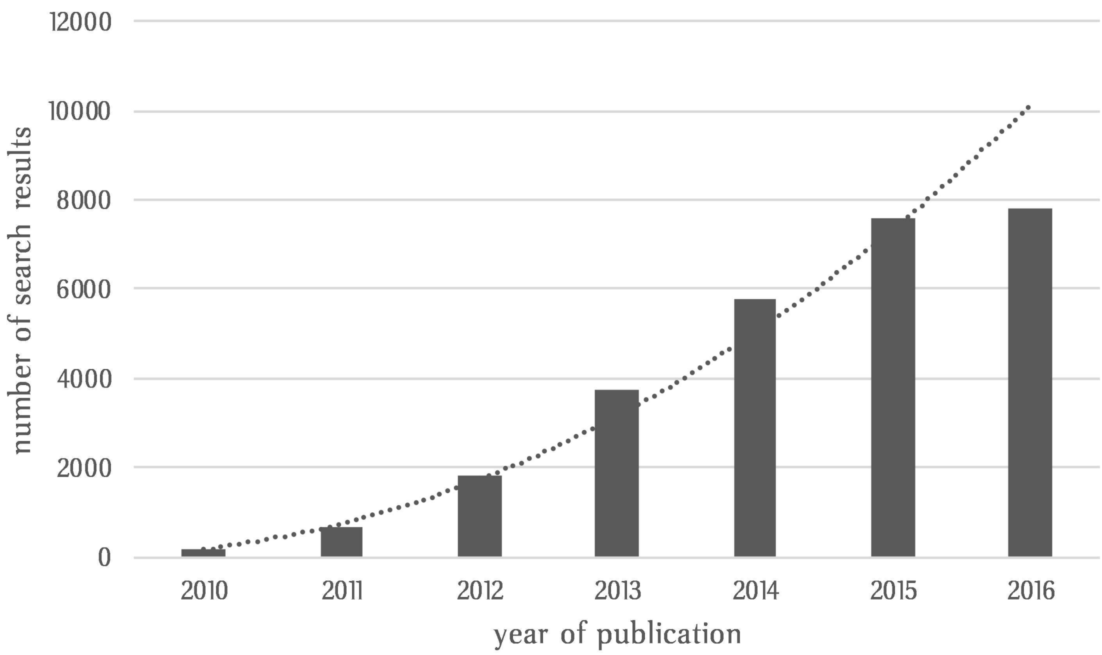


Figure 1.1: Hits for the term "gamification" on Google Scholar, filtered by year of publication, as of 19.01.2017. The dotted line shows an approximated quadratic growth trend (technically a power function, $y = 177.09x^{2.0794}$, $R^2 = 0.9899$). Own illustration.

that has yet to be properly validated by science (see chapter 3).[10] As we will show, scientific studies are often small-scale and in many cases not structured enough to make solid claims about the effectiveness of gamification. In the case of structured, large-scale approaches, reports of significant differences are rare. On one side of the conflict we have practitioners claiming the success of gamification, on the other side we have scientists who either have not yet been able to find proof for success (see e.g. Hamari, 2013) or, if they claim it, have study designs that are very hard to extrapolate from (see e.g. W. Li et al., 2012). In order for science to be able to support, or refute, the reported effects, structured research is needed. Such research begins with a solid understanding of gamification, as we can hardly compare results from studies that do not deal with the same subject.[11] As our initial research shows, the field of gamification is also too broad to be tested for as a single factor; the effects of badges may be very different from those of a leaderboard, for example. Even if we employ

[10] K. Duggan & Shoup (2013, p.14) make a similar claim.

[11] (Eickhoff et al., 2012) is an example of a paper included in the meta analysis by Hamari et al. (2014) that does not actually constitute gamification according to the stricter academic definitions.

the stricter scientific definitions of the term, there is still a broad range of game design elements that could potentially be used for gamification and a smaller, but still relevant, number of actually used elements. With a solid understanding of the core properties of gamified applications, we can proceed to design experiments for the proper measurement of the effectiveness of the various approaches. Here, we use the insight gained from the analysis of existing systems to develop an experimental setup and to make suggestions for measurement.

1.1 Gamification in Human Computer Interaction

As the examples in chapter 3 show, gamification is almost always discussed in the context of digital media. Either digital systems themselves are gamified (e.g. a social networking platform in (Farzan et al., 2008)) or digital systems are used to gamify an analog activity (e.g. a university course in (Akpolat & Slany, 2014)). None of the definitions we studied (see section 2.2) actually require the use of digital media, however. Deterding et al. (2011), for example, only specify a "non-game context". One can identify multiple reasons for the focus on digital media in practice. For one, computers make it much easier to track user behavior and to ensure that rules are followed, a human would have to do a lot of record keeping even for simple gamification. Furthermore, a computer is more or less impartial and many non-game context that one might want to gamify do not really support of a human judge. If we look at gamified fitness, for example (see Lister et al., 2014), it would be rather impractical to have a human observer with you whenever you train. The prevalence of digital media in discussions of gamification might lead one to make that part of a definition or at least to distinguish between digital gamification and analog gamification — just as Prensky (2003) considers it important to include the term *digital* in *digital game-based learning*. Similarly, most commonly used gamification elements (e.g. badges) are mainly used in digital games, suggesting that gamification is the transfer of elements from digital games to digital systems.

Arjoranta (2014) discusses this distinction in the context of games and points out that excluding non-digital games may lead to a loss of valuable insights. He quotes Aarseth on the topic, who warns that new technology can inflate our view of differences between media that are actually very similar to each other (c.f. Aarseth, 1997, p.14). We cannot see a reason for a fundamental distinction in either the source of gamification elements or the implementation between digital and non-digital systems. While the choice of medium can certainly be impactful and some forms of gamification are hardly feasible without computer assistance, the underlying principles remain the same. If a teacher hands out badges in class, for example, it should be considered gamification whether he or she does so through a digital system or in physical form. Similarly, lessons learned from non-digital games can still be very useful in gamification. In fact, our research on gamification in education has shown that analog versions of gamification have been used there for much longer than the term exists (see Broer, 2015). This is not to say that the choice of medium makes no difference. Möll (2014, p.151), for example, discusses the mediatization of poker and names the ability to observe your opponents as having an important impact on how the game is played. By removing face-to-face confrontation, the nature of the game changes. Höflich (2003) discusses the similarity of traditional mail and email in structure as opposed to the different practices of use. He writes that email and mail are not technically the same, but are very similar and should both be considered mail-like communication[12] (Höflich, 2003, p.42). Schönberger (2003, p.136) goes into detail about such practices, e.g. the difference in writing styles between an email and a letter.

Gamification deals with an area of study termed user experience (UX), defined in ISO standard 9241–110:2010 as “a person’s perceptions and responses that result from the use and/or anticipated use of a product, system or service” (International Standardization Organization, 2010, part 2.15). This can be seen in the goals of gamification that we extracted in section 4.1, where engagement and motivation feature heavily and where some authors consider the creation of a gameful experience at the center of gamification efforts (see e.g. Huotari & Hamari, 2016). When dealing with digital systems, user experience is seen as part of the field of Human-Computer Interaction (HCI) (Law et al., 2009). As such, UX — and with it gamification — is by nature multidisciplinary, drawing from computer science as well as social and behavioral sciences (see Carrol, 2002, p.xxvii). An understanding

[12]This is a slightly incorrect translation from the original German text in which he calls email a form of “briefliche Kommunikation” (Höflich, 2003, p.42). Another possible translation would be postal communication.

of digital media is essential for our work, but gamification deals with the experience of the users of the system and their resulting motivation. Motivation classically falls into the domain of psychology, but an increasing focus on the user in both HCI and information systems (see, for example, our analysis of gamification in information systems (Broer & Poeppelbuss, 2013)) has made interdisciplinary work essential. This is also evidenced in the variety of outlets that gamification research has been published in. Our research is based in the computer science area of HCI, but an understanding of the psychological factors that supposedly make gamification work is necessary. An analysis of the outlets of publications referencing gamification in Thomson Reuters' Web of Science (1429 hits for "gamification" as of 02.04.2017, see Table 1.1) shows the variety of areas in which it is discussed.[13] The top twenty sources are dominated by those that focus on education, but also include those that deal with the study of human-computer interaction, and that of games. The presence of the Games for Health Journal indicates that gamification is not limited to the field of education. Further down the list, interdisciplinarity becomes more apparent, including such sources as the International Journal of Market Research (5 publications), Frontiers in Psychology and Psychiatry (4 publications each), and the Journal of Nursing Regulation (3 publications), with a very long tail of sources with one gamification publication each.

In the study of gamification, this interdisciplinarity can be seen in the distinction between systemic and experiential perspectives, a key topic throughout our research. A designer of a gamified system is usually limited to changing that system itself — a systemic approach — but the desired outcomes are on the side of the user and therefore experiential and individual in nature. A practitioner designing a gamified system needs to understand the effects that individual design choices have on the experience of the user and the motivational power of those choices. Unlike other areas of HCI, such as the field of usability, we do not yet have enough information on these effects, especially when dealing with design elements extracted from games. If gamification is as effective at motivating and engaging

[13] An analysis like this is naturally skewed by special issues of journals and specific conference topics. An analysis by citation count shows similar interdisciplinarity, but a much lesser focus on education. That analysis has the issue of a high impact of individual publications, however, due to the low overall citation count. Also note that the Web of Science used here indexes far fewer sources than Google Scholar, which was used in figure 1.1.

Table 1.1: The 20 outlets with the highest amount of publications containing the term gamification, according to Thomson Reuters' Web of Science as of 03.04.2017. Results for conference series have been combined.

Outlet	Hits
European Conference on Games Based Learning	74
Int. Conf. on Education and New Learning Technologies (EDULEARN)	40
Computers in Human Behavior	31
Int. Conf. of Education, Research, and Innovation (ICERI)	25
Hawaii Int. Conf. on System Sciences (HICSS)	22
International Journal of Engineering Education	16
IEEE Frontiers in Education Conference (FIE)	15
Games and Learning Alliance (GALA)	14
IEEE Int. Conf. on Games and Virtual Worlds for Serious Applications (VS-GAMES)	14
Annual Symposium on Computer-Human Interaction in Play (CHI PLAY)	11
Computers & Education	10
Games for Health Journal	10
JMIR Serious Games	9
Serious Games (JCSG)	8
Int. Conf. on Interactive Mobile Communication Technologies and Learning (IMCL)	8
Multimedia Tools and Applications	7
IEEE/ACM Int. Conf. on Utility and Cloud Computing (UCC)	7
Int. Conf. on Applied Human Factors and Ergonomics (AHFE)	6
Elearning Vision 2020!	6
Journal of E-Learning and Knowledge Society	6

users as practitioners claim, it is bound to become integral to digital media design — yet no heuristics or guidelines exist. The goal of this research is to help to close this gap, by improving our understanding of those effects and providing tools to go even further in that direction. If we want to bridge the gap between practitioners and researchers, we need to show that gamification actually has the positive effects that it is said to have, such as increased participation and engagement or even improved learning outcomes. This thesis deals with the question of how to approach this question, beginning with an understanding of what gamification is and is not and ending in an experimental design that can be used to test the effectiveness of commonly used gamification elements. Section 1.3 contains a summary of our approach, details can be found in chapter 4 and chapter 5.

1.2 Education as an Application Context for Gamification

Education is one field that gamification is much discussed and applied in, as evidenced by the high number of publications about gamification in scientific outlets in the field (see table 1.1) and discussed in detail in section 3.2. It is also a field in which games have traditionally been of interest (see e.g. Gee, 2008)). Commercial products often focus on gamifying the classroom, such as Classcraft[14] or ClassDojo[15]. While these products are directly designed to add a virtual game layer on top of the reality of the classroom proceedings, existing educational applications are also introducing gamification. Learning management systems (LMS) — "web-based systems [that] allow instructors and students to share instructional materials, make class announcements, submit and return course assignments, and communicate with each other online" (Lonn & Teasley, 2009), for example, have been in use in both K–12 and higher education for many years now, with varying degrees of adoption. Many LMS are either described by their authors (e.g. Laskaris, 2015) or externally (e.g. Capterra, 2015) as being gamified. As in other contexts, these practical applications lack a theoretical foundation as well as proper (public) evaluation. As our analysis of gamification in LMS will show, these applications tend to focus on very basic features of gamification that have yet to be shown to be effective. This makes the field of education a very attractive area for further gamification research. For one, there is a relatively large body of work — both academic and practical — the analysis of which can help further our understanding of gamification. For another, the large amount of interest in both games and gamification in education indicates that their application is seen as promising by a variety of subject experts — potentially rooted in the somewhat antiquated view that games are for children. Finally, proper evaluation of the effectiveness of gamification requires a large number of users to test with, which education can provide. In Germany alone, about 8.4 million students were enrolled in schools of general education (Statista, 2016b) and an additional 2.5 million in vocational schools (Statista, 2016c). Unfortunately for HCI researchers, not all of these students use the same software systems in their education. The final decision over the software that is employed in a class often comes down to the teacher and even among

[14] www.classcraft.com
[15] www.classdojo.com

individual teachers that choice may differ depending on the subject matter taught. Further differentiation can be found on the school, state, and country levels respectively. As far as homogeneity of use goes, LMSs are among the more promising types of software used in education. For one, they are a rather mature technology. Browne et al. (2006, p.179), for example, already reported an adoption of virtual learning environments (their term for LMS) in 95% of UK higher education institutions. For another, the decision for a specific LMS is no longer mainly in the domain of individual teachers, but rather schools, or even states. These days, you will be hard pressed to find a higher education institution without a central LMS — e.g. all German universities with more than 10,000 students use one (Braungardt, 2014). A set of case studies comparing Germany to other countries showed a similar development for the use of LMS in schools in 2008 (Welling et al., 2008). They showed that other countries were already using centralized LMS solutions for K–12 education. Germany has followed suit since, spearheaded by federal state Baden-Württemberg, using the LMS Moodle (Breiter & Welling, 2008, p.33). Federal state Bremen is currently introducing itslearning as a centralized LMS solution (Landesinstitut für Schule Bremen, n.d.).[16]

While the base functionality of LMSs has been stable for many years now, new advances and new threats are arising constantly. Stantchev et al. (2014), for example, found that students prefer to use cloud file hosting services instead of LMSs for their cooperative work. García-Peñalvo et al. (2011) discuss options for combining LMS — usually made for formal learning — and personal learning environments (PLEs) (c.f. Attwell, 2007), usually made for informal learning. While institutions can enforce the use of LMSs, e.g. by requiring homework to be uploaded there,[17] ideally students would use LMSs because they help them learn. Usability and instructional design are two major factors in student adoption.

LMSs are particularly interesting for our research, because they have recently begun adopting gamification. Academic research on the topic is rare and no reliable numbers are available, but more and more developers at least claim that their LMSs are gamified. Andriotis (2014) describes the results of a survey on gamification in education done for

[16]It is of note that in Germany, federal states are independent in their management of education, among other things. (Bundeszentrale für Politische Bildung, 2015)

[17]And indeed, (perceived) mandatoriness can overshadow other factors for adoption (see e.g. McGill & Klobas, 2009).

TalentLMS, reporting a very high interest in gamified features. Unfortunately, no details about the study are known, making it of questionable usefulness. It does show the interest of LMS developers in the topic, however. Medved (2015) compared six gamified LMS for Capterra and our own research shows that most LMS include some form of gamification these days, especially points and badges (see section 5.1). While gamification usually does not directly improve usability or instructional design,[18] it is associated with increased adoption and sold as such in practitioner circles. One can assume, that the developers of the aforementioned LMS are including gamification to increase student engagement with the platform. The recent uptake of gamification and the growing tendency for centralized systems (bringing with it access to a large number of students through a central point of access) make LMSs an excellent field for the study of gamification.

1.3 Research Approach

The goal of our research is to provide instruments for the analysis of gamification, as applied to the context of education, especially LMSs, and to add to the growing body of knowledge about gamification in general. The first step towards structured gamification research is an analysis of the types of gamification in existence and their properties. We used two instruments to extract a list of properties that define the family of gamified applications. Following an approach we introduced in (Broer & Poeppelbuss, 2013), we analyzed existing definitions for the goals and means of gamification described therein: the why and the how of gamification as understood by a selection of authors from the field, both practitioners and researchers (see section 4.1). Furthermore, we asked experts to rate a large number of terms extracted from literature about gamification as to their usefulness for describing gamification, both in general and in a gamified system of their choice. These two approaches yielded an overview over gamification as experts see it, but not necessarily about the forms of gamification that are actually being used. Such an analysis is not only necessary in order to inform research, but also for the development of gamified applications in any chosen field. In both cases one needs to understand how and to what effect gamification

[18] Rather, gamification is likely more effective if part of clever instructional design.

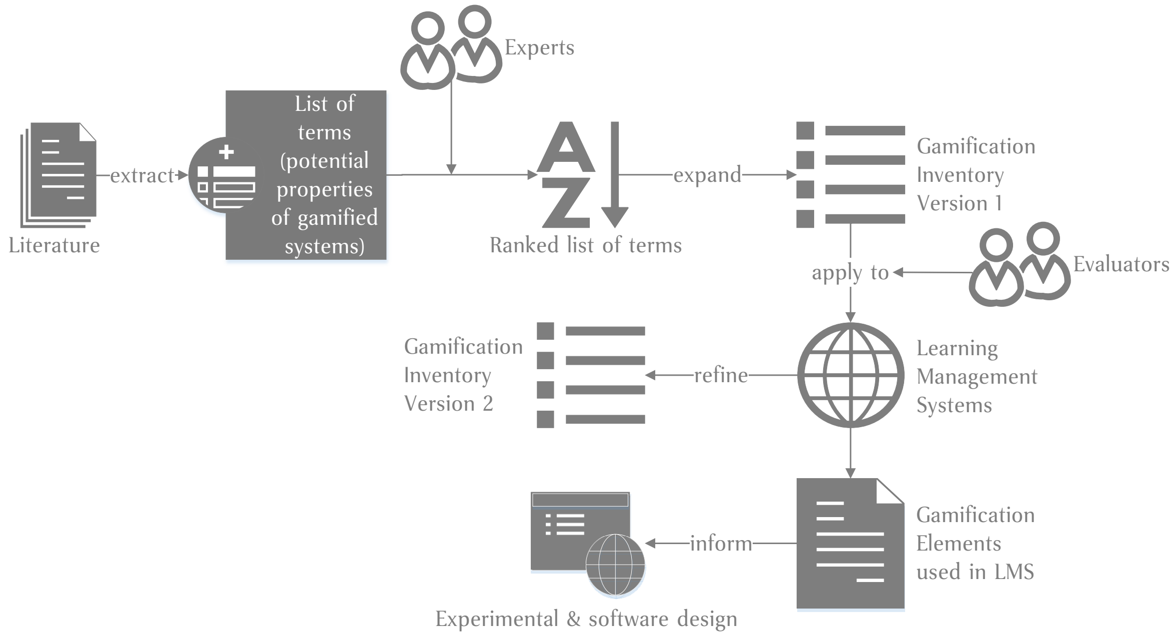


Figure 1.2: Structure of this research, resulting in an instrument for the qualitative evaluation of gamification, an experimental setup including software for testing gamification's effectiveness and an overview over the use of gamification in LMSs. Own illustration.

has been used previously in order to chose the correct elements for research or a newly designed system. Gamification being as broad as it is, a standardized instrument for such an analysis seems useful, both to guide researchers and to ensure compatibility between evaluations of different persons. If gamification was already well-defined and understood, such an instrument could be highly structured and quantitative in nature. As it is not, and as gamified applications vary greatly in their functionality and design, a more open, qualitative approach seems preferable. Similar to heuristic evaluations of usability (see e.g. Hollingsed & Novick, 2007), evaluators can be equipped with an instrument that guides their analysis, but will be required to adapt it to the system being analyzed and to employ their expertise for more detailed analysis.[19]

The set of properties indicative of membership in the gamification family we extracted from expert knowledge formed a basis for the development of such an instrument, the Gamification Inventory. Designed for the use of evaluators with basic knowledge about gamification, the initial instrument consisted of 38 items in five categories, reflecting the

[19]Heuristic evaluations have also attracted some interest in the field of game studies (see e.g. Nacke, 2009).

most commonly named terms in the previous evaluation. As the instrument was intended to be used for the evaluation of the use of gamification in a specific application or set of applications, testing and refinement required such a focus as well.

We applied our initial instrument to five LMSs with the double intent of testing the Gamification Inventory and discovering the gamified properties in use and updated it with our findings. With a firm understanding of the ways gamification can be used and is actually used in the field, one can begin to plan a proper quantitative evaluation of specific gamification elements. That includes an informed choice of these elements and their implementation — resulting in the requirements for a system with which to perform the experiment — and a proper experimental design that is suited to the context. In our analysis of existing quantitative research we have found a number of major issues that need to be addressed, among them the implementation of many changes at once (instead of a proper factorial design), the low number of participants — often caused by a lack of field access, and the short time frame. Especially the latter two issues are at least partially explained by the nature of academic research in small projects, such as a doctoral thesis. The time required to not only understand a chosen field but to actually create a gamified system and to then test it with a large number of users in the long term is simply higher than most such projects allow. We have designed such an experiment (see section 5.2) and implemented the software needed to complete it (see subsection 5.2.5), but have stopped short of actually performing the experiment. We discuss the issue of field access using the example of German schools in subsection 5.2.6. Our example implementation is set up for a 2x2 factorial design, testing the effectiveness of badges and points on student usage of a specific LMS feature, a competency grid for Moodle. As a result, we present a new, tested, and refined instrument for the qualitative assessment of gamification in a given system, an overview over gamification as it is being used in the most popular LMSs, requirements for an LMS to perform the experiment on, and an experimental setup to test the effectiveness of points and badges in schools, using custom add-ons to the competency grid for Moodle and to the corresponding mobile application. The following chapters present the theoretical and empirical underpinnings of our research, followed by two chapters interweaving methods and results to reflect the incremental process of our work.

2 Theory

A deep understanding of gamification from the point of view of HCI requires a multidisciplinary approach. Aside from theory on gamification itself, this chapter also deals with the — itself multidisciplinary — field of game studies as the foundation that gamification is built upon, and with the concept of motivation, from a psychological perspective. Elements of games are the building blocks that gamification is made of, as our analysis of a variety of definitions of gamification below shows. It is therefore not possible to understand gamification and its effects without first understanding games. In the first subsection below, we discuss a variety of definitions of games throughout the history of game studies. Two major points that shape our understanding are the characteristics of games — highly relevant to our attempt at using a family resemblances approach for the analysis of gamified systems — and the distinction between making something playful as opposed to making something gameful. Gamification, as we discuss it here, deals with the use of elements of games (ludus, in Caillois' words) and not those of free-form play (paidia).

Motivation, as the third major area of theory discussed here, is a goal common to most approaches to gamification, as we show later in our analysis of goals and means of gamification. Motivation is widely studied in psychology and it would be beyond the scope of this work to provide a deep analysis of the existing theories of motivation. In our work we have focused on self-determination theory (SDT) (Deci & Ryan, 1985b), as it is commonly referred to in works on gamification, as well as other areas of computer science, such as information systems.[1] We briefly present three additional theories of motivation that are more practical in nature, but also commonly referenced in works on gamification:

[1]We have shown in (Broer & Poeppelbuss, 2013) that information systems literature relies heavily on the same background in motivational theory that is used in gamification research.

Seligman's well-being theory, Pink's application of SDT to work contexts with a focus on autonomy, mastery, and purpose, and Fogg's behavior model. The latter three are rather useful in designing gamified systems, while SDT in its scientific accuracy is more useful for analysis. Especially the contrast between extrinsic and intrinsic motivation highlighted in SDT is at the core of much discussion of gamified systems. The common forms of gamification are extrinsic in nature (as both our analysis of the means of gamification and our related work section show), while the playing of games is usually intrinsically motivated.

Before we begin discussing the theoretical basis of this work, it is important to briefly discuss the inclusion of work by practitioners. Both games and gamification are fields that began in the world of practice and were only adapted as an object of scientific research later in their lifespan. Konzack (2007, p.110) marks a rise of video game research around the beginning of the century, a time that also saw the founding of the Digital Games Research Association (DiGRA) in 2003 and the first issue of the Game Studies journal in 2001. At that time, video games had been available for personal use for decades.[2] Even to this day, much of the body of literature in video games is based on personal experience and best practice, rather than empirical research (e.g. Salen & Zimmerman, 2004; Crawford, 1984). A similar phenomenon can be observed in gamification, although the history of academic research on gamification is almost as long as that of the term itself. Prominent books on the topic are all written with a practitioner's view (e.g. Zichermann & Cunningham, 2011; Zichermann & Linder, 2013; K. Duggan & Shoup, 2013; Kapp, 2012; Chou, 2015) although some of them include academic references. The two largest conferences in the field (Gsummit and Gamification World Congress) are run by practitioners, for practitioners as well. Research both on games and on gamification is therefore strongly rooted in a practical background and one would be remiss to discuss either without including that body of work. Unfortunately, practitioners often have little interest in the scientific method and their works are usually not peer reviewed. This is most evident in the question of gamification's effectiveness.

[2]It is difficult to decide on a date at which video games became mainstream. The first interactive computer game was developed in 1961, arcade versions of computer games began showing up in the early 1970s (including the famous Pong), and 1975 gave us the first game using a microprocessor. By 1980, companies were selling home versions of popular arcade games. (Kent, 2001, p.xiff) Loh et al. (2015, p.4) refer to the 1972 release of Pong as "the first landmark of digital games in modern history".

Many practitioners claim that it has been proven to work (see e.g Zichermann, 2014), which comes to no surprise since many of them make their living selling it. Unfortunately, as we will see in the review of empirical studies later on, science has yet to actually validate those claims. Any work dealing with games and especially gamification, including this one, therefore needs to find a way to combine the available knowledge from both sides, all the while keeping in mind the lack of empirical evidence for the practitioners' claims.

2.1 Definitions of Games

When studying definitions of gamification, one will often see references to definitions of the term game itself. Additionally, many attempts at creating lists of ways to gamify a system are founded upon existing definitions of what games are comprised of. As we will show in the following, it is far from easy to arrive at a definition for the term game that is both all-encompassing and narrow enough to still enable us to distinguish games from other concepts. Definitions have become more complex over time and more focused on structural elements of games as opposed to categorizations thereof. Historian Huizinga (2004) and philosopher Caillois (1979)[3] laid the foundation for our modern day understanding of games, but were not necessarily concerned with the issue of finding a perfect definition. Scholars have picked up on that work and attempted to provide more precise definitions. We will discuss a variety of definitions as they developed over time in this chapter in order to provide a basis for our further work. There are certain recurring elements in the definitions and authors like Juul (2005) and Salen & Zimmerman (2004) have created useful combinations of those. Nevertheless, some differences remain. Arjoranta (2014) has suggested a different approach to defining games that may help break the cycle of constantly redefining the term. In the context of this research, it is most important to understand games as part of gamification. It is not essential here to make exact distinctions between games and non-games. Rather, these definitions help us in understanding how to make a non-game context more gameful — to gamify it. Instead of finding a clear-cut definition of the term game, we are rather interested in a broad, but distinct set of properties of games that we can then

[3]Original publications in 1939 and 1958, respectively.

translate into gamification. Following Arjoranta's "Wittgensteinian" approach (Arjoranta, 2014), we are going to use definitions as "tools for practical purposes".

Most approaches to the definition of games discussed here are systemic in nature. They look for the elements (or properties) that in connection to each other define what games are. This is useful in various ways: For one, it allows us to look at a game as a set of rules which afford interaction with players. In this view, we do not need to discuss game artifacts (e.g. plastic figures or audiovisual content) or narative. Identifying elements of games suitable to gamification would then be as simple (relatively) as identifying the rules of a game and attempting to employ these elsewhere. For another, such systems can be analyzed objectively, as they do not take subjective experiences into account. Different players may interact differently with the same system and, in the case of more than one player interacting at the same time, the dynamic between players can greatly change their experiences. In either case, the system does not change. A game of Monopoly (if played correctly) is the same, no matter who is playing it or when, if looked at from a systemic perspective. Hunicke et al. (2004) discuss different perspectives of looking at games, referring to the "mechanics" of a game when discussing what we have been calling the systemic perspective here. The monopoly example from above also showcases the existence of these different perspectives ingrained in language: the term game in this context can refer to the physical artifacts that the game is played with ("I bought a new game today, it is called Monopoly."), the concept of the game as defined by its rules (Monopoly refers to all physical copies, no matter which artifacts are used therein.) and the process of playing the game ("Are you up for a game of Monopoly?"). While obviously related, the three uses of the term are sufficiently different to complicate definitions. The nature of the term game as part of natural language is at the heart of many issues with definitions. Wittgenstein (Wittgenstein, 2007) used the term to showcase his concept of family resemblances which we will refer to heavily in our search for a definition for gamification, and both Huizinga (2004) and Caillois (1979) refer to the use of the term game in various languages. German, the language of Wittgenstein's original work, even complicates things more with no distinction between the terms play and games. We will discuss the advantages and disadvantages of using a systemic perspective in our discussion of gamification in later chapters.

2.1.1 Structured and Unstructured Play

A key point of discussion with high relevance to our work in gamification is the question of free-form vs. structured play, of paidia vs ludus, a scale introduced by Caillois (1979). The more strict the rules of a game are, the closer to the ludus end of the scale it is. A game with no rules (if such a thing existed) would be considered paidia, paidia therefore being not the opposite but the absence of rules. This view becomes more complicated, when the same scale is used to distinguish between games and toys, between playful and gameful systems or experiences. In one of the foundations of scientific literature on gamification, Deterding et al. (2011) use a two-dimensional scale to situate gamification, one axis of which has *gaming* and *playing* as its poles (Deterding et al., 2011, p.13). The argument there is the same as Caillois': The gaming side is very structured, rule-based while the playing side is free-form. One can find rules as a pivotal characteristic of games in many definitions of the term. Salen & Zimmerman (2004, p.50), Mayer (2014, p.6), and Juul (2005, p.36) mention them directly, Crawford talks about games as "formal system[s]" (Crawford, 1984, pos.163) and Suits about "lusory means" (Suits, 2005, p.34).

In order to keep such definitions useful, it appears necessary to restrict our understanding of the term rules. A sandbox (be it an actual, real, sandbox, or a digital world in which players are free to do what they want) should rather be found on the paidia side of the scale. The player is provided with tools (toys) and is free to employ them to his or her heart's content. Such systems lack properties such as a quantifiable outcome (Juul, 2005, p.36) and they certainly have fewer rules than systems one would find on the other side of the scale. Yet they do have rules (laws of nature, program code, even certain outside restrictions — such as a mother forbidding her child to eat the sand in the box). Making the distinction of whether something constitutes a game or not simply based on the amount or strictness of rules available would not be very objective, as neither is actually measurable.

Unfortunately, Salen and Zimmerman's distinction between constituative, operational, and implicit rules (Salen & Zimmerman, 2004, p.130) does not really help with narrowing the definition of rules to something more useful to the ludus and paidia scale. A separate, unpublished study of ours[4] employed Kringiel's catalogue for game analysis (Kringiel, 2009) to compare two pairs of games, looking for factors that made them successful in electronic sports (Laudan, 2016). The study ran into difficulties identifying the rules of the examined games and concluded that computer games vary drastically from other games in this regard. A board game, for example, usually comes with a set of rules that describe exactly how the games is supposed to be played. This is necessary, because the game is run by humans and they would not be able to play the game as intended if they did not have access to all the rules. In computer games, rules are enforced by the computer and inseparable from the rest of the code that runs the game. For example, a board game's rules may state that players are not allowed to move their figurines onto water tiles. A computer game could have the same rule, but it is not explicit to the player anywhere, instead, the computer will simply prevent the player from moving their avatar into water.

Computers, as those responsible for preserving the correct game state, are capable of handling virtually infinitely complex sets of such rules and it could be argued that the whole program code constitutes the rules of the game. If all program code is part of rules, however, all digital systems have rules, making rules useless for distinguishing between games and non-games. Yet, the difference observed by Caillois clearly also exists in computer games, we even use the aforementioned analogy of a sandbox for a certain type of rule-less games. It seems imperative, then, to capture the essence of Caillois' distinction in a form that is suitable for analyzing video games. We will have to find the types of rules that move a game either to the ludus or the paidia side of the scale.

If we compare video games that are considered more free-form to those that are considered less so, two major differences can be observed. One is that of goals or objectives — a property of games discussed by Suits, and in a way also Salen and Zimmermann and Juul, when they speak of quantifiable outcomes. Paidia activities lack rules that define such goals or outcomes. Examples would be the child playing house or a player playing Minecraft (see e.g. Duncan, 2011). Both activities have rules, but none that relate to an outcome.[5] It

[4]The corresponding bachelor thesis is available upon request.

[5]Note that, as with many games, Minecraft offers a variety of modes for play, some of which have at least some outcomes — such as the death of the player character in hardcore mode. We refer to the game's creative mode here.

is important to note that a paidia activity can still be endowed with such rules regarding the outcome by the players themselves. Simply kicking around a ball with a friend would be considered paidia, but add two goals and a scoreboard and you have a much more ludic activity.

The second area in which free-form and structured games differ is in the limitations put onto the player. Many video games restrict the player's movement in order to control the story told in the game or restrict resources. Some video games go so far as to essentially restrict the player to a single corridor to move through. (This can have the form of an actual corridor in a game, but there are many other forms of restricting the player to a linear path.) In gamer jargon, this is often referred to as "railroading" (see e.g. Stenros et al., 2011). There is only one trail for the player to follow and nothing they do changes the outcome. Interestingly, following this trend to its final conclusion, one ends up in a narrative instead of a game. Putting no limitations on the player leaves us deep in paidia territory — and therefore we would be dealing with play, rather than a game — while adding too many restrictions reduces interactivity and creates narrative instead. Aarseth has described this phenomenon as a spectrum between a narrative pole and a ludic pole (Aarseth, 2012). We have illustrated these differences as two dimensions in figure 2.1. The horizontal dimension shows the amount of restrictions put on the player and is equivalent to Aarseth's spectrum. The ludic pole is on the left, the narrative pole on the right.[6] The vertical dimension depicts the role that goals play in the game. Games on the bottom have little to no goals for the player, while those on the top rely heavily on goals. The figure illustrates that Caillois' scale is a collapse of at least two dimensions. The classic examples of a free-form open world type game (bottom-left quadrant) and a very regulated, goal oriented game (upper, right-hand quadrant) are not the only possible games. One can have a very open world, for example, and still provide very detailed goals to the player.

[6]As depicted, games closer to Aarseth's ludic pole are actually closer to paidia than ludus. An increase in restrictive rules moves a game closer to Aarseth's narrative pole and towards the ludus side of Caillois' scale. This is a function of using the same term (ludus/ludic) for two different concepts. This also shows that even this two-dimensional depiction is a simplification. While highly narrative games tend to put more restrictions on the player, restrictions themselves do not necessarily cause narrative. Proponents of interactive storytelling attempt to create narrative without restricting the player. See (Crawford, 2005, p.49ff) for a discussion of plot versus interactivity.

Of the two types of rules that can increase ludus,[7] those adding goals seem to be more useful for our distinction, which is indubitably the reason why they appear in other definitions of games as separate concepts. Restrictions, on the other hand, seem dangerous in so far that too many of them can make an activity less of a game rather than more. Nevertheless, structure requires boundaries and Suits emphasizes the importance of less efficient means (lusory means).

Overall, our understanding of games, and therefore gamification, falls on the ludus side of Caillois' scale and on the upper two quadrants of our own two-dimensional representation. As we will see later, the narrative side of games is not recognized as important for gamification by a good number of experts, underlining Bogost's statement that we discussed above of procedural rhetoric being the key differentiator between games and other media. Also very important for our future work is the takeaway that the concept of rules is difficult in computer games and will therefore be hard to transfer to gamification. As we will see later on, different understandings of the term can lead to issues in structured analysis.

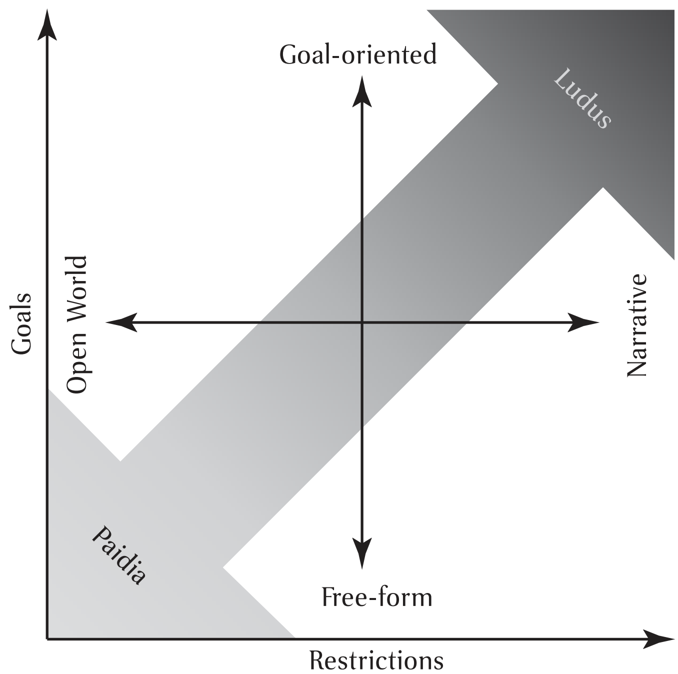


Figure 2.1: Paidia and ludus as a combination of at least two dimensions. Own illustration.

[7]Technically, there is no such thing as a measure of ludus. When we refer to an increase in ludus, we are actually discussing a shift on the ludus–paidia scale towards the ludic pole.

2.1.2 Characteristics of Games

Most modern definitions of games discussed here have in common that they describe necessary characteristics of games. Juul, for example, considers systems lacking one of his characteristics to be borderline games; those lacking more to not be games at all (Juul, 2005, p.44).[8] Crawford mentions a scale of "gaminess" (Crawford, 1984, pos.280)[9] on which one can only reach high marks by adhering to his set of characteristics. In a way, this is due to the attempt to create what Arjoranta calls a real definition — a definition that describes all games and only games, based on attributes shared by all games (Arjoranta, 2014). Finding such a definition can be very helpful to distinguish between different fields of study and to ensure that one is talking about the same subject when discussing games, but it is really not all that helpful for an understanding of gamification. If gamification, as we will see below, deals with elements of games only, one could assume that one could simply take the distinct properties identified by the definitions discussed here and use them individually. If the absence of a single one of those properties (let alone more) can turn a game into a non-game, however, we cannot assume that any of the positive effects we ascribe to games will be preserved in such a transfer.

Much more important than a list of hard and fast criteria would be a list of properties commonly found in games — but not necessarily all games. Games being as different as they are (which both Caillois and Wittgenstein emphasize), any property common to all of them is bound to be very general and essential to any game-like experience. Take Crawford's interaction, for example. If interaction is a measure of "gaminess" (Crawford, 1984, pos.280), a system without interaction is not "gamy" at all — no matter what other properties of games we add. No amount of rules, for example, can make a non-interactive system game-like. Taking a step back from real definitions allows us to employ more specific sets of elements of games, as they do not need to be found in every game. Deterding et al. (2011) refer to "elements characteristic for games", actively discouraging the use of elements

[8]This is a slight simplification. Some characteristics seem to weigh more than others. Free-form play, for example, only lacks one of his characteristics but is not considered borderline, while chance-based gambling lacks two characteristics and is still considered borderline (Juul, 2005, p.44).

[9]We are referring to the Kindle ebook version of the text here which does not have page numbers. We list the positions in the Kindle version instead.

common (all games have them) or unique (only games have them) to games. There are only a few attempts to catalog game elements. Mostly one can find short lists of important elements, such as the "ten ingredients of great games" by Reeves & Read (2009).

In a separate unpublished[10] study by Sarabi (2014), we attempted to use a catalog of questions for game analysis derived from Kringiel's work (Kringiel, 2009) to analyze three games of different popular genres in order to see if it was possible to obtain a comprehensive list of game design elements employed in those games. The result was a long list of elements at different levels of abstraction that could partially be mapped to gamified applications. The study also made clear, however, that not even the application of a very detailed, structured instrument was sufficient to extract all game design elements from a game. Further work in this area in combination with multiple experts performing the analysis might lead to a sufficiently complete list, however, that could then be categorized and adapted to a specific level of abstraction. One issue with such structured extraction is the diverse nature of games. Our studies that used Kringiel's catalogue of questions found that its applicability differed greatly depending on the subject studied, even though it was designed to be general in nature. Also, some questions are still not detailed enough and concepts not well defined. An extreme example here is the definition of what constitutes a rule, as discussed above.

Overall, the study of game definitions reveals a set of properties of games that one needs to understand when studying, or designing, gamification. Rules and goals play an important role in making something (more of) a game, as does interactivity and, with it, conflict. Suits also left us the idea of overcoming unnecessary obstacles as a key element of games, which is especially interesting in the context of gamification. Is the tension of gamefulness with the need to get real-life results a hindrance to gamification or can we maybe leverage existing obstacles by framing them as unnecessary? As important as this understanding of what makes games work is, these definitions are not suited as a basis for finding game elements to use in gamification. First steps in such a direction have been taken, but for now all work on gamification is based on a limited understanding of which parts of games we can actually use. Further work described below concentrates on identifying which elements are already used, how effective they are, and where room can be found for innovative approaches.

[10]The corresponding bachelor thesis is available upon request.

2.1.3 Game Definitions in Detail

The definition of the term game is a long discussed topic within game studies. We have addressed the major points relevant to this research above, but a good understanding of the concepts behind them requires a more in-depth analysis. This chapter introduces a set of definitions from literature in historical order; chosen for their prominence in game studies literature and uniqueness. There are many other variations on these (e.g. Abt, 1987; Myers, 2009; Tavinor, 2008; Waern, 2012; Frasca, 2007) but they do not differ in such a way that they would add substantially to the discussion. We will discuss Huinzinga's and Caillois' foundational work that predates video games and Suits philosophical treatment of the matter that did not take the (then) emerging phenomenon of video games into account. All following definitions build upon the classics but mainly concern themselves with video games. Video game pioneer Crawford discussed definitions in his 1984 book, as well as in later works. Both Juul and Salen and Zimmermann attempted to provide definitions as a synthesis of previous definitions such as the ones discussed here. Arjoranta refers back to Wittgenstein's discussion of the term game and his notion of family resemblances in an attempt to break away from real definitions of the term, providing the theoretical foundation on which most of our research is built. Finally, we discuss Mayer's definitions of learning environments as games, since our work is focused on a learning context. Game-based learning is not the focus of this work, however, and a full discussion thereof would not be helpful for understanding gamification. Mayer's work is included purely as an example of the application of classical game studies to the field of learning.

Huinzinga

Scholarly work on games predates video games by many years. While not the first scholar to tackle the concept of games, Huinzinga's "Homo Ludens" (Huizinga, 2004) is often used as a basis for discussing the definition of games in the field of game studies. Huinzinga focused on play as an essential element of culture, looking beyond the biological function of training infants for life's challenges. Huinzinga names a set of formal properties of games. For one, play is free and ordered play is no longer play (Huizinga, 2004, p.16). Later definitions use different wording, but this characteristic is common to many of them. The idea that play is voluntary and not something that we have to do is an important characteristic that should not be forgotten when attempting to use play in other contexts — such as gamification.

Secondly, play is not ordinary — or real — life. Huinzinga makes a clear distinction of play happening when we leave ordinary life and spend time in a separate temporary sphere (Huizinga, 2004, p.16). Being outside of real life, play is situated “outside of the process of satisfying needs and desires” (Huizinga, 2004, p.17, translated by the author). This externality, the strict separation of play and real life, is difficult to translate to gamification, as we will see later on. In a way, it is contrary to the whole concept of gamification.

Huinzinga’s third property of play is that of finiteness and limitedness. Play happens inside certain borders of time and space. All play has an end at a certain time and due to its temporal limitation is repeatable. Huinzinga underlines repeatability as one of the key properties of play (Huizinga, 2004, p.19).
Related to temporal and especially spatial limitations is another characteristic of play: order. Following Huizinga (2004, p.19), “play demands order”. Later definitions tend to use the term “rules” instead of referring to order, but the principle remains the same. Play (or, if we refer to Caillois’ distinction: games) without order is not play. Rules, or structure in general, are essential for play to take place. In his summary characterization of play, Huinzinga names the lack of material interest and profit as a key characteristic of play (Huizinga, 2004, p.22). This is connected to the separation of play and real life mentioned above, but more limited and easier to reconcile with the idea of gamification. Huinzinga provided many baseline concepts for the discussion of games, even if later definitions have modified his in most places. The direct use of Huinzinga’s work is complicated by the lack of a distinction between play and games that can so often be found in later work.

Caillois

Caillois’ work, like that of Huinzinga, predates video games. Caillois builds on Huinzinga’s work and attempts to categorize games. He identifies two important distinguishing factors that are cited in games research to this day. For one, he categorizes games into four types:

Games of competition (agon), games of chance (alea), games of mimicry, and games of vertigo (ilinx). Modern video games can still be analyzed along these categories (with some difficulties in the latter category), but usually do not clearly fall into one category or the other. Many games are games of competition but include an element of chance, for example. (Caillois, 1979, p.36)

Arguably more important to modern game studies and especially gamification is the difference between structured play, governed by rules (ludus) and open, free-form play (paidia). Ludus is closer to many common definition of games (rules are very often a requirement for something to be considered a game) while paidia is more akin to playing with toys. This distinction led to the creation of the term *gamefulness* as counterpart to *playfulness*. Some authors even go so far as to differentiate between *gamification* and *playification* (Scott, 2014, p.401). The distinction between ludus and paidia is the main advance that Caillois' work introduced over that of Huinzinga. Not all play is structured and governed by rules in the same degree. A child playing house operates under only the loosest of constraints. He or she will attempt to mimic the behavior of humans and toys will usually be limited to certain rules adapted from real life as part of that mimicry, but otherwise he or she will do as he or she pleases. Play that is more on the free-form side of the scale (paidia) is also much less finite in temporal terms. While it will come to an end eventually, the temporal limitations are usually dictated by factors outside the game — e.g. loss of interest on the side of the child, or dinner being served. Ludus, on the other hand, is usually limited by internal constraints, such as a win condition (e.g. He or she who reaches the goal first, wins, ending the game).

As one can see in Caillois' use of a scale from paidia to ludus, this distinction is not a binary one. Play is not always strictly free-form or governed by hard-and-fast rules. Most forms of play happen between the two poles. This uncertainty in definitions is inherited by later attempts of distinguishing between playfulness and gamefulness and plays a major role in our discussion of gamification and serious games in subsection 2.2.1. No matter which definition one adheres to, there are always examples of play that are on the border of being considered games. We have expanded upon the distinction between ludus and paidia above and introduced the idea that the scale is at least two-dimensional.

	AGÔN (Competition)	ALEA (Chance)	MIMICRY (Simulation)	ILINX (Vertigo)
PAIDIA Tumult Agitation Immoderate laughter	Racing, Wrestling, Etc. } not regulated Athletics	Counting-out rhymes Heads or tails	Children's initiations Games of illusion Tag, Arms Masks, Disguises	Children "whirling" Horseback riding Swinging Waltzing
Kite-flying Solitaire Patience Crossword puzzles LUDUS	Boxing, Billiards Fencing, Checkers Football, Chess Contests, Sports in general	Betting Roulette Simple, complex, and continuing lotteries*	Theater Spectacles in general	Volador Traveling carnivals Skiing Mountain climbing Tightrope walking

Figure 2.2: Caillois' categorization of games. Image taken from (Caillois, 1979, p.36)

Suits

Bernard Suits brings a very interesting perspective to the discussion of the definition of games in the concept of the lusory attitude. As they are for many other authors, rules are central to his definition of games. The important difference being his stress on the rules enforcing less efficient means (lusory means) as a way to arrive at a goal, and games being made possible by the acceptance of those rules by players in order to be able to play (aforementioned lusory attitude). Essentially he is saying that a game is only a game if we purposefully restrict ourselves to a limited set of means so that we can play a game. (Suits, 2005, p.34)

Juul later argues that Suits' idea of less efficient means does not fit many games and questions whether it applies to digital games at all (Suits, 2005, p.34). His argument has some merit, but it seems as if most, if not all examples could still be broken down to an attempt to overcome unnecessary obstacles and the lusory attitude in general. There are some rather famous examples making fun of gamification for it's shallowness, such as a button that simply gives the player a million points. Obviously such a game is not fun for anybody (and we could certainly question whether it is a game at all.) Suits' work does well at explaining why this is not fun, but moving around falling blocks in Tetris to achieve the same score can be: While both games may have the same goal of achieving a high score, Tetris gives the player much less efficient means to do so, putting unnecessary obstacles in the player's way. Not only do you have to jump through the hoop of moving falling blocks around instead of simply clicking a button, those blocks even come in different shapes, appear in different orientations, and start falling faster and faster the higher your score gets. If questioned, a player of Tetris might well state getting a high score as the goal of the game, yet choses to play that game instead of the much simpler game of the 1 million point button. This is Suits' lusory attitude at work — the player is playing the game because it puts those obstacles in his or her way and only because of those does it actually become a game.

One can question the usefulness of this aspect of Suits' definition for actually distinguishing games from non-games, but it is still an important characteristic to keep in mind. The concept is closely related to those of challenge, difficulty, conflict (Crawford, 1984, pos.280), and flow (Csíkszentmihályi, 2008) in that it stresses the importance of making games non-trivial to the player. When attempting to introduce game elements to non-game contexts, lusory means and lusory attitude provide an interesting tension to productivity. Especially in work environments, efficient means tend to be valued highly and the introduction of unnecessary obstacles would run counter to such efficiency.

Crawford

Crawford's definition of games is often reduced to four factors: representation, interaction, conflict, and safety (Juul, 2005, p.30). These terms contain many other elements found in later definitions as well. Crawford describes a game as a "closed formal system that subjectively represents a subset of reality" (Crawford, 1984, pos.163). That phrase contains many important concepts. A closed system is internally complete and does not need to reference the outside world. This is similar, if more technical, to Huinzinga's separation between play and the real world. The game functions in itself and does not interact with the real world. The term formal refers to the need for rules in a game, a requirement common to most of the definitions discussed here. Unlike his predecessors, Crawford chose to define a game as a system, a collection of interacting parts. We will further discuss the use of the term in the following section on Salen and Zimmerman's definition.

Finally, there is the subjective representation of a subset of reality that Crawford mentions. While the concept of a fictional game representing reality is difficult and not very important to our work here, it should be noted that Crawford differentiates between the subjective representation in games and the (attempt at) objective representation in simulations. While games, according to Crawford, represent reality in a way, they do not attempt to actually model reality. Deliberate simplification as opposed to accurate modeling is key to game design and therefore an important characteristic of games (Crawford, 1984, pos.185). Crawford (1984, pos.207) also stresses the importance of fantasy as "the key agent in making the game psychologically real".

The second key characteristic of games Crawford is usually quoted for is interaction. While not as ripe with additional hidden concepts as representation, interaction is equally difficult to tackle. In his original 1984 work, Crawford considers interactivity to be the degree in which a system responds to the player's input. He considers puzzles to be non-interactive, while games are. (Crawford, 1984, pos.214), In a later work, Crawford defined interaction as "a cyclical process in which two actors alternately listen, think, and speak" (Crawford, 2003, p.5). Interaction, Crawford suggests, "provides a useful index of 'gaminess"' (Crawford, 1984, pos.280).

The third element identified by Crawford is conflict. It arises from obstacles actively placed in the player's way (be it by another human or by a computer). Passive or static obstacles, on the other hand, can be found in non-games such as puzzles (Crawford, 1984, pos.280). Following the argument, "conflict can only be avoided by eliminating the active response to the player's actions" (Crawford, 1984, pos.287), meaning a removal of interactivity and therefore making the game a non-game. If conflict follows directly from interaction as Crawford states (and no interaction can be had without conflict), it might be possible that the two terms describe identical concepts. Crawford does not discuss this in further detail.

The final element of games mentioned here is safety. Unlike other authors, Crawford does not require absolute detachment from real-life consequences however. His definition simply states that the game must be safer than the reality it is representing (Crawford, 1984, pos.302). A game, following Crawford, might still have consequences beyond itself such as a loss of dignity or being rewarded (or failing to achieve said reward.)

Salen and Zimmermann

Salen and Zimmermann's definition of "A game [as] a system in which players engage in artificial conflict, defined by rules, that results in a quantifiable outcome." (Salen & Zimmerman, 2004, p.80) is an attempt at synthesis of previous definitions. As such, many elements of their definition have been discussed previously and are simply connected differently here. One key point they make is that games, to them, are systems — "sets of parts that relate to form a whole" (Salen & Zimmerman, 2004, p.50). This understanding of games being made up of elements is critical to many definitions of gamification. In fact, much of this work will deal with the question of identifying such elements and transferring them outside games.

Salen and Zimmerman maintain Huinzinga's separation of games from the real world, framing them as artificial in nature. Similarly, they adopt the necessity of rules for games to exist. Their definition of rules is rather broad, however. Rules of digital games include the program code itself, but are not limited to it. The authors also mention operational and implicit rules. If the program code itself is to be considered a set of rules, it seems impossible to create a digital game that does not include rules. With such a definition, all digital play would fall on the ludus side of Caillois' scale. Expanding this concept into non-digital games, one could even argue that the laws of nature limit any sort of free-form play, essentially making all play ludus. This results in an interesting addition to the understanding of the paidia–ludus scale. If rules are expanded to include what Salen and Zimmeman call the "constituative rules" (Salen & Zimmerman, 2004, p.149) of the game, no play can ever have absolutely no rules. The paidia end of Caillois' scale is therefore a theoretical point of absolute zero that can never be reached. Describing play as paidia, or free-form, therefore really only means that the game has relatively few rules, or less ludus.

Salen and Zimmerman's definition includes a point that is implicit in most other definitions, the need for at least one player interacting with the system. Interestingly, their definition is vague on whether they consider the player part of the system or not. Players are necessary for a game to be a game, but at the same time they are not actually part of the game (since they interact with it.) As a systemic definition, this is questionable. A game does not have players, but rather the potential for players to interact with it. Games afford playing (for more on affordances see section 4.1), but does a game cease to be a game when nobody is playing it? The latter would be in line with Huinzinga's limitation of play in time, but it seems problematic in a definition focusing on games as systems (rather than activities or experiences). The ambiguity of the terms play and game further complicates all attempts at definition. One would commonly consider a box of Monopoly a game (essentially a system of rules and artifacts affording play). At the same time, one would speak of playing a game of monopoly, referring to the activity of playing (using those rules and artifacts).

Salen and Zimmerman incorporate Crawford's conflict as central element of games. (Salen & Zimmerman, 2004, p.80) Conflict can exist between players, but also between a player and the game system. When playing the game, the players process is impeded by obstacles that he or she needs to overcome. For the purposes of this definition, it is not relevant whether these obstacles are created by the game system or by other players. Overcoming these obstacles (or failing to do so) leads to an eventual end of the activity, with — according to Salen and Zimmerman — a quantifiable outcome. This incorporates multiple previously named properties of games. For one, there is Huinzinga's limitation in time. Games, if defined this way, need to end. If there is no clearly defined end to a game, it is not a game. Secondly, this end does not just mark the termination of the activity of playing, but includes a quantifiable outcome. One has not just played a game, but there is a result summarizing how successful one was at playing. Interestingly, Salen and Zimmerman name this property — and not rules — as the one "usually distinguish[ing] a game from less formal play activities" (Salen & Zimmerman, 2004, p.80).

Juul

Juul's classic game model (see figure 2.3) is an attempt at a synthesis of many other definitions of games. He names six important characteristics of games: rules, variable and quantifiable outcome, valorization of outcome, player effort, player attached to outcome, and negotiable consequences (Juul, 2005, p.36). Rules reflect Huinzinga's demand for order and their existence clearly places games (as Juul sees them) on the ludus side of Caillois' scale. Following Juul, play that is not governed by rules (Caillois' paidia) cannot be considered a game. The need for outcomes to be quantifiable has also been stated by Salen & Zimmermann and even Caillois already required games to be uncertain. If the outcome was already predetermined, the game itself would lose its purpose. Outcomes, according to Juul, also need to have values attached to them and matter to the player emotionally. Not all outcomes can be of identical quality, some should rather be considered better by the player than others. In other words, players should be able to strive for certain outcomes. This striving is reflected in another of Juul's properties of games, the requirement for player effort. All these requirements for outcomes clearly point us in the direction of ludus instead of paidia once more. Quantifiable outcomes can only exist, if rules exist that define what an outcome is. Similarly, in play that does not adhere to rules, the player can essentially choose an outcome instead of striving for it. Play without rules therefore does not require player effort and is not challenging.

Juul's final requirement is that of negotiable consequences. Unlike Huizinga, Juul does not require a full separation from real life, but the potential for such separation. The classic example here would be a game of poker. Such a game can have meaningful real-life outcomes if played for money, but it does not need to. Poker, according to Juul, does not become less of a game because of the possibility of having consequences. The important factor is that these consequences are negotiable — that play can also be fully separated from real life, if desired. Following Juul, all full games fulfill these characteristics. He names borderline cases, i.e. Sim City, which fulfill most but not all characteristics and gives examples of non-games that lack more than one of them. It would be a mistake to use Juul's model as a basis for a list of game elements, as his intent was not to describe the characteristics a game can have, but rather all those that a game must have. Games can have many additional characteristics that are not described in Juul's model — these are simply not necessary for something to be considered a game. It is an interesting exercise to apply Juul's classic game model to gamified systems in order to see where such systems differ from games. As we have seen, Juul's model falls clearly on the side of ludus on Caillois' scale. In our own interpretation of the distinction seen in figure 2.1, Juul's games would be found in the upper two quadrants. It remains to be seen, if it is useful to concentrate purely on ludus when discussing gamification or whether free-form play can be useful as well.

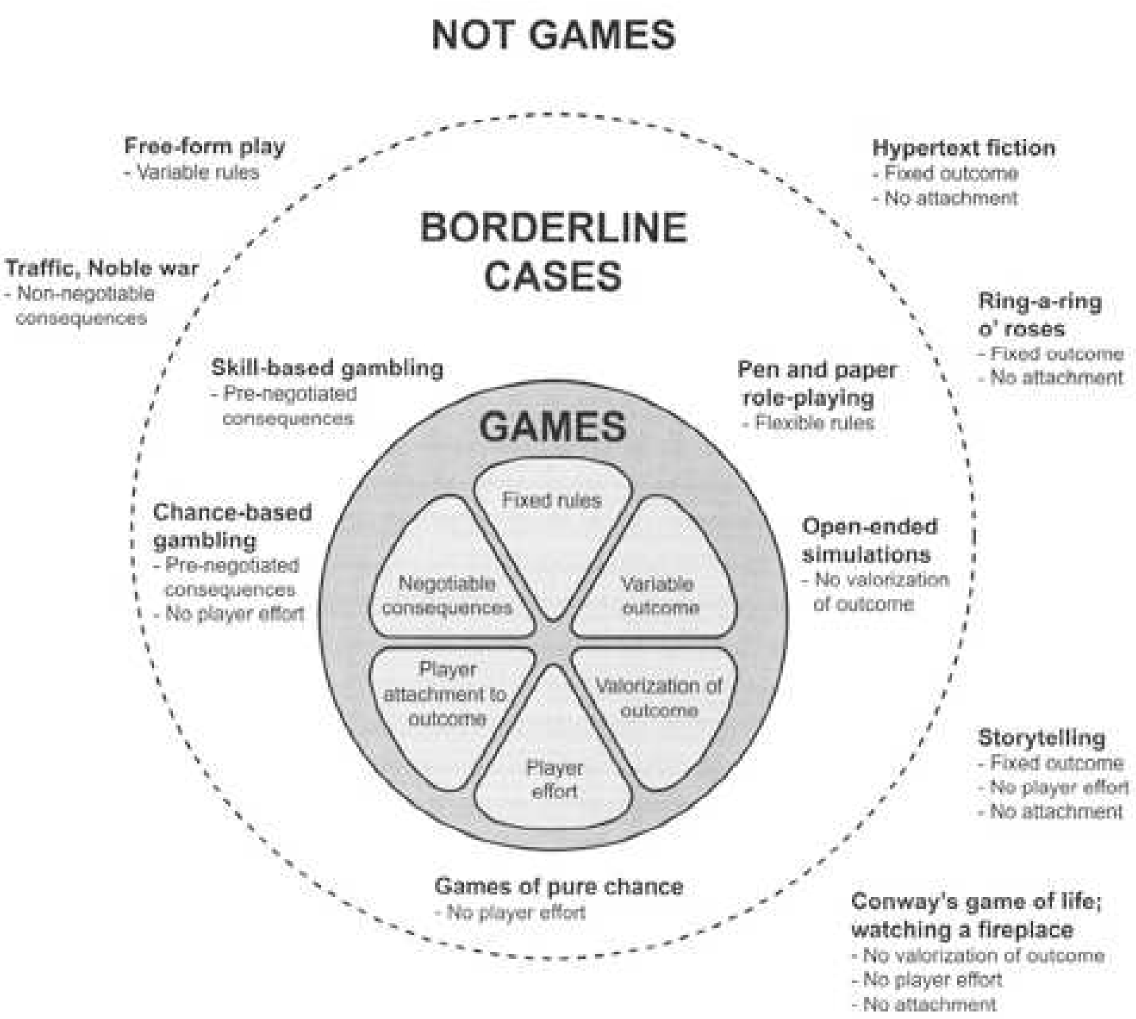


Figure 2.3: Juul's classic game model. Image taken from (Juul, 2005, p.44)

Arjoranta

Some authors discussing the definition of games like to bring Wittgenstein's Philosophische Untersuchungen into the discussion. Wittgenstein does indeed refer to commonalities between different types of games in his work (Wittgenstein, 2007, §66) but does so as an illustration of family resemblances rather than to actually discuss games as such. The idea behind family resemblances does fit the predicament of defining games well, however. Wittgenstein points out that there are cases (games being one) in which we use one term for a variety of things that are not actually defined by a characteristic (or set thereof) common to all of them, but rather a variety of features that are shared between some, but not all members of the family. Arjoranta refers back to Wittgenstein as a foundation of his argument against what he calls the *common core approach* (Arjoranta, 2014) of game definitions. The common core approach is exemplified by Salen and Zimmerman as well as Juul in the previous chapters. Both definitions attempt to combine relevant elements of older definitions into a new, more precise, all-encompassing definition. Arjoranta argues, with reference to Wittgenstein, that we should look at family resemblances instead, at "features that connect some, but not necessarily all, games." (Arjoranta, 2014) Instead of providing a list of properties that games must have, such a family resemblances approach would rather focus on a much larger list of properties that games can have. We have adopted this approach to the term gamification and describe it in more detail later on.

Mayer

Mayer (2014) names five defining characteristics for learning environments to be called games: Games are

- rule based, meaning that "events occur within a causal system based on a knowable set of rules" (Mayer, 2014, p.6),
- responsive, meaning that the "environment allows for players to act, and responds promptly and saliently" (Mayer, 2014, p.6),
- challenging, meaning that the "environment provides opportunities for success on tasks that are difficult for the player" (Mayer, 2014, p.6),
- cumulative, meaning that the "current state of the environment reflects player's previous actions and allows for assessment of progress toward goals" (Mayer, 2014, p.6),
- inviting, meaning that the "environment is interesting, appealing, and fun for the player" (Mayer, 2014, p.6).

Mayer's definition differs from the previous ones in a few key areas. For one, he focuses on games for learning, games that are "intended to promote a change in the learner's knowledge or skill." (Mayer, 2014, p.8). For another, the definition includes a few characteristics that the previously discussed ones did not. Especially the idea that games have to be inviting is as unique as it seems obvious. It is certainly a characteristic that would be named by players if asked what defines a good game. It is, however, a subjective characteristic. A game may be appealing, interesting, and fun for one player but not another. We have seen a similar phenomenon in Juul's classic game model — a definition that allows for the same system to be considered a game under certain circumstances, but not under others. In Juul's case, however, a complete description of the circumstances (especially the negotiability of outcomes) still allows for an objective decision. In Mayer's definition, only the players themselves can really know whether what they are playing is actually a game. This makes Mayers' definition useful as a set of characteristics to aim for when creating a game, or indeed gamifying a system, but less so in the analysis of existing systems. Mayer combines his characteristics with some game elements adapted from Kapp (2012). Responsiveness, for example, is reflected in the game elements of feedback and replay, the cumulative nature in reward structures. Again, these are useful as a design guideline — if one wants to make a game more appealing one could try adding storytelling or improving aesthetics — but less so as an analytical tool. The combination of Mayer's work with the other authors discussed above forms the basis for our understanding of what games are and at the same time is one of the two major starting points for our work in creating an instrument for the qualitative analysis of gamification. The core concepts from the definitions above, combined with the goals and means of gamification we identify below are the source of the original terms used in our expert survey.

2.2 Gamification

Gamification as a field of study is so young that almost every larger work on the topic begins with an attempt at defining gamification. Some authors do so in a very structured and narrow way, while others prefer broader, more inclusive definitions formed from practical experiences. Unfortunately, authors have yet to agree upon a common definition, making research on gamification that much harder. Before one can find a good definition for gamification, it seems proper to define attributes of a good definition. According to Locke, a definition of a concept "accomplishes two things: (a) it ties the concept to reality, and (b) it distinguishes the concept from other concepts" (Locke, 2003, p.416). He also notes that a good definition “identifies only the most fundamental differentiator” (Locke, 2003, p.416). Proper scientific communication would therefore not only require scholars to agree on one definition of gamification to use, such a definition should also clearly distinguish gamification from related concepts (such as serious games) while at the same time being as succinct as possible. We will discuss the distinction between gamification and similar terms later in this chapter.

We have briefly discussed the issues with real and nominal definitions (Arjoranta, 2014) when discussing games above — the existence of a large variety of different expressions that make the identification of “some attributes that are in some way essential to the object being defined” (Arjoranta, 2014) problematic. Real definitions are inflexible and do not cover Wittgenstein’s family resemblances well. Our following analysis will, just as in the case of game definitions, not focus on creating a new, all-encompassing, real definition of gamification. We will rather attempt to summarize the discourse on what is currently being considered gamification and attempt to identify elements commonly associated with gamification instead of those common or unique to it. A nominal definition (a definition agreed upon by convention) can be very useful for scientific discourse and we will generally follow Deterding et al.’s meaning of the term gamification as “the use of game design elements in non-game contexts” (Deterding et al., 2011) in this work, but include findings from other authors as well.

We have analyzed a variety of definitions, both from a scientific and from a practical background. The selection of definitions is by no means all encompassing, as definitions of gamification are almost as common as published research thereon. Rather than providing a complete overview of definitions, we chose a set of definitions that are commonly referred to in literature and that are rather distinct from each other, so as to be able to extract a broad feature space for synthesis. We have analyzed three peer-reviewed, scholarly definitions by Deterding et al., Huotari and Hamari, and Blohm and Leimeister and four definitions from books on the topic of gamification by Zichermann and Cunningham, Zichermann and Linder, Duggan and Shoup, and Kapp. While the former focus on analyzing gamification, the latter deal with the question of how to properly create gamified systems. Our analysis was focused on identifying the goals and the means of gamification discussed by the various authors, a distinction we introduced in (Broer & Poeppelbuss, 2013).

You will find a detailed analysis of the goals and means in section 4.1, showing in which ways the examined definitions of gamification overlap as well as those in which they diverge. We will provide a brief summary of the results here, to present our overall understanding of gamification, however. Most definitions consist of a description of the means that are used for gamification as well as the goals that one wishes to achieve. Concerning the means, it seems prudent to only include very general means in the definition and to support those with a more detailed non-exhaustive list of potential tools (specific means), if one is looking to create a definition. We have summarized our findings in table 4.3, table 4.4 and table 4.5 in section 4.1, listing the goals/means mentioned in each analyzed paper, ordered by frequency of mentions. Unsurprisingly, all authors agree that games have to play a part in the definition of gamification. The range of which parts of games define gamification varies widely though. Most academics seem to favor the narrower definitions of either *game design elements* or *game mechanics* in their definitions, while practitioners seem to favor very broad terms such as *game thinking*. It seems that academic communication is best served by a more narrow definition of the term, so as to be able to distinguish between gamification and related terms such as *serious games*.[11] This would also fulfill the requirement for specificity described by Locke (2003).

The more detailed lists of possible means to be used for gamification differ from author to author, as seen in table 4.5. It is clear that badges, levels, leaderboards and points are indicators of gamification as they show up in almost all of the examined definitions.[12] Many other elements found in games are employed as well, however, and there are likely more that simply have not been tried yet. No definition can give a comprehensive list of means to use in gamification. The fact that gamification is so often associated with the first four means on our list has led to a dispute over whether one should rely on a different term (e.g. gameful design) to describe systems that do not rely on them. Unless one can show a clear distinction between these four means and all the others found in the analysis, it might be prudent to summarize these four most common tools as a subset of gamification

[11] See subsection 2.2.1 for a discussion of these related terms.
[12] Kapp does not include leaderboards and Deterding et al. do not include points specifically.

instead – the term *pointsification* has found some use in online and academic discussion (e.g. Nicholson, 2012), usually attributed to (Robertson, 2010).

There is even less consensus where the goals of gamification are concerned, not even on the question of whether goals should be part of a definition at all. Deterding et al. do not include goals into their definition and the other authors are divided depending on the context that they studied, as evident in table 4.3. This indicates that one has to be very careful when including goals in a definition. If one does so at all, one should limit oneself to very generic goals, such as *engagement* or *gamefulness* — both of which come with issues of their own. Further empirical evidence is required to provide a truly useful definition of gamification and eventually a model or categorization of different approaches to gamification. So far we have shown that any such definition is likely to be somewhat similar to Deterding et al.'s "use of game design elements in a non-game context", eschewing the mention of goals as they are not necessary in distinguishing gamification from other concepts. Goals are important for both research on and implementation of gamification, but their use in definitions has been shown to be problematic at best.

While we cannot find a nominal definition of gamification, the definitions do not vary so much that the term becomes useless. It is clear from all authors' writing that gamification deals with making something (be it *learning* (Kapp, 2012) , *value creation* (Huotari & Hamari, 2012), or a generic non-game context (Deterding et al., 2011)) more of a game. Most of the authors discussed here see gamification as the process of taking something from games (be it *game design elements* (Deterding et al., 2011; Blohm & Leimeister, 2013), *game mechanics* (Zichermann & Cunningham, 2011; K. Duggan & Shoup, 2013; Kapp, 2012), or the very broad *game thinking* (Zichermann & Cunningham, 2011; Kapp, 2012)) whereas Huotari & Hamari (2012) stress the importance of gameful experiences being created, no matter the means. Which parts of games are used differs from definition to definition, but badges, levels, leaderboards, and points are basically agreed upon. Similarly, the reasons for doing so vary, but come down to motivation and engagement in most cases.

2.2.1 Distinguation between Gamification and Related Terms

There are a variety of terms similar to gamification that are often confused with it: *serious games*, *games with a purpose*, *game based learning*, *games for learning*, *edutainment*, and *exergames*. All of these, and gamification, have in common that they refer to a use of games that goes beyond pure entertainment. They present different approaches, however, and should not be confused. Serious games may be the oldest of the terms and its definition is unfortunately also subject to discussion. The earliest available definition goes back to Abt, who requires serious games to "have an explicit and carefully thought-out educational purpose" and not to be "intended to be played primarily for amusement" (Abt, 1987, p.9). Later definitions of the term have extended beyond educational use. Zyda (2005), for example, defines a serious game as "a mental contest, played with a computer in accordance with specific rules, that uses entertainment to further government or corporate training, education, health, public policy, and strategic communication objectives." Even this definition seems oddly specific in its broadness as it names a variety of fields that serious games can be used in but by naming them implicitly excludes other fields. For the purposes of this work, we will follow Micheal and Chen's definition of the term instead. According to them, serious games are "games that do not have entertainment, enjoyment or fun as their primary purpose" (Michael & Chen, 2006). It should be noted, that this definition is not without contention. Loh et al. (2015, p.14) stress the importance of performance/skill improvement in serious games and distinguish them not only from entertainment games but also "message broadcasters/edutainment". This distinction is certainly necessary in research on game-based learning, but it leaves us without an overarching term for games with a non-entertainment purpose; hence our reliance on Micheal and Chen's definition in the following.

Similar to gamification, serious games therefore combine a non-game context (e.g. training) with games. Deterding et al. very clearly state a separation in their definition however: Where serious games refer to the use of complete games, gamification refers to the use of elements from games. Other authors are less clear on this distinction. It is useful for academic purposes to distinguish between the two approaches since the results are quite different. Contemporary gamified applications with their points and badges would clearly fail to be considered games in most of the classic definitions, while serious games very clearly are games. Loh et al. (2015, p.9) discuss this, saying “Gamification is not a game at all! Instead, it borrows from the concept of game mechanics to motivate people to continue certain behaviors [. . .] through points systems, badges, or monetary awards”. The distinction is not always this clear, however. Let’s assume, as an example, that we could take an existing game and break it down into its elements (a difficult task, but not hard to imagine in the abstract), then we would take all these elements and apply them to a non-game context. We would end up having reassembled the game, all the while following Deterding et. al’s suggested approach of using game elements only. At which point did we stop creating a gamified system and started creating a game instead? Deterding et al. already depicted the difference as a gradient, not an absolute (Deterding et al., 2011), which is just not reflected well in their use of “whole game” as a descriptor. We would suggest that the two ends of the scale should not be “whole” and “parts”, but rather “serious” (or a similar term) and “game”. Serious games and gamification would then be different expressions of the same phenomenon. Figure 2.4 shows this concept with examples for various types of serious games and gamification along the scale; we call it the serious–game scale. Both would be found somewhere in between the poles (not fully serious, not full games) with gamification being rather closer to the serious pole, while serious games are further towards the game side of the spectrum. Essentially, gamification would take a serious system and make it more game-like, while serious games take a game, and make it more serious.

This view also allows for mixed forms, which clearly exist. There are various examples out there, that are hard to classify as either a serious game or gamification — such as the protein-folding application fold-it[13] (Cooper et al., 2010). Depending on your point of view, fold-it can either be considered a protein folding application with game elements (such as points and a leaderboard) or a puzzle game that delivers protein folding data. For the purposes of this thesis, we will consider systems that are clearly tilted to one end of the scale or the other appropriately, and err on the side of inclusion in borderline cases. It should be noted that there are authors that do not consider this distinction to be relevant

[13]http://fold.it

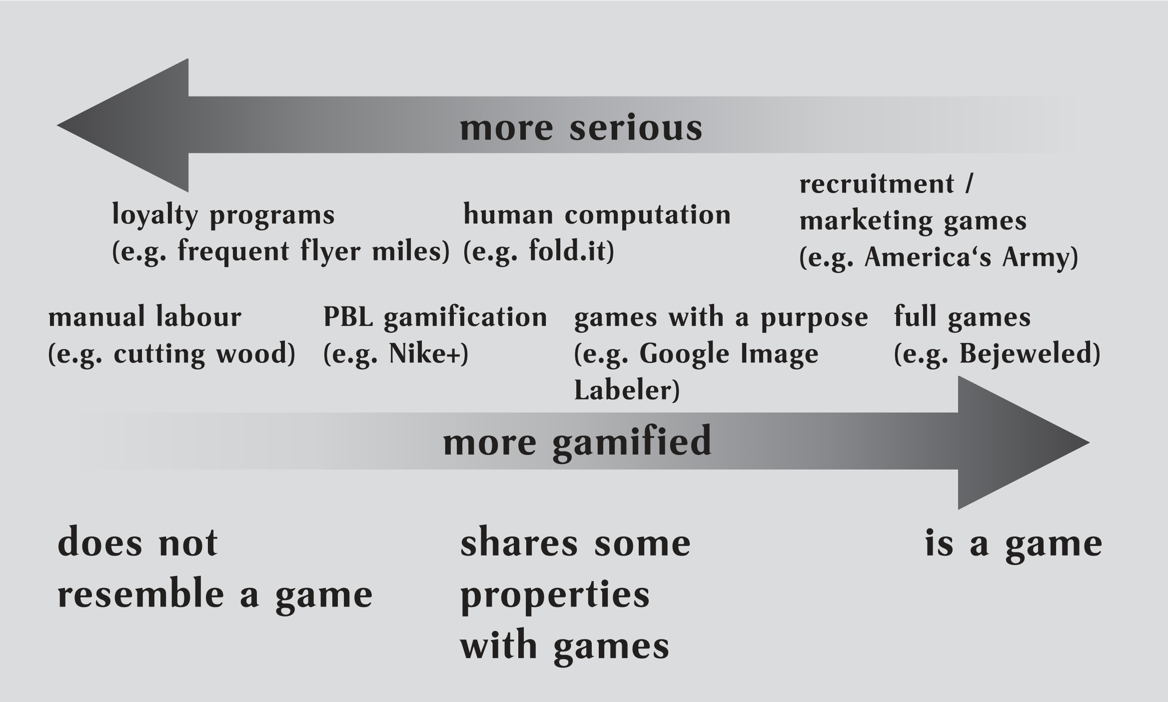


Figure 2.4: The serious–game scale. Gamification and serious games approach a synthesis of a non-game context and games from two sides of the same spectrum. Example placements are approximate, the further right an example is, the more game-like it is. The vertical axis has no meaning. Own illustration.

at all. Kapp (2012, p.16–17), for example, considers serious games and gamification to be essentially the same. Considering the ambiguity shown in the serious–game scale, there is certainly some merit to this. From a process point of view, however, there is a difference between taking a game and adding serious elements to it and taking a non-game context and adding game elements to it. When we refer to gamification, we will therefore refer to systems that can be found between the center and the left side of the scale in figure 2.4, systems that have been enhanced with game elements.

The other terms mentioned above can be considered subsets of serious games. Games with a purpose are a form of crowdsourcing, or human computation, that employs games as motivators. von Ahn (2006) first described this concept and gave image labeling as an example. Two players would be paired in a game, shown the same image and would each have to guess what words the other would use to describe the image. While this is purely a game from the users' perspective, the operator of such a system would gain valuable information about the contents of an image — a task that computers are not yet very good at. Such data can be very useful, for example for a search engine.[14] The aforementioned fold-it is also often mentioned as an example, although it differs in the transparency of its use. It is very clear to the users of fold-it that they are helping to fold proteins instead of simply playing a game.

Oh and Yang studied a variety of definitions for the term exergame, and proposed to define exergames as "an experiental activity in which playing exergames or any videogames that requires [sic] physical exertion or movements that are more than sedentary activities and also include strength, balance, and flexibility activities" (Oh & Yang, 2010, p.10). In essence, exergames are games used to promote exercise or physical activity in general. Exergames have been created for a variety of target groups and purposes. Many studies focus on the elderly, such as the work by Brox et al. (2011) on increasing seniors' physical activity through the persuasive effects of video games; a study of the impact on cognition by Anderson-Hanley et al. (2012), or the work by Agmon et al. (2011) on improving elderly user's balance. The target group is not the main differentiator between exergames and gamification, however. Exergames are rather found at the "game" end of the scale we proposed above, focusing on full games that make the user exercise while playing rather than enhancing exercise with game elements, as seen in the examples studied by Lister et al. (2014).

One area in which games are being studied extensively is that of education and informal learning. We have already discussed that serious games originated in this area, but there is a variety of other terms that cover games with an educational intent. *Game based learning* is probably the most commonly used these days, but *games for learning* and *edutainment* are also mentioned. Breuer & Bente (2010) provide a good overview over these terms and their differences. For the purposes of our research, however, these terms are similar enough (and indeed overlapping, as seen in Bente and Breuer's work), that we will simply refer to *game based learning* to cover all of them. Prensky (2001) prefers to discuss *digital game based learning*, but a limitation to digital systems seems unnecessarily restrictive. (See section 1.1

[14] As evidenced by the fact that Google licensed von Ahn's system and used it to improve image search results for five years. The company later launched an application with a similar approach for collecting landmark data for its mapping service. The mobile game Ingress rewards users for submitting pictures and descriptions of local landmarks (Celino, 2013).

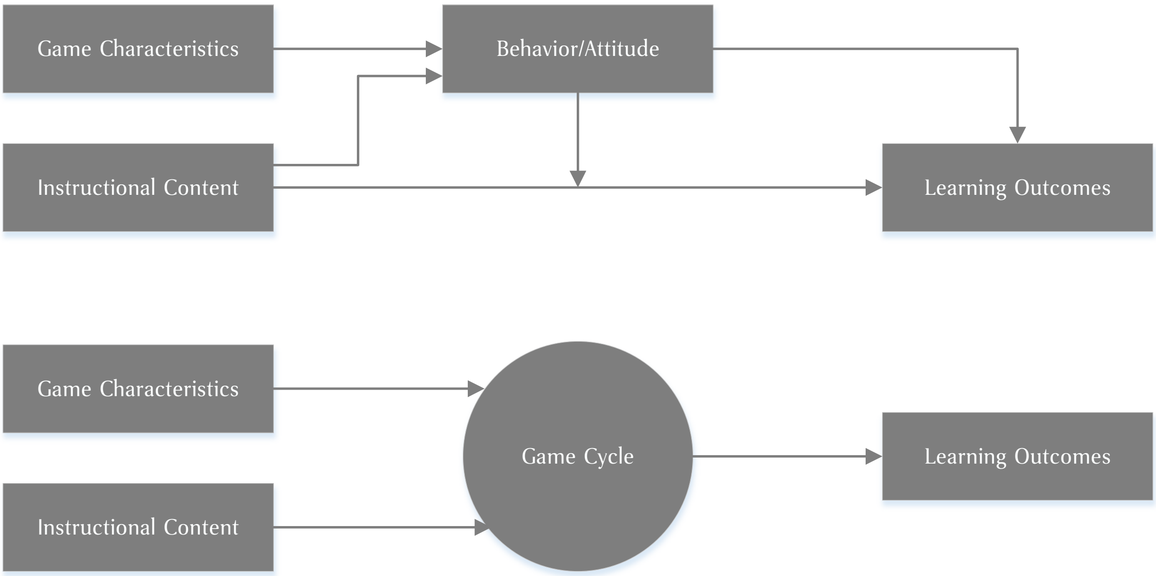


Figure 2.5: Top: Theory of gamified learning, adapted from (Landers, 2014). Bottom: Input-Process-Outcome Game Model, adapted from (Garris et al., 2002). It is proposed that game characteristics in gamification have an indirect impact on learning outcomes, while their impact is direct in serious games.

for a discussion on the relevance of the medium.) Game based learning, as we use the term here, therefore refers to any use of games to foster learning. On the serious–game scale discussed above, game based learning would be situated on the side of games rather than the serious side. Game based learning employs (more or less) full games that educate or train. This is not limited to games that were actively created to include educational content, but also includes games that are used to teach certain skills. E.g. one might use a massively multiplayer online game (MMO) to teach management and teamwork.

Landers (2014) proposes a theory of gamified learning that distinguishes between gamified learning and serious games. He defines gamification of learning as "the use of game elements, including action language, assessment, conflict/challenge, control, environment, game fiction, human interaction, immersion, and rules/goals, to facilitate learning and related outcomes" (Landers, 2014, p.757) in a synthesis of the definition by Deterding et al. (2011) and a taxonomy of game attributes by Bedwell et al. (2012). The major distinction Landers makes between gamification of learning and serious games (for learning) is the way in which they use game characteristics to impact learning. He adapts an input-process-output model from Garris et al. (2002), stating that instructional content and game characteristics are combined to form the game cycle which in turn directly causes learning. In Landers' theory of gamified learning, on the other hand, instructional content and behavior/attitude both cause learning, where instructional content influences behavior/attitude and the latter has a moderating effect on the learning effect of the former. Game characteristics only come into the picture in so far as they influence behavior/attitude, having only an indirect effect on learning outcomes. Landers (2014, p.760, emphasis in the original) states that "[t]he goal of gamification cannot be to *replace* instruction, but instead to *improve* it. If the instructional content does not already help students learn, gamification of that content cannot itself cause learning."

Both the mediating and the moderating paths in Landers' model rely on gamification causing a targeted behavior/attitude and on that in turn either causing learning (Landers names time on task as an example) or moderating the learning process. This creates two possible points of failure for a gamification effort: gamification could be ineffective at causing the target behavior, or the target behavior might not affect learning outcomes (through either moderation or mediation.) Importantly, this means that the measurement of learning outcomes in gamified learning can be less useful in determining the effects of gamification as a tool in general. One might employ an effective gamification strategy (as in: the employed gamification elements cause the desired target behavior), yet not see any effect on learning outcomes. The same can also happen if the instructional material was ineffective to start with (c.f. Landers, 2014). We discuss Landers' theory further in the context of collecting data for our planned experiment in subsection 5.2.3.

2.2.2 Structuring Gamification

Some authors go beyond definitions and attempt to structure gamification. These attempts at structure take various forms — the authors refer to their work for example as a "design space" (Schering, 2014) and a "framework" (Chou, 2015). They are all attempts at dissecting gamification and providing a structure to its analysis or construction. They differ from the definitions discussed above in that they provide more structure (as opposed to simple lists of goals and means) and at the same time do not attempt to provide a nominal or real definition of gamification. Some of these are intended for guidance in the process of gamification itself, while others are aimed at categorization. All of them are created from theory and personal experience and not sufficiently validated empirically, as far as published material goes.

Kappen and Nacke originally introduced the kaleidoscope of effective gamification (Kappen & Nacke, 2013) as a tool for employing gamification in a business context, but have since also employed it in their research (Kappen et al., 2016). They present gamification as a set of interconnected layers converging on a core they call *effective gamification* (see figure 2.6. On the outside of their model is a layer of (perceived) *fun* that the user experiences, followed by a *game design process* layer with *principles*, *mechanics*, *models*, *patterns*, and *interface* as its components, followed by a layer of *game experience*, comprising *achievements*, *challenges*, and *actions*. The final layer surrounding the core is the *motivated behavior* layer, built on SDT's intrinsic and extrinsic motivation. Intrinsic motivation is represented by SDT's classic combination of *autonomy*, *competence*, and *relatedness* while extrinsic motivation comprises *badges*, *points*, *leaderboards*, *incentives*, and *rewards*.

The connections between and the order of the layers remain somewhat vague in the original publication. It seems logical that the game design process layer informs the design of a game experience in the respective layer which in turn leads to specific design decision that motivate behavior either intrinsically or extrinsically. The elements of the extrinsic motivation section, however, could also be seen either as game mechanics or even game design patterns. Additionally, while the individual elements subsumed under extrinsic motivation are almost directly useful when informing game design, competence, autonomy, and relatedness are in themselves target states that require their own affordances which are not specified in the original publication. In (Kappen et al., 2016) the authors expand the inner layer — motivated behavior — with what they call *themes*. They see competence as either engagement, achievement, or performance based. Autonomy is reached through customization, purpose, or independence and relatedness through relationships, sharing, and preferences.

Schering (2014) introduces a design space for gamified applications that is intended for

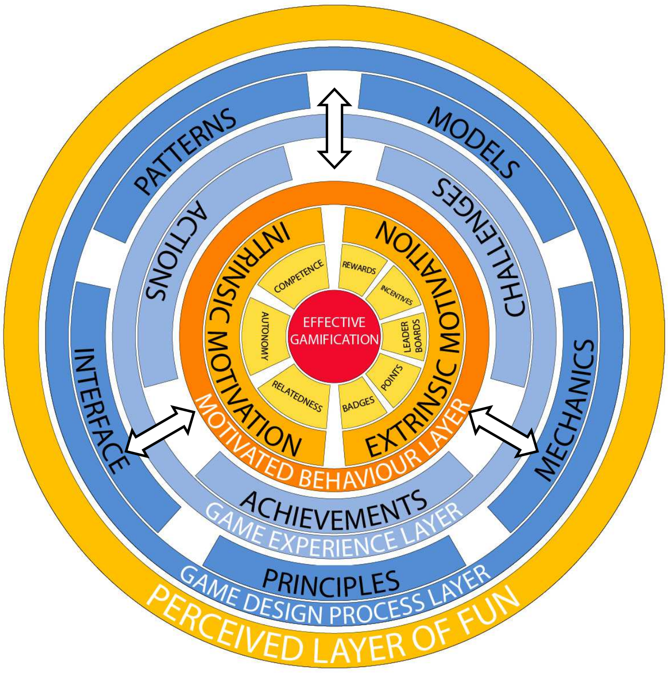


Figure 2.6: The Kalaidoscope of Effective Gamification. Image taken from (Kappen & Nacke, 2013)

the categorization of gamified systems. Such a categorization, according to the author, is necessary as a basis to identify the effectiveness of different approaches to gamification. Schering's design space consists of eight dimensions: "(1) kind of motivational structure, (2) strength of gameful character, (3) location of performed tasks, (4) adaptivity to changing skills, (5) adaptability through players, (6) level of integration into work, (7) addressed range of player drivers, [and] (8) strength of social interaction" (Schering, 2014, p.364). Schering has applied her model to 23 gamified applications, resulting in a strong overlap, signifying strong similarities in approaches to gamification. Essentially, the model describes eight properties that a gamified system can have (to varying degrees). E.g. it can focus on intrinsic or extrinsic motivation (dimension 1, kind of motivational structure — Schering refers to the poles as "meaningful" and "reward driven" instead) and it can adapt more or less to the capabilities of the user (dimension 4, adaptivity to changing skills). While certainly useful for categorization, some of the dimensions seem too broad for meaningful separation. Especially the first dimension seems in need of elaboration. As the section on motivation above has shown, there are various ways of motivating users with varying effects and simplifying them to a single scale risks combining very different gamified systems into the same category.

Chou approaches gamification from a practitioner's perspective and therefore does not provide much substantiation for his resulting structure, the Octalysis framework (Chou, 2015). Nevertheless it is often referenced and detailed and deserves attention. His framework comprises eight dimensions as well (modeled after what the author calls the "8 core drives of gamification"), yet it differs greatly from Schering's: meaning, empowerment, social influence, unpredictability, avoidance, scarcity, ownership, and accomplishment. Each dimension has a variety of game mechanics associated with it that appeal to the respective core drive. E.g. points appeal to the accomplishment drive while group quests might appeal to the social influence drive. Chou's framework is intended to be used in one of two ways. One can use it for the analysis of existing systems to see how (and how much — Chou attaches value to the dimensions) it has been gamified, but one can also use it to inform the design process with certain drives in mind. Chou further suggests to use the framework to analyze the drives being appealed to in different types of users and in different stages of using the system, suggesting that you need to provide gamification for each type of user in each possible stage of using the system. Chou uses Bartle's player types for this separation (see Bartle, 1996).[15]

These attempts at structuring gamification are especially interesting for a comparison with our own approach of using family resemblances. A more structured comparison would be very interesting for future work, such as analyzing the same set of gamified systems with different approaches, clustering the results and identifying similarities and differences. As it stands, these approaches are not suitable to our further work, however, demanding the development of our own structure.

2.2.3 Gamification Elements and Design Principles

A term that has become common in gamification literature, especially among German researchers, is that of the *gamification element* (see e.g. Korn, 2012; Burkhard et al., 2013). Unfortunately, the term is not usually defined well. If we use the definition of gamification by Deterding et al. (2011), a gamification element might be a game design element that has

[15]We discuss categorizing players on the basis of their motivation to play in subsection 2.3.3.

been transferred to a non-game context. We will understand it that way for the remainder of this work. A gamification element is a clearly separable element of a non-game system that is commonly used in game design. Essentially, the term aligns well with the previously discussed *means for gamification*, but always refers to an implementation. Means, as we introduced them in (Broer & Poeppelbuss, 2013) are a broad category, encompassing gamification elements but also broader overarching concepts. A leaderboard, therefore, will be considered a gamification element while competition will not, as leaderboards are one implementation of the *design principle* (see Dicheva et al., 2015) of competition. These two categories seem to match Blohm and Leimeister's (Blohm & Leimeister, 2013) approach, with gamification elements equating their *game mechanics* and game design principles equating their *game dynamics*. We will discuss a few common design principles for gamification and example gamification elements in the following sections. You will note, that the design principles are not completely disjunct and that in many cases certain gamification elements can be used in various ways, combining multiple design principles. The principles are also non-exhaustive. Since we do not (and may never) know all possible gamification elements, it seems difficult to create a complete list of the overarching principles. Additionally, even existing principles are not agreed upon in their definition; the distinction between *progress*, *accrual grading*, and *visible status* by Dicheva et al. (2015), for example, is rather arbitrary and depending on context. Other principles mentioned by the authors, such as *storytelling*, may even not be considered gamification at all, depending on the definition one abides by. As we cannot provide an exhaustive list, we will introduce the principles most commonly found in literature here.

Competition

Competition is only one of a variety of means for gamification discussed above, but it is one of Calliois original four types of play and an element that can be found in most modern games (digital or otherwise) that include more than one player. Vorderer et al. (2003) call it *social competition*, defining it as "a process which develops by competitive actions performed by individuals or social entities in order to maintain their own interests to the disadvantage of others.” They specify that competition between individuals should be called social competition, because their understanding of competition itself also includes other forms of play, such as fighting a computer controlled boss monster. Their understanding of competition aligns best with what other authors have referred to as conflict (e.g. Salen & Zimmerman, 2004). Hartmann (2009) describes the difference between two types of competition: that in which one tries to beat a certain self-applied standard of performance, and that in which two participants[16] attempt to achieve incompatible goals. He calls the second form social competition as well. According to him, all computer games include the first type of competition, but only the ones that provide social competition can be called competitive Hartmann (2009, p.211f, translation ours). When we discuss competition here, we will be referring to conflict between players, or social competition.

Two reasons make competition stand out as a gamification design principle worthy of detailed discussion. For one, we have shown in section 4.1 that leaderboards are among the most common means for gamification discussed in literature and they are driven by competition: The main functionality of a leaderboard is to show different participants in relation to one another. Users are enabled to compare their progress to that of others so that they are motivated to improve and beat them. Being the best, or at least better than somebody else, is the core motivator of a leaderboard — classic competition. Vorderer et al. essentially describe the functionality of a leaderboard in their further discussion of social competition: “Each social entity monitors the process by ongoing evaluations of the ‘status quo’. These evaluations include the perception about how the current individual’s position is in contrast to the positions of the others and what tendency is expected for the further process of the competition.” Vorderer et al. (2003)

The second reason for a prominent discussion of competition is that actual negative effects of including it in a gamified application have been shown. In their study of the effect of competition and collaboration on learning in an educational mathematics game, ter Vrugte et al. (2015) found that the effect of competition on collaborative learning was positive for above-average students, but negative for below-average students. This is consistent with

[16] He later refers to these as “social entities” (Hartmann, 2009, p.211, translation ours) and includes simulated characters with artificial intelligence among those as well.

the findings of Schubert et al. (2014), who also found a negative impact of competition on a subset of the participants. Eck & Dempsey (2002) studied the effect of coaching on the transfer of mathematic skills in a simulation game, employing both a group with competition and one without. They found positive effects of coaching in the group that did not include competitive elements, but negative effects in the group that included competition. Finally, Mayer (2014, p.154ff) summarizes that they could find no “strong and consistent evidence to support incorporating competitive features into an educational game”.

As with all other gamification design principles, we do not yet have sufficient data to fully understand the effect of competition on the users of a gamified environment. Existing studies suggest, however, that different users may react to competition differently and that competition can have a significant negative effect on users. As seen in the results of ter Vrugte et al., competition is especially problematic with low-performing users. This is not surprising — seeing that you are one of the best performing users of a system can be motivating, while seeing that you are among the worst might lead to disappointment and a feeling of never being able to catch up. A trend seen in computer game design is the so-called no-disincentive leaderboard (Zichermann & Cunningham, 2011, p.50). Instead of showing the top users of an application and the active user’s potentially very distant positioning, only the part of the leaderboard surrounding the user’s position is shown. That way, all competitors are close by and rising on the leaderboard is made to seem possible. We do not know of any studies measuring the effect of no-disincentive leaderboards on user motivation, however.

Rewards

Even more so than competition, gamification elements based on the design principle of rewards are ubiquitous. As we have shown in section 4.1, points and badges are commonly named means in definitions of gamification and empirical studies of gamification almost always include such elements as well or even solely focus on them. Hamari et al. found points, badges, or rewards being used in 15 of the 24 studies they reviewed and Dicheva et al. found points in 22 of their 34 studies (not to mention other reward-based elements.) The general idea behind the rewards design principle is that users are given rewards for actions deemed desirable by the system's designers. These rewards do not have to have material value, and indeed most implementations do not include material (or tangible) rewards at all. Instead, users are given virtual rewards, most commonly in the form of points or badges. Points are usually used to show progress — a user with many points should be considered more successful at using the system in question than one with few points. Points are not always actually bound to success (however one might define it), but to action instead. Users might be rewarded with points for adding content to a system, for example, without a check for the quality of that content. In some cases, points are used as a currency that can be spent — either on actual, tangible rewards or on intangible rewards within the system, such as visual changes to an avatar. Points are usually used as frequent reinforcement, with small point gains happening at a high frequency.

Badges (also known as achievements) on the other hand are designed to happen less frequently but to have a larger impact on the user. Badges serve as indicators of milestones in the user's progress — in a video game one may attain badges, for example, for beating a difficult boss, completing a section of a game, or performing a specific action a certain number of times. Hamari & Koivisto (2015) describe badges on a systemic level as consisting of "a signifying element (the visual and textual cues of the badge), rewards (the earned badge), and the fulfillment conditions which determine how the badge can be earned". Often times, badges not only mark such milestones after the fact, but also serve as indicators of goals for the user. Games, and gamified systems, achieve this by making attainable badges visible and in some cases showing the progress towards them. Not all games follow this pattern, however. Some game designers see value in surprising the user with badges, encouraging exploration. Unlike points, badges make use of user interest in collecting (c.f. Toups et al., 2016).[17]

Both points and badges can be used to signify status, if they are visible to other users of the system. Initiatives, such as Mozilla Open Badges[18] even attempt to make such status

[17]Badges should be considered part of the "memory" type of value in the taxonomy by (Livingston et al., 2014), as it was used in (Toups et al., 2016). As the authors note, collecting goes far beyond the way badges work, indicating a lot of potential for further gamification.

[18]http://openbadges.org/

visible outside of the gamified system itself. They see an attained badge as proof of personal achievement that can be valuable to show to others. In the case of Mozilla Open Badges, the focus is on badges earned for learning, such as the completion of online courses. Badges, according to the initiative, are “an online representation of a skill you’ve earned” (Mozilla, 2016). Even within a single system, status can be an important reward to users, however. One of the most referenced examples for gamification is Foursquare, a mobile application that used to allow users to check-in to locations one visited and to earn badges for such check-ins. The gamified features of the application have now been moved to a separate application called Swarm, but the general functionality remains intact: Checking in to a variety of locations can earn you badges, and frequently checking in to the same location can make you its (virtual) mayor for a while.

The Foursquare example shows, that rewards and competition often work hand-in-hand. Most leaderboards, for example, work on the basis of point rewards. Instead of simply using points as an indicator of personal progress, however, leaderboards put them in the context of other user’s points, causing competition. Similarly, badges can be based on competition — for example when only the first person to reach a goal is awarded a badge.[19]

In addition to rewards of status, users can also be offered more direct rewards for their participation. These can be either tangible or intangible, although tangible rewards tend to come with real costs. Tangible rewards often take the form of loyalty programs — such as the Starbucks mobile application that offers free coffee to the user once a certain amount of drinks has been purchased. Loyalty programs are often seen as precursors to gamification (Zichermann & Linder, 2013, p.37f) and tangible rewards tend to be the remnants of that earlier age. There are not many gamified applications that offer intangible rewards either, although they are common in games. The digital card game Hearthstone[20] offers examples of different types of intangible rewards in games. For one, there are rewards that help you progress within the game itself — e.g. players are rewarded with cards that they can then add to their deck to improve it. For another, there are cosmetic rewards

[19]An example from games would be World of Warcraft’s server first achievements — badges only given to the first player to reach a certain goal. E.g. the first player to reach level 80 — then the maximum level — on each server would get a badge.

[20]http://www.battle.net/hearthstone/

— rewards that change the look of parts of the game but do not affect the game itself. In Hearthstone, a player who has won 500 games with a specific hero gets access to an animated, golden version of that hero. The former type of reward is difficult to transfer to a gamified application, as it is usually not in the interest of the system's designer to limit users in their capabilities. E.g. it makes little sense to make chapters of a textbook only available to those members of a class that reach a high score on a test. That does not mean that it is without merits, however. Euclid: The Game[21] is not only an example for an application that is somewhere in between a serious game and gamification, it also exemplifies the use of in-game rewards well. Users are rewarded with advanced geometrical tools once they have shown that they understand how to solve problems without them, making it more comfortable to solve more complicated puzzles.

As a rule, rewards are a form of extrinsic motivation and therefore potentially subject to the overjustification effect (see subsection 2.3.1). If the user is performing certain actions in order to get a reward, he or she is not doing because of the activity itself. Such extrinsic motivators rely on the user's interest in being rewarded — if rewards are or become meaningless to the user, there will be no motivational effect. A second positive effect that rewards can have on the users is guidance. Both points and badges, if used appropriately, can guide users through their use of the system in question. We have already mentioned that badges can serve as indicators of milestones, separating a large task into more achievable chunks. They can thereby make goals both more specific and attainable, as described by Locke (1996).

Phillips et al. (2015) present a classification of video game reward types based on focus groups and validated by an expert review of sixty video games using the classification. They found that rewards of access, facility, sustenance, glory, praise, and sensory feedback were all used in the games analyzed. Rewards of facility were most present (36.72% of categorized instances), followed by rewards of glory (20.10%). The least common rewards were those of sustenance (8.30%) and praise (5.35%). Interestingly, the rewards of access, facility, and sustenance all directly impact the future playing of the game, while glory, praise, and sensory feedback do not. The cards players can earn in Hearthstone are rewards of facility, as are

[21] http://euclidthegame.com

the tools one can earn in Euclid: The Game. The golden heroes in Hearthstone could be considered rewards of glory. While this taxonomy has yet to be applied to gamified systems, our following review of empirical studies shows mostly rewards of glory being used.

Much criticism of modern games, especially massively multi-player online roleplaying games (MMORPGs) , discusses them as *skinner boxes*,[22] referring to a branch of psychology studied by Skinner (1938) called behaviorism. Skinner discussed operant conditioning, the idea that behavior can be shaped by feedback through external stimulants. The term skinner box reduces Skinners complex arguments to the idea that individuals can be trained to do things through external feedback. Just like the rat in Skinner's operant conditioning chamber, players of certain games are made to perform activities purely for the sake of the associated reward, the critics say. Yee (2001) discussed this early in his MMORPG studies and the term has spread since (see e.g. Siang & Rao, 2003). One important element of operant conditioning is the effect of the reward schedule, e.g. random or fixed intervals. While this thesis does not address ethical issues with gamification, it should be noted that careful design of reward schedules can have a strong impact on player (or user) behavior. This can be used to improve engagement, but one must keep ethical issues in mind.[23]

Goals

Goal setting has been the object of much psychological study in the past five decades (see e.g. Locke, 1996) and it is highly relevant to both game design and gamification. If we go back to the game definitions we've discussed, an element that is mentioned repeatedly is that of a goal (or outcome). Mostly, they refer to the final goal of the game (e.g. beating the final boss), but games often incorporate smaller intermediate goals as well that guide the player and keep him or her interested. World of Warcraft (WoW)[24] is the most successful MMORPGs of all time and is still among the top overall MMOs (Statista, 2016a).[25] One major innovation

[22]The first explicit behaviorist was not Skinner but Watson (see Skinner, 2011, p.4).

[23]E.g. some modern free-to-play games have implemented almost lottery-style random interval reward schedules that encourage the user to spend money. Similar methods can probably be applied in gamification.

[24]http://www.battle.net/wow/

[25]Measurements of MMO popularity are difficult, as the games use different business models — e.g. free-to-play vs. subscribtion based games and numbers provided by the games' developers are not always reliable. An alternative to those numbers is the tracking of activity in online social media related to the game. http://mmo.blakey.co/, for example, tracks the activity for different MMOs on the online discussion platform reddit.

that WoW introduced over previously established MMORPGs was the massive use of quests to guide players (see e.g. Ashmore & Nitsche, 2007; McNamara, 2004). Other MMORPGs would generally simulate an open world and offer some quests to players, but WoW was designed in such a way that players were never left without a quest — and therefore a direct, attainable goal to follow. This has since been imitated by many other games, not just in the MMORPG genre. With quests like those, players always have something to strive for, an attainable goal, a known next step.

Koestner & Hope (2014, p.411) conclude that "goal strivings are most successful and adaptive when they are aligned with intrinsic, need-satisfying aspirations; when they are based in autonomous motivation; and when they are supported by empathic rather than directive others". Locke (1996), in a summary of 30 years of goal-setting research, names 14 findings, multiple of which support such a quest design from a psychological point of view. One major takeaway is the importance of the specificity of goals. Another finding stresses that individuals need to find goals to be both important and attainable. Especially attainability is reflected in the subdivision of final goals into more immediate subgoals. The user might not see the final goal as attainable, but each small goal can be designed to be so, increasing the commitment to goals (Locke, 1996, p.119).

Similarly, feedback about the progress towards a goal increases the effectiveness of goal setting (Locke, 1996, p.120). Subgoals can provide such feedback, but many games go further and also include progress measurements for those subgoals. A simple quest in WoW might require you to kill a number of specific enemies, for example. The progress towards this goal is shown as a counter in the quest description and progress is also visualized on screen whenever another monster of the right type is killed.

This design pattern can easily be transferred from games to non-game contexts. The process of writing an essay, for example, could be gamified by dividing the required work into small steps with clearly defined, measurable outcomes. These can then be presented to the user as quests. As briefly mentioned in the section about badges, they can also be used for goal setting. Instead of giving the user in the above example a set of quests, one

At the time of this writing, WoW is the clear leader with more than double the activity of the second place game Guildwars 2.

could also create a variety of badges to be earned — such as a badge that requires him or her to add at least 15 references to the document or to include specific sections (such as an introduction.)

Clear goals are also one of the elements of the flow experience described by Csíkszentmihályi (2008) as the "optimal experience". While the concept did not originate in games, it has been adapted to games in order to improve the user experience (e.g. Sweetser & Wyeth, 2005). The idea of creating a flow experience in a non-game activity is therefore not gamification, but games can teach us much about how to create such an experience. Clear goals are one aspect of flow that games incorporate very well. Interestingly, Dicheva et al. (2015) distinguish between goals and quests as two different design principles although quests are simply a way of formulating goals. We consider quests as a gamification element, whereas goals are a design principle.

Feedback

Three of the definitions for gamification we studied (see section 4.1) mention feedback as a means of gamification and none of the approaches discussed in subsection 2.2.2 do. This is far from unanimous, but it can be assumed that the inclusion is rather a question of whether feedback is a game element, not one of its general importance. Psychology has shown a positive effect of feedback on task performance on average (Kluger & DeNisi, 1996) and that it can positively influence intrinsic motivation (Ryan & Deci, 2000). Moreover, the temporal proximity of feedback (that is, rapid or instant feedback) has been shown to improve task performance (see e.g. Wu et al., 2012; Kettle & Häubl, 2010). Games are excellent at providing feedback quickly and visibly. Mistakes are often immediately obvious and punished by a loss of resources, points, or the avatar's life. Similarly, positive actions are often immediately rewarded with in-game advantages, points, or badges. Many non-game contexts on the other hand only provide delayed feedback — e.g. diets or physical exercise. Adding instant feedback to such activities may help with motivation and task performance. A variety of feedback mechanisms from games may be helpful in such an attempt. Gee (2005) has discussed feedback at the intersection of games and learning under the headline of “pleasantly frustrating”. According to him, well designed games make failure acceptable through making “learners feel – and get evidence – that their effort is paying off in the sense that they can see, even when they fail, how and if they are making progress” (Gee, 2005, p.10). Feedback is also at the center of flow theory (Csíkszentmihályi, 2008, p.54) as one of the eight components of the flow state, the optimal experience. While not originally developed in the context of games, the ideas behind flow have been picked up in academic discussion thereof (see e.g. Chen, 2007) because it is very useful in describing the optimal state of somebody playing a game and the steps it takes to achieve it. Csíkszentmihályi notes, that feedback has to be “logically related to a goal in which one has invested psychic energy” (Csíkszentmihályi, 2008, p.57). This does not have to be a goal that is provided by others (e.g. the game) but could also be a goal set by the player him or herself. In essence, feedback provides the player with information about whether whatever he or she did was useful in getting towards a goal they value. The technology they use enables video games to give rapid feedback on player inputs, a feature often lacking in real-world scenarios. The past years have seen an increase in research in using rapid feedback technology in education (see e.g. Cotner et al., 2008) and even the anticipation of rapid feedback has been shown to increase test performance (Kettle & Häubl, 2010).

For many non-game contexts, rapid feedback should be automated as it can be very difficult to have humans give that feedback. (Either because of the sheer number of simultaneous actions that require feedback — e.g. many users in a single system, many students in one class — or because manual assessment requires time — e.g. feedback on a class paper.) In order for computers to be able to give meaningful feedback, quality must be measurable. The difficulty of such measurements differs greatly depending on context. It is relatively simple, for example to give feedback to a user of a gamified running application. Measuring distance traveled and speed is trivial, and the user can be given essentially instead feedback thereon. In education this is much harder, but possible in some fields. There has been a lot of success in automated assessment of programming assignments, for example (see Ihantola et al., 2010), whereas the computer-based assessment of music competencies is still in its infancy (see e.g. Finken et al., 2015). If one wants to give rapid feedback on the quality of interactions with a gamified system, one must therefore employ knowledge specific to the field of application.

A simpler approach that can for example be seen in our evaluation of gamification in learning management systems (LMSs) in section 5.1, is to give feedback not on the quality of interactions but on their quantity. If one wants to encourage students to submit homework through a gamified system, for example, one might simply reward the act of doing so, instead of judging the quality of the submission. As our analysis of empirical studies below shows, this has the potential to lead to increases in the frequency of the desired behavior. However, the same studies show that the quality of submission may be reduced thereby. In subject areas where automated assessment is possible, employing such a mechanism for rapid feedback may be one of the more promising gamification approaches to date. In other areas, targeting specific behaviors with rapid feedback may also work, but one needs to be careful not to sacrifice quality for quantity.

Freedom to Fail

Many of the game definitions we discussed stress the importance of games being separate from reality (or, as Juul put it, to have negotiable consequences). This implicitly includes the property of many games that the user has the freedom to fail. Most games encourage experimentation and risk through mechanics that reduces the negative consequences of failing. Generally, any consequences from failures stay within a game (unless external consequences have been negotiated — e.g. playing for money). Furthermore, most games allow you to simply try again, either by restarting the game or by simply going back (through reloading or other measures). Many non-game contexts behave differently and failure has direct and irreversible consequences. An example from education is the exam. Students are rarely encouraged to try an exam, see how much they already know and given the opportunity to re-take it when they are better prepared. Usually, an exam is taken once and the results thereof become part of your academic record, causing anxiety. A similar effect can be seen in some games, when failure is permanent, such as in the case of online leaderboards. This effect is sometimes referred to as *ladder anxiety* (Lobel et al., 2013), a term coined from the term used for leaderboards in some games. In either case, anxiety can be prevented if failure does not have consequences. In education, for example, gamification can reduce the impact of failure through shorter feedback cycles (see J. J. Lee & Hammer, 2011) and simply through allowing students to re-take exams as often as they want to. Such an approach has obvious difficulties in implementation, however, as resources to grade exams or homework are not endless (automated evaluation is almost a must here) and equally one needs to prevent students from learning the test instead of the subject matter.

Another approach for allowing failure in education is to change the grading system. Standard grading systems that operate on averages essentially start the student out at the best possible grade which they can only worsen (or keep) with each grade they get. A bad exam will reduce your final grade, whatever you do. It has been suggested (and tried in small scale, see e.g. Dougherty, 2015; Simões et al., 2013) to replace grades with experience points. With this change, students are working towards a goal instead of trying to keep a level of quality and even a less-than-stellar performance in an examination will still get them part of the way to their goal. This change alone is purely one of perception, but if one combines it with optional assignments, failure suddenly does not feel as bad. If a student gets only a few points on a piece of their work, they can decide to hand in additional work to make up for the missing points. Freedom to fail is certainly one of the most difficult concepts to transfer from games to non-game contexts as it is the separation from the real world that makes games games in the first place. Education is a relatively easy example only because it, similar to games, follows a set of more or less arbitrary rules. The eventual outcome of education — learning — is its main part that needs to stand up to the scrutiny of the real world, everything else can be molded by the educator. In theory, an educator can decide to remove all consequences of failure and there are various approaches to education that eschew grades, for example (c.f. Lübke, 2013). Obviously, regulatory and other external circumstances limit educators in their freedom.

Social Elements

Social elements can influence motivation and performance in a variety of ways. Ryan and Deci discuss relatedness as one of three basic psychological needs and as a requirement for intrinsic motivation (Ryan & Deci, 2000) and those stages of external motivation with a more internal perceived locus of causality (PLOC). Social relations play an equally important role in Seligman's well-being theory. Unsurprisingly, Yee also found social relationships to be one of three main components of motivations to play online games, subdivided into socializing, relationship, and teamwork (Yee, 2007). If social relationships play a major role in motivation in general and the motivation to play games in particular, it seems only logical to use this fact in gamification as well. The most basic form of supporting social relationships in an application is to simply include capabilities for communication. Chat, messages, and friends lists are all commonly seen in gamified applications. It is questionable, however, whether this can be considered gamification, as neither feature is particularly specific to games. Instead, gamification should look not simply at communication capabilities, but at the way that games foster the building of relationships. MMORPGs have been especially successful at this, often requiring cooperation and teamwork in order for players to succeed and creating social relationships through the necessity for interaction. Ang & Zaphiris (2010) have analyzed social roles in these structures. The need for teamwork and cooperation is one game design element that could be transferred to non-game contexts to foster social relationships. Ducheneaut & Moore (2004) have suggested that meeting spaces are instrumental in supporting the creation of social relationships in online games — that is places that the game makes players spend time at, which they can then use to socialize with others. While gamified applications do not usually offer spaces in the sense that a virtual world does, examples of similar functionality can already be found. Codingame[26], a platform for gamified programming, for example, automatically makes users join chat rooms related to the task they are currently trying to solve.

Zichermann & Cunningham (2011) describe social engagement loops as a gamification element. The idea behind such loops is that the actions of the user trigger reactions of other users that then trigger reactions of the original user once again, keeping them engaged with the system. A non-gamified example is the use of twitter:[27] a user may be motivated to sign up for twitter and post something on there. This is a desired behavior that the developers

[26] www.codingame.com
[27] www.twitter.com

of the system would like to reinforce. If all that twitter did for the user was to offer a way to publish his or her messages, however, the user might lose interest quickly. Twitter incorporates additional functionality, in allowing other users to interact with a message: they can reply to it or retweet it (that is, forward it to other users) or mark it as a favorite. When these interactions happen, the original author is notified and thereby enticed to return to twitter and interact with the response. This forms a (theoretically) infinite loop of interaction that keeps users engaged. Games on social networks have attempted to create similar loops in the past by, for example, encouraging users to ask their friends for help in the game.

Autonomy

As we will discuss below, a feeling of autonomy is often seen at the core of intrinsic motivation. Games offer such autonomy to their players in varying degrees, ranging from very linear, story-driven games to those that are specifically designed to allow the player as much freedom as possible. We have discussed these variations and Aarseth's narrative and ludic poles earlier in subsection 2.1.1. Naturally, the idea of autonomy is interesting to those who attempt gamification, as intrinsic motivation can be a very worthy goal (see below) and as it is something that is inherent to games. Autonomy can come in many forms, but it always means that users are not forced into a specific action but do it of their own volition. In education, such choice may include that of topics to focus on, or tasks to do. As with many other gamification design principles (and the related elements), it can be difficult to specifically pinpoint games as the origin of the use of autonomy. Not only is it mentioned in a variety of works on motivation that were not written specifically for games or gamification (e.g. Deci & Ryan, 1985b; Seligman, 2011), it has also been used in non-game areas, such as education, before the term gamification was even coined (see Broer, 2015). Nevertheless, there is a lot that we can learn from games about autonomy, or, more importantly as we will discuss below, the feeling of autonomy. While games certainly provide autonomy to the player to varying degree, it can be very difficult to provide true autonomy in a narrative context, as we have discussed earlier. Since narrative is usually not generated by a computer but rather has to be handcrafted by humans, it is difficult to deal with the exponential growth of decision trees in narrative computer games. Even in a game of purely binary decisions, the number of potential outcomes grows quadratically with each decision. Computer game designers apply a variety of tricks to make the player feel in control of the narrative when most decisions actually have little to no impact. Crawford (2005, p.123ff) discusses a sample of these strategies (and why they do not work in interactive storytelling). One example are foldback schemes, that is leading branches of a decision tree back to a common node. Games designed in such a way provide multiple paths to the same (intermediary) goal, giving players the feeling of autonomy, especially if they do not play the game repeatedly and therefore do not notice that their choices do not matter. In an example from education, one might give students the choice between different publications to base a presentation on, knowing that all those publications include the same core message one wishes to discuss in class. Similarly, a math teacher may offer students the choice between different problems as homework, where all choices promote the same learning outcomes. In a final example, one might offer extra credits to students while using a rank-based grading system. Extra credits assignments are being presented as being optional, but the rank-based grading gives those who do the extra assignments a clear advantage, directly impacting the grades of those that do not.

As briefly noted before, providing an illusion of autonomy only works as long as the user does not see through that illusion. Realizing that one was tricked might even lead to negative consequences. In game design, games are often abandoned once they are fully understood — a concept at the core of Koster's theory of fun (Koster, 2005). In non-game contexts, this may not be an option, one would not want students to stop studying once they have understood that they are not actually autonomous in their learning, for example. There is little to no research on the impact of such a removal of perceived autonomy, but practitioners should keep the potential harm in mind. Unlike games, most non-game contexts have non-negotiable outcomes. It is safest, therefore, to give actual autonomy to one's users. If perceived autonomy is to be employed, consequences should be considered carefully.

2.3 Overview on Theories of Motivation

In addition to the distinction between gamification elements and gamification design principles we discussed above, we have introduced our own distinction between goals and means of gamification in section 4.1. Motivation was only one of a variety of goals of gamification found in that analysis. Engagement was mentioned twice as often and problem solving, user behavior, and value creation as often as motivation. Yet, we will argue that motivation is at the core of what most attempts at gamification intend to achieve. Most approaches to gamification deal with the question of getting users to do something they would not previously have done. K. Duggan & Shoup (2013) even explicitly mention the driving of user behavior as a goal of gamification. Engagement, for example in an online community, is all about getting users to create content (e.g. comments, entries in a wiki, links, or votes) and the steering of user behavior is clearly about enticing users to do things differently. According to Ryan and Deci, “[T]o be motivated means to be moved to do something” (Ryan & Deci, 2000, p.54). They define the state of not being inspired to act as unmotivated, while someone who is motivated is “energized or activated towards an end” (Ryan & Deci, 2000, p.54).

This is essentially the major goal of gamification, to move someone from a state of amotivation (not motivated at all to do something specific) to a state in which he or she is motivated to do so. A classic example from online communities that predates the term gamification would be post counters and ranks based on activity within the community. Many online bulletin boards implement a simple display of the number of posts made by individual users and grant titles to users based on this number. This provides an easy way for veterans to show their status and at the same time is supposed to motivate newer users to post more. A user that was originally content with simply reading what others wrote and not motivated to ever add content of his or her own, could be moved by such a system to start posting as well. Excluding *gameful experience* and possibly *learning* from our list of goals of gamification in section 4.1, all other goals clearly relate back to motivation. This makes an understanding of motivation and the body of knowledge about how human motivation is created and lost essential to anyone dealing with gamification.

The following chapters will briefly introduce different theories of motivation, both on an academic and a more practical level. Ford (1992, p.153) describes 32 theories of motivation, making it impossible to discuss all existing approaches within the scope of thesis. We will instead focus on SDT (Deci & Ryan, 1985b) and theories built upon it for psychological explanations of motivation. SDT's prominence in gamification research makes it a natural choice, although it would certainly be interesting to evaluate other theories of motivation in games and gamification with a proper psychological background. In addition to SDT, we will briefly discuss three additional works on motivation due to their prominence in works on gamification: Seligman's well-being theory, Pink's book on employee motivation, and Fogg's behavior model.

2.3.1 Self-determination Theory

Self-determination theory by Deci & Ryan (1985b) is probably the most often referenced theory on motivation in gamification research. According to the authors, motivation varies in two dimensions. For one, there is the level of motivation — one is not simply motivated or not but one can be more or less motivated to do something. For another, there is the orientation of motivation, describing whether the motivation is based on internal or external reasons (Ryan & Deci, 2000). The two basic types commonly discussed on the latter dimension are extrinsic and intrinsic motivation. The former is based on a separable outcome (such as being paid for work), while the latter arises when something is "inherently interesting or enjoyable" (Ryan & Deci, 2000). Following their 1985 publication of SDT, Ryan and Deci have further broken down extrinsic motivation into subtypes of motivation. Vallerand (2000) proposes a compatible model that discusses motivation on three hierarchical levels — global, contextual, and situational — all on the scale from amotivation through extrinsic motivation to intrinsic motivation. Other authors rely on a similar distinction. Davis et al. (1992, p.1112), for example, describe extrinsic motivation as referring to "the performance of an activity because it is perceived to be instrumental in achieving valued outcomes that are distinct from the activity itself", while "intrinsic motivation refers to the performance of an activity for no apparent reinforcement other than the process of performing the activity per se". SDT has since been applied and evaluated by a variety of authors in a variety of fields. E.g. parenting (Grolnick et al., 1997), language learning (Noels et al., 2000), exercise (Edmunds et al., 2006), medicine (Williams et al., 2004), and, highly relevant to this research, education (Reeve, 2002). We will rely on the original authors' work for our overview of SDT to provide a consistent summary.

An intrinsically motivated task is one in which a person engages without any influences that do not directly relate to the task. Ryan & Deci (2000, p.56) refer to the "ubiquitous readiness [of humans in their basic state] to learn and explore" as an example of intrinsic motivation. To them, intrinsic motivation exists in the connection between the individual and the activity. While a task can be of such nature that it is intrinsically motivating to many people, intrinsic motivation is always based on both the individual and the activity. It would technically be incorrect to speak of an intrinsically motivating activity, as one can really only find activities with a high likelihood of being intrinsically motivating to a person.

Deci and Ryan name two primary measures for intrinsic motivation. One is based on an experimental setup giving participants a free choice to (continue to) do something or not. The assumption being that a participant continuing to perform a task even though they have been specifically told that they do not need to do so anymore (removing, in theory, all extrinsic motivation) is intrinsically motivated.

A second measure comes in the form of self-reports, such as Ryan and Deci's Intrinsic Motivation Inventory.[28] The inventory has been validated in various settings, both by the authors and by third parties (e.g. E. McAuley et al., 1989; Tsigilis & Theodosiou, 2003).[29] It includes seven subscales, one of which is directly related to the measurement of intrinsic motivation, called the interest/enjoyment subscale. This gives us a direct link between intrinsic motivation and enjoyment — or, as it is more often called when discussing games, fun.[30]

[28] http://www.selfdeterminationtheory.org/intrinsic-motivation-inventory/

[29] Markland & Hardy (1997) point out some potential weaknesses of the intrinsic motivation inventory (IMI). They posit that its self-reported nature may lead to incorrect classification of extrinsic motivation as intrinsic — i.e. users may report enjoyment in an activity but actually enjoy the rewards that come from performing it. (Markland & Hardy, 1997, p.30)

[30] A large variety of instruments for measuring motivation in specific areas have been developed based on the IMI. Choi et al. (2010) evaluate a measure for schizophrenia research, for example, and F. Li (1999) one for exercise.

Intrinsic motivation is often considered to be the better, more valuable form of motivation. Ryan and Deci state, for example, that "intrinsic motivation results in high-quality learning and creativity" (Ryan & Deci, 2000). Vallerand goes so far to conclude "that intrinsic motivation produces the most positive consequences, whereas certain types of extrinsic motivation (especially external regulation) and amotivation produce the most negative ones" (Vallerand, 2000, p.314). It is this qualitative difference, that makes the study of SDT so important to those interested in gamification. Games provide a high amount of intrinsic motivation, as they are usually not played for external reasons. Huinzinga's separation of the real world from the world of the game even mandates an exclusion of external influences. While there certainly are potential extrinsic motivations for playing games (as seen in Juul's negotiable consequences), and while one might even leverage that certain elements of games are extrinsically motivating in nature, even though they are included in the game, games are mainly played for enjoyment and therefore gameplay is generally intrinsically motivated.

Many non-game activities, and especially those likely to be gamified, tend to be extrinsically motivated, however. In theory, students should go to school because they enjoy learning. In practice, many students go to school because they have to and learn because they require good grades. Similarly, there are certainly those that do sports for enjoyment, but many others do so to keep their body in shape and keep a certain level of fitness. There are cases in which these extrinsic motivations are not enough to prompt action, which is where gamification comes in. Gamification attempts to raise motivation to a level at which the user will actually perform the desired action.[31] As games are very good at creating intrinsic motivation, the assumption would be that we can use lessons learned from games, or elements of games, to also increase the potential for intrinsic motivation in other activities.

Ryan and Deci name two psychological needs whose fulfillment can lead to intrinsic motivation — competence and autonomy. According to a subset of SDT, cognitive evaluation theory, competence and autonomy are inherently linked in their potential of affording intrinsic motivation. Feeling competent in an activity alone does not lead to intrinsic motivation unless one also has a feeling of autonomy in the activity (Ryan & Deci, 2000). It is important to note here that Ryan and Deci refer to feelings of autonomy and competence,

[31]See also our discussion of Fogg's Behavior Model below.

not actual autonomy and competence. It is irrelevant whether one is actually competent at an activity, it is only important that one feels that way in order to be motivated. Gamification can never directly increase the competence of a user, but it can attempt to make the user feel more competent — through feedback or adaptive difficulties for example.

In a way, extrinsic motivation is the opposite of intrinsic motivation. Where intrinsic motivation is based on an individual's enjoyment of an activity, extrinsic motivation comes from external consequences. Both types of motivation are sides of the same coin, however. To a practitioner implementing gamification into an existing system, it may not matter whether the gain in engagement was intrinsically or extrinsically motivated as long as both methods result in the same gain. There are reasons to prefer one type of motivation over another and we will discuss these below, but one must not forget that one is dealing with motivation (as opposed to amotivation) in both cases.

Deci and Ryan subdivide extrinsic motivation into four categories on a scale from an external PLOC to an internal one (Ryan & Deci, 2000). The PLOC is a description of what one perceives the origin of motivation to be. As such, the PLOC can be either internal ("I wanted this to happen, so I did it"), or external ("Someone else made me do it") (see De Charms, 2013). Extrinsic motivation with a fully external PLOC is called external regulation and associated with extrinsic rewards or punishments (Ryan & Deci, 2000, p.61). Here, the external source of motivation is obvious to the subject. If he or she goes to work solely to be paid for it, the PLOC is fully external. Ryan and Deci also discuss a second type of extrinsic motivation that has an external PLOC, if less so than the previous one: introjection. Introjection is associated with ego involvement or approval (Ryan & Deci, 2000, p.61). E.g. one might go to work because others would consider one lazy if one did not go. The source of motivation is still somewhat external, but it is the subject him- or herself that comes to the conclusion. Pressure from outside is implied, but not actually delivered.

Further down the scale is identification, the first type of motivation with a somewhat internal PLOC. In identification, one consciously values the activity or performs it in order to arrive at goals one has set by oneself (Ryan & Deci, 2000, p.61). Going back to an earlier example, doing sports because they help one stay healthy and having identified health as a valuable goal is being motivated through identification.

The final type of extrinsic motivation is integration, with a fully internal PLOC and difficult to distinguish from intrinsic motivation. Behaviors of this type are part of one's self-definition (Gillison et al., 2009, p.310). According to Ryan and Deci, these behaviors "are still extrinsic because behavior motivated by integrated regulation is done for its presumed instrumental value with respect to some outcome that is separate from the behavior, even though it is volitional and valued by the self." (Ryan & Deci, 2000, p.61)

This differentiation of extrinsic motivations is important to keep in mind when discussing gamification. Often, extrinsic motivation is equated with rewards given to the player (such as points and badges). While that is clearly part of extrinsic motivation, it is only a part of the available spectrum. The further one intends to move the PLOC, the more difficult it becomes to create affordances for these motivations. Creating external regulation is easy, as it only requires rewarding or punishing the player. Integration on the other hand would require complex means that enable the player to identify valuable goals and to integrate them with their own personal goals.

Ryan and Deci give some pointers as to how to promote internalization, that is moving away from a fully external PLOC to a more internal one. A key aspect they underline is that of the psychological need for relatedness, pertaining to the connection to others. Ryan and Deci state that, in the case of an activity that is not intrinsically motivating, "the primary reason people are likely to be willing to do the behaviors is that they are valued by significant others to whom they feel (or would like to feel) connected" (Ryan & Deci, 2000, p.64). The previously discussed needs of autonomy and competence do not only lead to intrinsic motivation, according to the authors they are also required to internalize past the level of introjection.

While we have previously said that extrinsic motivation is not necessarily negative as it is still motivation, one should carefully observe whether it is the right choice of motivation to foster. A controversial topic in motivational psychology is the overjustification effect (Tang & Hall, 1995). Various meta analyses have come to different conclusions regarding its existence, but the latest studies seem to support it (Deci, Ryan, & Koestner, 2001). The overjustification effect refers to the tendency of extrinsic rewards to undermine intrinsic motivation. Many studies have shown (as summarized in the meta analyses mentioned above) a decrease in intrinsic motivation after offering participants extrinsic rewards. Many of the experiments followed the same pattern: participants were split into two groups, one of which was offered a reward for some activity in the first half of the testing session. In the second half, this reward was no longer offered. Participants in the group that never got any reward were often more likely to continue with the task than those who were rewarded earlier. Normally, one would assume the motivation of the rewarded group to be higher while a reward is offered and to drop back to the level of the control group when rewards are no longer available. What researchers found instead, was that the prior availability of rewards somehow reduced the intrinsic motivation to perform the task.

The overjustification effect was only found in studies with tasks that were "interesting" in so far as they afforded intrinsic motivation in the first place. In tasks that were not intrinsically motivating ("boring"), no undermining effect could be found (Deci, Ryan, & Koestner, 2001, p.46). Assuming that these findings are valid, one could argue that gamification aimed at extrinsic motivation is fine if the gamified activity is boring, while one has to be careful if the activity was already intrinsically motivating. One example many people can relate to is the way one's position towards learning changes when starting (and advancing in) school. The inherent desire for learning in humans previously postulated by Ryan and Deci often changes into something completely different when confronted with institutionalized learning. Students stop learning out of interest in learning (intrinsic motivation) and start learning only when extrinsically motivated (e.g. by the pressure of grades). Lepper et al. (1973) have shown the overjustification effect in a school setting. Their experiment showed students that expected a reward to be significantly less motivated intrinsically.

Behavioral economists have studied this effect under the name of "crowding-out effect" (Weibel et al., 2014, p.72). Explanations for its existence include the informational character of rewards (that is, one expects that rewards are only given for tasks that are not inherently enjoyable (see Weibel et al., 2014)) and Lindenberg's goal-framing theory in which extrinsic rewards shift the frame of motivation from a hedonic frame to a gain frame (Weibel et al., 2014, p.76). Lindenberg (2001) also argues, however, that pure use of hedonic frames may not be ideal for a productive environment, but rather a succession of gain and normative frames with a hedonic frame as a background.

SDT provides an excellent background for understanding the motivating effects behind gamification. When studying existing gamification, most implementations fall on the extrinsic side of the scale, usually with a clearly external PLOC. Most attempts at gamification will stop at rewarding users for desired behavior and go no further. Introjection can also be found, especially in cases where leaderboards are implemented or the ability to brag about collected badges to others. Implementations with an internal PLOC are rare and less simple to identify. Through this analysis alone, SDT theory manages to highlight some of the major shortcomings of current approaches to gamification. Where games are mainly intrinsically motivating,[32] gamification does not manage to repeat that and often employs extrinsic motivation instead. While not necessarily bad, no amount of extrinsic motivation can replicate the intrinsic motivation created by games. The idea of taking the best from games to use elsewhere is therefore flawed if one sticks to extrinsic motivation only. SDT introduces three psychological needs that improve internalization and may lead to intrinsic motivation: autonomy, competence, and relatedness. We will revisit these concepts later when discussing motivational theories built upon the work of Ryan and Deci.

2.3.2 Other Commonly Referenced Theories of Motivation

While SDT is the main theory of motivation referenced in gamification literature, there are many others. It is not within the scope of this work to provide a comprehensive discussion thereof; for an overview over 32 theories of motivation from a psychological perspective see (Ford, 1992, p.153). This chapter focuses on three more practical approaches to motivation: Seligman's well-being theory and McGonigals adaption thereof, Pink's work on motivation in the workplace, and Fogg's behavior model. All three have in common that they are frequently used in game studies and gamification literature and that they provide useful extensions to SDT.

[32] It should be noted that not all games are purely intrinsically motivating. One trend that can be observed in modern video games is the addition of progression elements (as well as the by now well established use of badges). Many modern games add an external layer of progression on top of their core gameplay to keep players interested. Especially games with a classic match structure (such as first person shooters) keep adding additional rewards that persist outside the match.

Well-being Theory

In his well-being theory, Seligman introduces five elements that together define well-being. These five elements are positive emotion, engagement, relationships, meaning, and achievement (Seligman, 2011, p.16). While not a theory of motivation, Seligman's work has influenced that of McGonigal, an influential public figure in gamification. She proposes the use of Seligman's five elements (PERMA) as a better alternative to the predominant use of extrinsic rewards in gamification (McGonigal, 2011). She proposes four integral factors for creating happiness (and thereby motivation) "satisfying work", "the experience, or at least hope, of being successful", "social connection", and "meaning" (McGonigal, 2011, p.49).

The connection is not difficult to make — if these elements promote well-being and we can build them into whichever non-game context we are looking to improve, that in turn should improve the well-being of the users and increase their motivation for using our system in the first place. Looking at the five elements in sequence, positive emotion is both the easiest to understand and the hardest to pin down. Positive emotions (Seligman names happiness and life satisfaction) are certainly laudable goals, but tell us little about how to actually build gamified systems. User engagement is a much-studied phenomenon that eventually relates back to motivation. Lehmann et al. decribe it as "the quality of the user experience that emphasizes the positive aspects of the interaction, and in particular the phenomena associated with being captivated by a web application, and so being motivated to use it" (Lehmann et al., 2012). O'Brien et al. name a set of characteristics in their definition of engagement as "a quality of user experiences with technology that is characterized by challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect" (O'Brien & Toms, 2008).

Relationships show an actual overlap with SDT, stressing the importance of others in creating valuable (motivating) activities. Seligman defines meaning as "belonging to and serving something that you believe is bigger than the self" (Seligman, 2011, p.17) and states that it has both subjective and objective components. While not directly represented in the types of motivation in SDT, one could easily consider a feeling of meaning to create at least identification as conscious valuing of an activity. Interestingly, this would make meaning an extrinsic motivation still. Seligman's final element is achievement or accomplishment, which can clearly promote SDT's feeling of competence. Achievement for achievement's sake (as Seligman describes it) could be considered intrinsically motivated. Achievement often comes with extrinsic motivation as well, as one often enjoys to be seen as competent by others. Overall, PERMA was not meant to be a theory of motivation nor a recipe for gamification, but it does include some helpful approaches and recipes to design better systems, be they gamified or not.

(Relatedness) Autonomy Mastery Purpose

Similarly to SDT, Pink brings us three key elements to motivation. Unlike Ryan and Deci, he looks at the question of motivation from a practical, rather than an academic purpose. His three core elements of motivation are autonomy, mastery, and purpose (Pink, 2011, p.87). Some overlap with SDT is immediately obvious: both contain the element of autonomy (and use similar definitions for it) and Pink's mastery can easily be mapped to the competence in SDT. His third element cannot be found in SDT, but in Seligman's well-being theory instead: purpose, or, as Seligman calls it, meaning. The idea that you are performing an activity for a reason larger than yourself. Pink relates purpose to having goals one means to reach with an activity, but goes on to describe that not all goals are equally satisfactory. He distinguishes between *profit goals* — becoming wealthy, achieving fame, and *purpose goals* — helping others to improve their lives, learning, growing (Pink, 2011, p.142). Pink reports about a study by Deci and Ryan on the results of achieving those goals, finding that achieving purpose goals generally increased happiness, while achieving profit goals did not necessarily do so. This leads to the conclusion that including goals with a purpose beyond personal gain into a system could increase motivation as users strive for increased happiness. It is curious, that Pink, while referring often to Deci and Ryan, does not include relatedness in his key elements. When his work is referred to in gamification literature, this is often rectified by expanding the list of key elements of motivation to four: relatedness, autonomy, mastery, and purpose (RAMP) (Marczewski, 2012, pos.1080).

Fogg's Behavior Model

Previously we have been discussing motivation in a vacuum, assuming that increased motivation for an activity would lead to one performing that activity. Obviously, this is an oversimplified approach as it does not allow for different levels of motivation (one does not just need motivation, but enough motivation) nor for other obstacles to following through on motivation. One attempt at modeling human behavior in more detail comes from Fogg (2009) in form of the Fogg Behavior Model (FBM). According to Fogg, three factors lead to the actual performance of an activity (he refers to activities as behaviors): one requires sufficient motivation, one needs to be able to perform the activity, and there needs to be a trigger for that activity.

This model clearly includes the previously discussed need for motivation (with three subcomponents: pleasure/pain, hope/fear, and acceptance/rejection) but introduces two additional factors. For one, one needs to be able to perform the activity. This factor is not as binary as it sounds, rather Fogg refers to how good one is at performing the activity. According to him, ability has six subcomponents, depending on time, money, physical effort, brain cycles, social deviance, and non-routine. Fogg describes a trade-off of motivation and ability, meaning that if an activity is more difficult for an individual to perform, the motivation required to do so needs to be higher. Fogg includes a 2-dimensional scale with ability on one dimension and motivation on the other. The higher the combination of both, the more likely one is to actually perform the activity.

His third element, the trigger, is less well defined. The idea behind it is that, given sufficient motivation and ability, we still need such a trigger to actually start the activity, essentially a reminder that the activity is open to us. A simple example might be that of drinking coffee. If one is tired or full after dinner, the motivation to drink coffee might be higher than otherwise. If one is in a restaurant and has sufficient funds, accessing coffee is very simple and as such one might be more likely to actually drink one than one would be if one had to prepare it first. This represents ability — one must be more motivated to drink coffee if it is more difficult to access. Third, one might not order coffee after dinner even though one is motivated and able to, unless somehow reminded of the possibility. In our

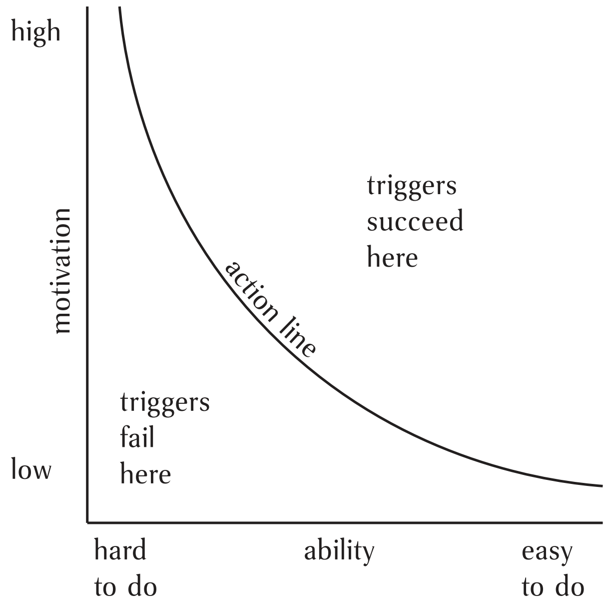


Figure 2.7: Fogg's behavior model. Adapted from (Fogg, 2007)

example, the waiter might approach after clearing the dishes and offer to bring coffee. Fogg distinguishes between three types of triggers — signals, which are essentially reminders that an activity is available, facilitators that highlight the ease of performing an activity, and sparks that give reason for doing so.

2.3.3 Motivations to Play

So far, we have discussed motivation on a more general level. Especially interesting for us, however, are motivations to play games as these are what gamification is trying to leverage. Why do users choose to play games and why do they choose the games they do? This is of obvious interest to game designers as they need to make their games appealing and the concept translates well to gamification. If we employ gamification to cause certain behavior, an understanding of what makes users play games should also help to identify the elements from games to transfer elsewhere in order to cause motivation for that activity. Two scholars in game studies are especially prominent for their work in understanding motivations to play and we will briefly discuss both below.

Bartle (1996) studied the motivations of users to play multi-user dungeons (MUDs), early versions of modern day MMORPGs. His often-referenced paper is based on observation within such games and he often reiterates that the results are limited to these very specific games. Bartle found four types of players within MUDs, categorized according to their motivations to play. He used a two-dimensional model for the categorization. One axis describes whether the player is acting or interacting, the other differentiates between the players and the world. Achievers want to act on the world (e.g. kill a strong boss monster), killers want to act on other players (e.g. kill them), explorers want to interact with the world (e.g. be surprised by it), and socialisers want to interact with other players. Bartle then goes into detail about how these populations interact with each other, but the categorization and the idea that players are playing the same game with different motivations is the major takeaway for our studies.

Yee (2007) examined a similar phenomenon in MMORPGs , but used an inventory of 40 items with 3000 respondents for his analysis. Principle component analysis revealed three major motivations with a total of 10 components that themselves had three or four subcomponents each. The major component of *achievement* was divided into *advancement*, *mechanics*, and *competition*; the *social* component comprised *socializing*, *relationship*, and *teamwork*; and *immersion* comprised *discovery*, *role-playing*, *customization*, and *escapism*. Subcomponents for *advancement*, for example, were *progress*, *power*, *accumulation*, and *status*. Yee's work provides empirical support for Bartle's observation that there is a variety of motivations among players of the same game. Seifert & Jöckel (2009, p.303ff, translation ours) used Yee's instrument in their own work and, through factor analysis, identified four dimensions of game experience — *challenge*, *fellowship*, *thrill*, and *relaxation* — as well as 11 motives for playing MMORPGs (in this case: WoW.) The factors they identified were, in order of explained variance: *achievement/success*, *community*, *competition*, *exploration*, *integration with reality*, *aesthetics* (in the colloquial meaning of the term), *role-playing*, *leadership*, *playing alone*, *game mechanics*,[33] and *escapism*. They concluded that they have shown a plausible connection between motives for playing video games and the game experience. It should be noted that MMORPGs offer a large variety of game experiences and therefore enable a variety of affordances that may lead to players choosing to play the game. Other games are

[33] This factor refers to the interest in understanding the inner functionality of the game, items refering to using external tools for planning, and the interest in knowing a lot about the game's mechanics.

often more linear in the affordances they offer.
These findings (especially Bartle's work) have found widespread adoption in studies on gamification. In many cases (such as in Chou's work discussed above) Bartle's player types are used unmodified, even though they were never intended for use outside of MUDs and despite their lack of empirical foundation.

Marczewski (2015, pos.802ff) has consulted Bartle and presented an adaptation of his player types to gamification. His framework includes six player types, four of which are aligned with the RAMP factors for intrinsic motivation discussed above (relatedness, autonomy, mastery, and purpose): *socialiser*, *free spirit*, *achiever*, and *philanthropist*. The user types of *disruptor* and *player* are aligned with *change* and *reward* respectively. The latter two types are then again subdivided into four types. Both use a *user-system* axis but the player subtypes are also differentiated through an *acting-interacting* axis while the disruptor types are either *black hat* or *white hat* instead. Marczweski summarizes all types in what he calls the gamification user types dodecad. While Marczewski's work is not publicly based in empirical evidence, attempts have been made to validate the it. The research design and number of participants in (Gil et al., 2015) were not sufficient for a proper verdict on the applicability of Marczewski's categories. (Tondello et al., 2016) claim validation, however, the associations between player types and game elements are weak for three of the six proposed types, suggesting room for improvement in the typology. Even so, all three classifications can help us understand, which game elements are best suited for use in gamification. The presence of these very differently motivated types of users also raises a problem that we will discuss further in section 3.3.

Human action is driven by motivation. That motivation can be either extrinsic or intrinsic in nature, with extrinsic motivation generally being easier to achieve and intrinsic motivation being more valuable. Many authors therefore discuss the question of how to improve intrinsic motivation. Four factors that stand out are (perceived) relatedness, autonomy, mastery, and purpose. By increasing either (or all) of theses factors, one should be able to increase intrinsic motivation for an activity. This in turn should increase the likelihood of the user actually performing the activity. According to Fogg, a pure focus on increasing motivation would be inefficient, however, as ability and triggers also matter. Therefore

Figure 2.8: Marczewski's Gamification User Types Hexad. Image taken from (Marczewski, 2015, pos.839)

making an activity easier to perform should also increase its usage, as should providing relevant triggers. Good gamified applications involve all three: A fitness application for example may use badges for (extrinsic) motivation, include videos of exercises to perform (ability), and remind the user through notifications that it is time for exercise (trigger). Gamification usually focuses on raising motivation, but there are also examples of game elements that function as triggers, for example. SDT stresses the importance of perception in its focus on autonomy, competence, and relatedness as the three driving forces behind intrinsic motivation, meaning that a designer of a gamified system does not necessarily need to actually provide for these three needs but rather needs to make users feel autonomous, competent, and related to others. He or she must be careful, however, that users do not feel tricked, causing a backlash against the system. The theories of motivation we introduced here provide us with practical guidelines for creating gamified systems, with properties that seem useful for gamified systems to have, and with a background in front of which analysis of experimental results can take place.

3 Related Work

As we have shown above, the number of empirical studies examining gamification has risen drastically over the past few years (see e.g. Hamari et al., 2014) and they are published in a large variety of outlets. The few meta studies that exist each only cover a small part of those publications. The following chapter contains an examination of the results of those meta studies as well as results from additional empirical studies on the topic. We have concentrated on studies that have the potential to give insight into the effects of gamification.[1] Since gamification in education is a specific focus of this work, we will treat the studies that focus especially on that field in a separate chapter. Section 3.3 summarizes the findings from the reviewed empirical work with a focus on informing our methodological approach.

3.1 Empirical Studies of Gamification

One major meta study on the effects of gamification in general comes from Hamari et al. (2014) and attempts to answer the question whether gamification works or not. The authors come to the conclusion that gamification does indeed work. Out of 24 peer reviewed empirical research papers reviewed, 22 included quantitative results. 2 of those reported purely positive results, 13 are listed as partially positive and 7 as only presenting descriptive statistics. Hamari et al. discuss a variety of caveats, however, that put the results in a less positive light. The authors mention 8 methodological shortcomings in the reviewed works: small sample sizes, lacking use of validated instruments, lack of control groups, multiple

[1] We considered studies whose title or abstract suggested work on the effects of gamification. If the initial impression turned out to be misleading, the studies were excluded. The list of additional studies is not exhaustive (see section 6.1) and merely supports the re-analysis of the studies discussed in the meta studies.

affordances studied at once, use of purely descriptive statistics, very short experiment time frames, lack of clarity, and no use of multi-level measurement models (Hamari et al., 2014). And indeed, reading the papers listed as having positive results can lead to quite a few questions about the validity of the results. Since a study that is methodologically weak cannot provide strong support to the results it presents, we have analyzed the same studies again, in order to identify how many of them actually provide positive results and whether we could accept the claim, that gamification does indeed work. These results have originally been published in (Broer, 2014). We focused on the following four major questions in our repeat analysis and answered them for each of the papers that Hamari et al listed as having positive results:

1. Pertinence: Does the paper actually deal with gamification? Are the instruments employed appropriate to give insight into the effectiveness of gamification?
2. Positive results: Does the paper report significant positive outcomes regarding the use of gamification?
3. Negative results: Does the paper report significant negative outcomes regarding the use of gamification?
4. Unsupported hypotheses: Does the paper report hypotheses that could not be confirmed?

As we have discussed above, we do not consider the use of full fledged games (and anything that is close to the *game* pole of the proposed serious-game dimension in subsection 2.2.1) to be gamification. The use of full games is different enough from the use of individual elements that one cannot abstract from results of one to the results of the other. Studies discussing complete games were therefore marked as non-pertinent to the question at hand. Studies that did not test hypotheses regarding the effectiveness of gamification or used insufficient empirical means to do so were likewise excluded. Statistically significant results that support gamification having a positive effect from the perspective of the person employing gamification were marked as positive results. If the goal of gamifying a system was to increase participation, for example, results that showed significant increases in participation for the experimental group were marked as positive. Likewise, results that supported the opposite of the intent behind the specific instance of gamification were marked as negative. If the paper included (usually positive) hypotheses about the effectiveness of gamification that were not supported by data, those were marked in the fourth category.

Figure 3.1 shows the distribution of our ratings of the results of the 15 papers marked by Hamari et al. as positive or partially positive. Seven had to be marked as non-pertinent. A prominent example is the work by Eickhoff et al. (2012), employing a full-fledged annotation game in their treatment group. While their results are promising, their treatment is better described as a game with a purpose (von Ahn, 2006) rather than gamification. A different, but equally non-pertinent example is (Hamari & Koivisto, 2013). Their study deals with the motivations for the use, not the effects of gamification. In a final example, W. Li et al. (2012) used experimental setups for their groups that diverged so much from each other that it is difficult to pinpoint their results to specific sources. Of the remaining eight papers, three reported significant negative effects in addition to positive ones. Domínguez et al. (2013) report that the experimental group of students did significantly better on certain tasks but also significantly worse on a final examination and on participation. They conclude that "quantitative analysis suggests that cognitive impact of gamification over students is not very significant" (Domínguez et al., 2013). Guin et al. (2012) report that their gamified

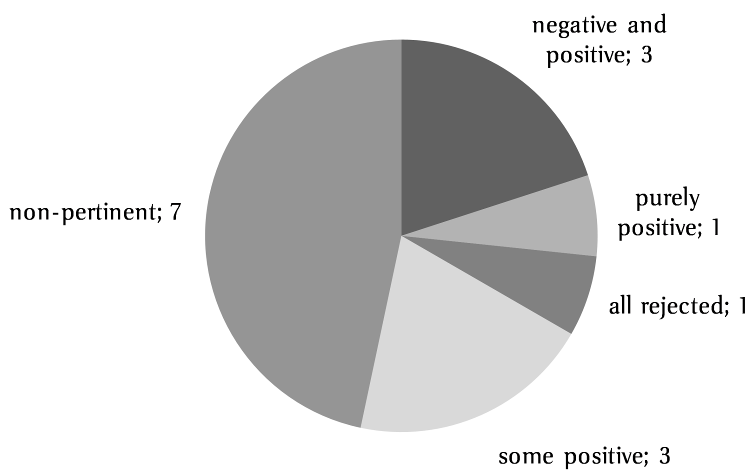


Figure 3.1: Revised rating of paper results. Own illustration.

experimental setup was significantly more interesting than other versions, but they also report a significant decrease in speed and in task completion rates. The study by Hamari (2013) did not confirm any major hypothesis. The only significant correlations found were between user activity and the amount of times the users viewed their own badges or those of others. It seems easy to attribute these correlations to the fact that more active users are more likely to find and use a badge system in the first place.

The second study that showed purely positive results dealt with the removal of gamification from a previously gamified system (Thom et al., 2012). The authors show a significant decrease in activity following the removal. It is unclear whether this decrease means that gamification had a positive effect or whether it was the removal of expected rewards that resulted in reduced motivation. Such an effect (a reduction of intrinsic motivation through extrinsic rewards, often referred to as the overjustification effect) has been extensively studied in psychology (see subsection 2.3.1). Three papers reported a mix of positive results and unconfirmed hypotheses. Denny (2013) saw an increase in student activity, but no significant differences in question authoring, response quality, and perceived learning value when gamifying an online learning tool. Farzan et al. (2008) saw an increase in the amount of content shared, but not in the amount of users sharing content in the long run. Finally, Hakulinen et al. (2013) measured an improved time management in their experimental group, but could not confirm significant differences in individual submission scores, final scores, or carefulness in completing tasks. Overall, evidence of gamification's success is hardly conclusive. 3 out of 15 papers marked as positive in Hamari et al.'s review can be considered net-positive without caveats. All three could only confirm part of their hypotheses.

One of the conclusions of our study was that the studies reviewed by Hamari et al. are not, in fact, sufficient to conclude that gamification works. Some positive effects have been sighted, but many hypotheses could not be confirmed and in most cases the effectiveness of gamification was not actually measured properly. A review by Keusch & Zhang (2015) came to a similar conclusion, reviewing 14 studies of gamified surveys. They found that many of them report positive results in self-reported outcomes (e.g. enjoyment, interest) but few also had valid positive results in actual behavioral outcomes. In many cases, comparisons between gamified and non-gamified behavior were biased due to intrinsic changes to the survey design (e.g. rephrasing of questions.) Some negative outcomes, such as an increase in response time, were also reported. The authors question the usefulness of gamification in survey design and request a better theoretical foundation for such experiments. We will use the following section to examine additional studies that were not part of the previous meta analysis to see, if additional positive results can be found or if the effects of gamification are still largely unknown.

While academics are still discussing the merits of gamification, many practitioners are already sure that it works and either side is critical of the other's views. This apparently changed in August 2014, when Gabe Zichermann posted that gamification was finally scientifically validated (Zichermann, 2014). In his post, Zichermann refers to a study by Mekler et al. Mekler et al. (2015) studied the influence of individual gamification elements on task performance as well as intrinsic motivation and competence need satisfaction in an image annotation task. They found significant positive differences for all three experimental conditions compared to the control condition. When users were given points for completing tasks they completed significantly more tasks than without gamification but significantly less than they completed when levels or leaderboards were used. The authors surveyed a relatively large number of participants (n=273), with roughly 70 participants in each condition. Participants were screened for their orientation using the 12-vignette General Causality Orientations Scale (GCOS; Deci & Ryan, 1985a) and labeled as either control or autonomy oriented. Autonomy oriented participants were found to complete more tasks on average than control oriented ones. Participants' motivations were measured using the intrinsic motivation inventory (IMI) (see e.g. Ryan et al., 1991). No significant differences in intrinsic motivation could be found between the groups. Finally, Mekler et al. measured task quality and found a significant inverse correlation between task quality and quantity but no significant differences in task quality between groups.

Overall they have shown that, for their specific case of image annotation, the common gamification elements of points, levels, and leaderboards could all increase task quantity in the short term without a reduction in task quality and without changes to intrinsic motivation. This is certainly a positive result, regarding the question of gamification's effectiveness and one of the few studies to do so with scientific rigor. One must not make the mistake, however, to abstract from this very limited in scope experiment to all of gamification. It is interesting that no decrease in intrinsic motivation could be measured when extrinsic rewards were introduced, as it supports earlier findings in motivational research. The over-justification effect is not relevant when a task is already of low intrinsic motivation — such as the image annotation task used in the study.

T. Y. Lee et al. (2013) studied the effects of gamification on participation in a crowdsourcing task. Participants (n=437) had to identify Twitter accounts owned by employees of IBM in a longitudinal study over 6 months. In a 3x2 design with a control group the authors investigated the impact of individual achievement, social achievement and a combination thereof; both with and without gamification. Gamification was represented by levels in the individual achievement condition and a leaderboard in the social achievement condition. They found significant increases in average task completion in the gamified conditions with an order of magnitude difference between the non-gamified individual achievement condition and its gamified counterpart. Lee et al. specifically mention that a subset of users in the gamified conditions reacted very well to the stimulus but unfortunately did not provide supporting numbers. While the crowdsourcing task in the experiment is one that is mainly aimed at raw output numbers, other uses for gamification may require a more even distribution of success factors. These results indicate that gamification only works for a subset of users. The reliability of the results is somewhat questionable, as the researchers also found a significant difference between the control group and the non-gamified social achievement group which is difficult to explain and suggests other influencing factors.

Hamari & Koivisto (2015) used a survey among users of a gamified fitness application to perform an analysis of factors that are related to attitude towards the system and its continued use. Employing a model based on technology-acceptance theories and an instrument comprised of various instruments measuring attitude, continuance intentions for systems use, enjoyment, ease of use, playfulness, usefulness, recognition, and social influence. They could show positive associations with attitude for usefulness, recognition, and social influence and positive associations with continued use for usefulness (indirectly through attitude), ease of use, enjoyment, playfulness (indirectly through attitude), and attitude. The authors propose that hedonic factors (e.g. those stemming from gamification) tend to have an influence on continued use through rather subconscious means while utilitarian factors tend to have a conscious influence on attitude towards the system. They conclude that while utilitarian factors can have an influence on use intentions, those are usually moderated through attitude while the hedonic factors have a direct influence on use intentions.

Interestingly, the authors frame ease of use (a utilitarian factor) as a "lack of negative affective experiences" (Hamari & Koivisto, 2015, p.427), behaving similar to enjoyment (a hedonic factor) which is essentially a measure of positive affective experiences. Enjoyment therefore complements ease of use in its measurement of affective experiences, suggesting that a system focusing purely on hedonic or utilitarian aspects will forego some possibilities for improved affective experiences and thereby continued use. Of additional interest is the lack of association between (social) recognition or social influence and continued use. While the authors could show such an association with attitude, that association did not carry on to continued use. This may have an impact on the design of gamified systems as designers often attempt to include social aspects and the lack of effect on continued use would not necessarily show when surveying or interviewing users (as the social factors are associated with attitude, which is what you would likely extract from such studies). While this study does not speak to the effects of gamification (or lack thereof), it does provide an interesting perspective on how gamification can influence continued use and how measurements of attitude may underestimate that influence.

Allam et al. (2015) studied, among other factors, the effect of gamification on a web-based intervention for rheumatoid arthritis patients. The randomized controlled trial found positive effects of gamification on empowerment, physical activity, and health care utilization (that is, a decrease in the latter) when compared to the control group. Social support features were found to have similar results, but significantly fewer visits to the website than the gamification group. All measures (except for visits to the website) were self-reported and the authors acknowledge that this is a limitation to the study. The authors employed a mix of common gamification elements with badges, points, and leaderboards all present. The main effect measured in this study seems to be the effect of gamification on interaction with the website, although the increased interaction did not lead to an increase in knowledge about the subject of rheumatoid arthritis. This is consistent with other findings, such as those by Domínguez et al. (2013) who could also measure increases in activity but not learning outcomes.

Morschheuser et al. (2015) studied the influence of points and leaderboards on bank employees in an information acquisition task. In a between-subject design, participants (n=68) were given a number of articles to read at their own pace and choice to prepare for a meeting, each article followed by a quiz on that article's content. Participants had the option to stop what they were doing and take a coffee break at any time, ending the experiment. Participants in the experimental group were given points for reading and doing well in quizzes and were shown a leaderboard. The authors have shown significant differences in all measured variables — total time spent, number of articles read, time spent per article, number of questions answered and answered correctly, as well as the percentage of questions answered correctly. This suggests a short-term positive effect of a points-and-leaderboards setup on user participation. The authors do not discuss within-group differences but the listed standard deviations are much higher in the experimental group in most cases, suggesting that the employed gamification elements have had a large impact on some users but little to none on others.

Meske et al. (2015) performed a market review of gamification software for leading social collaboration tools. In their investigation of 6 such products (e.g. Chatter by Salesforce, Yammer by Microsoft) the authors found a strong focus on quantitative performance, rewarded with badges and points and shown in leaderboards. Levels, challenges/quests and rewards/feedback were found in some applications and community collaboration in only one. The authors criticize the lack of long-term engagement through gamification and a complete disregard of the concept of flow.

3.2 Gamification in Education

As in the previous chapter on empirical studies of gamification in general, we will begin our examination of gamification in education in particular with a meta study. Dicheva et al. (2015) systematically mapped 34 studies and categorized them. Unlike Hamari et al. (2014), Dicheva et al. did not put a strong focus on the results reported in the studies but rather looked at the educational contexts that gamification has been used in and the game elements employed. They categorized the reviewed papers according to a set of design principles: *goals*, *challenges and quests*, *customization*, *progress*, *feedback*, *competition and cooperation*, *accrual grading*, *visible status*, *access/unlocking content*, *freedom of choice*, *freedom to fail*, *storytelling*, *new identities and/or roles*, *onboarding*, and *time restriction*.[2] The authors note that some of these principles already existed in education before gamification[3] while others have been "borrowed from video games" Dicheva et al. (2015, p.78). They also provide some game mechanics that are commonly used for implementation. Of the design principles identified, some were much more common than others. These are (highest amount of mentions first) *visible status*, *social engagement*, *freedom of choice*, *rapid feedback*, *freedom to fail*, *goals/challenges*, and *storyline/new identities*. All others were mentioned in less than three of the reviewed studies. A similar spread can be seen in the identified game mechanics, with (highest amount of mentions first) *badges*, *leaderboards*, *points*, and *levels* being frequent while the rest was negligible.

The results reported in the 34 studies were mostly positive (18 studies), followed by mixed results (nine studies). Only one study reported negative results, according to Dicheva et al., but seven studies did not report results at all. It is of note, that the papers marked as having positive results included those that relied on self-reported measures and that the quality of experimental design was not evaluated. We have already discussed the work of Denny (2013) in the previous section and did not categorize it as purely positive. We will examine the studies with positive outcomes of quantitative measurements in the following, as well as additional studies that were not included in Dicheva et al.'s analysis.

[2]These categories are not fully congruent with the ones we introduced earlier. See subsection 2.2.3 for an explanation.

[3]We have discussed this in more detail in (Broer, 2015).

Falkner & Falkner (2014) reviewed 10 studies of badges in computer science education and found that "we do not yet have enough evidence to clearly indicate that badges, by themselves, make any contribution to student engagement or effects in forming a learning community" and go so far to say that additional research suggests that badges should not work in computer science education at all.

Anderson et al. (2014) experimented with the use of badges in a massively open online course (MOOC). When comparing a course with badges enabled to two previous courses without badges, they found strong increases in activities that lead to the attainment of badges (i.e. votes in a forum), but not in activities not related thereto (i.e. posts in a forum). Their results come with two important caveats: for one, the experimental setup was not particularly well controlled; other factors than just badges might have changed between the two courses without badges and the one with badges enabled. One factor is the number of participants in the third course, which was almost double that of the previous courses. It is not infeasible for the number of participants to have an effect on individual participants' activity levels. For another, the researchers were working with a heavily tailed distribution in which a small number of participants was responsible for most of the measured activity. As we have seen earlier, gamification seems to have a tendency to strongly motivate a small subset of users and may have done the same here, increasing the activity of the already very active users while not affecting others.

Akpolat & Slany (2014) gamified a university course on software development. They introduced competition between teams and weekly challenges. They found, that students were more likely to focus on the content of the weekly challenge than on other programming practices. They also reported positive self-reported engagement and learning success when compared to other, non-gamified courses. Unfortunately, no control group was employed and it is unclear, whether the non-self-reported results actually speak to the effects of gamification. (e.g. an instruction to focus on a certain topic for a week may have had the same effects as measured here.)

Barata et al. (2013) tracked a number of statistics over five years for a master level multimedia content production course. The course was gamified for two of those five years. Metrics included both student activity (e.g. posts, material downloads) and grades on the final exam. While the authors name a variety of significant differences between years, no consistent improvement could be found between the gamified and the non-gamified years. The amount of variance in the measurements suggest that other factors were influencing the results, such as changes in the course content. Only material downloads and posts show consistent significant differences between gamified and non-gamified years. This suggests that the chosen form of gamification (experience points that directly affect grades, challenges) has increased participation in activities required to attain those points.

In a longitudinal study of the effects of bagdges and leaderboards in classrooms, Hanus & Fox (2015) studied differences in a variety of dimensions between a gamified experimental group and a control group (n=80). They measured the video game habits of the participants following Riddle (2010), intrinsic motivation using the intrinsic motivation inventory (see e.g. Ryan et al., 1991), class satisfaction, class effort, and social comparison using self-developed inventories of five, four, and six Lickert scale items respectively, learner empowerment following Weber et al. (2005), and final exam scores. They could show that the experimental condition significantly reduced intrinsic motivation and class satisfaction over time. They also showed that the students' level of intrinsic motivation mediated the effect of course type (experimental/control) on final exam scores. Their hypothesis that their form of gamification would increase social comparison was partially supported by their results, as was the hypothesis that learner empowerment would be lower in the experimental group. The authors could not show a significant negative effect on learner effort in the experimental group, however.

The authors admit some limitations to their study, mainly the fact that participation in the gamified environment was mandatory which in itself might have lowered intrinsic motivation. Unlike the results reported by Mekler et al. (2013), Hanus and Fox's results support the negative impact of extrinsic rewards on intrinsic motivation suggested by cognitive evaluation theory. Their experimental setup unfortunately does not allow a discrimination between the effects of badges and that of leaderboards. We can therefore not say, if the negative effects are caused by the pressure of social comparison or by the extrinsic rewards.

O'Donovan et al. (2013) gamified a university level course in broad scale, employing a system including virtual rewards such as points and badges, competition in the form of leaderboards, a virtual currency, and strong thematic elements with a story and visuals oriented at the steampunk genre of fiction. The system presented students (n=34) with quizzes and puzzles based on lectures and course material and rewards were based on their results. Evaluation of the system was done without a control group, instead comparing results to both previous year participants and other courses in the program. O'Donovan et al. report that students perceived gamification to increase understanding and engagement and that of the employed gamification elements, students perceived quizzes and lecture attendance (in that order) to contribute most to their learning. The authors further report a significant difference in course grades compared to the previous year, but add that lecturers did change between years and that their chosen form of gamification allowed for some improvement of course grades (O'Donovan et al., 2013, p.249). The lack of a proper control group, the connection of gamification to course results, and the reliance on self-reported perceived advantages of gamification unfortunately led to little empirical evidence for the success of gamification. The most helpful takeaway from O'Donovan et al.'s study is the ranking of different gamification elements according to their perceived contribution to learning. (In order from best to worst: quizzes, lecture attendance, in-class exercises, puzzles, and game story (O'Donovan et al., 2013, p.248).)

Domínguez et al. (2013) and de-Marcos et al. (2014) report results from an experiment in which three similar undergraduate IT courses were equipped with different kinds of supporting tools. One received a gamification plug-in to use in their learning management system (LMS), one received access to a social networking site to use for course purposes, and a control group received neither. As Domínguez et al. report data from the same experiment but exclude the social network group in their data, we will focus on de-Marcos et al. here. All relevant data reported by Domínguez et al. can also be found in the later publication. De-Marcos et al. report significant increases in post-test results on learning objectives when comparing the gamification group with the control group, but also significant decreases in the final examination score and in participation. Interestingly, the social media experimental group scored better, often significantly, than the gamification group on almost all measures. The authors discuss that the practical nature of the interventions led to improved scores in tests requiring practical application, while the traditional learning approach was better suited as preparation for the final multiple-choice exam. De-Marcos et al. further describe their approach to gamification as fostering competition and thereby decreasing participation. The positive results for both interventions on all aspects except for participation and the final exam could be considered an indicator of both methods having merits, but they might also simply be an effect of offering a different type of learning at all (measurement bias). Overall, the results regarding gamification are rather mixed and a potential negative impact of fostering competition is shown. The positive results for the social network intervention group suggest that the social factor should not be ignored in gamification either.

Iosup & Epema (2014) report results from gamifying three consecutive iterations of the same undergraduate course on computer organization as well as one master level course. They employed the usual gamification elements of points, badges, levels, and leaderboards as well as onboarding in form of an entry quiz, social engagement loops in the form of teamwork, and content unlocking. Unfortunately, their evaluation did not make use of a proper control group; the only data for comparison is an older iteration of the same course by different staff. Measurements for comparison focus on successful completion rates. As in the case of O'Donovan et al. (2013), virtual rewards in the gamified course had direct impact on course results, further reducing the validity of the results. Iosup and Epema provide an interesting case study and anecdotal results from employing gamification in a higher education course, but do not provide information about the effects of gamification.

Leong & Luo (2011) introduced a gamified learning platform to an introductory computer science course. They added the usual suspects of points, badges, levels, and leaderboards to their course as well as restructuring assignments into smaller "missions" and adding a discussion system. According to self-reports of students using the system, the main impact of gamification was improved engagement with the course. The splitting of larger tasks into smaller missions, though not necessarily gamification, was also seen positively. Overall, Leong and Luo only bring anecdotal evidence to the question of the effects of gamification. They did not use a control group and report very few actual figures, greatly limiting the validity of their evaluation.

Filsecker & Hickey (2014) studied the effect of extrinsic rewards (in the form of badges) and competition (in the form of a leaderboard) on elementary school students in their use of an educational video game environment. In a quasi-experimental setup the intervention group was awarded badges if certain competencies were reached and was allowed to move a physical representation of their in-game avatar on a leaderboard inside the classroom. The control group simply played the educational game without these additions. The authors could not find a significant difference in self-reported intrinsic motivation of the students nor in their use of feedback pages but did show significant gains in learning as measured by the students' score in a problem solving assessment.

In two similar studies Hew et al. (2016) observed the effects of points, badges and leaderboards on student participation and learning outcomes in an experimental setup (n=22 and n=43, respectively). In both studies, the experimental group showed significantly greater participation (measured in the number of comments posted) as well as significant increases in the quality of hand-ins and the difficulty of self-chosen tasks. The first and the last outcome were directly rewarded with points and indirectly with badges. The authors could not find significant differences in post-test scores for subject knowledge. Further results indicate that the effect of gamification varied strongly among the students, with a majority of students (>60%) only viewing their badges and position and leaderboards two times or less.

In a similar study in a statistics course using Moodle, Huang & Hew (2015) found that gamification in the form of points, badges, and leaderboards led to an increase in participation in a message board for the course, in the viewing of course content, and in extracurricular learning activity (pre and post course tasks) but not in in-class gains. Learning gains from the pre and post course activities were not measured. The study took place in a course with a single three-hour class session. The authors note that a large subset of the treatment group was motivated by the gamification elements employed but that some students were not.

In a large (n=379) , longitudinal (10 weeks) setup, de-Marcos et al. (2016) measured the effects of educational games, gamification and social features on student learning in an undergrade ICT course. They employed four experimental conditions — the educational game "Ribbon Hero" (which has been considered gamification elsewhere, e.g. by W. Li et al. (2012)), a gamification plug-in for the employed LMS Blackboard 9, a social networking platform based on Elgg, the same platform with added gamification (points, badges, leaderboards), and a control group. The authors found that all experimental conditions generally outperformed the control group in intermediary, practical assignments but that the control group significantly outperformed the educational game, gamification, and social conditions in the final examinations. There was no significant difference between the social gamification condition and the control group in the final examination. Among the experimental conditions, no clear leader emerged in the practical assignments, with rankings varying from assignment to assignment.

The authors argue that the results indicate that different forms of support are of varying effectiveness depending on the type of examination. It could also be argued, however, that the experimental conditions motivated student engagement during the course but made no difference (or even negatively impacted) the motivation to prepare for the final exam. In general, the results suggest either a disconnect between the tests for practical skills and those for conceptual knowledge, or ineffective long-term memorization in the experimental conditions. If we assume such a disconnect, it is not difficult to explain the lack of improvement in conceptual knowledge through experimental conditions — they simply did not alter the activity required for a high score. Neither rote learning nor attention in class were particularly supported or motivated in the experimental conditions. Negative results could be explained by a sense of security caused by high scores in practical exams that was not appropriate for the final exam, thereby reducing preparatory effort.

Attali & Arieli-Attali (2015) present two studies on the effect of gamification on test performance. In both studies, the intervention group (or groups) were awarded points for the correct responses to questions with the amount of points given being dependent on the response time. Faster responses would gather more points. The first study employed two different intervention conditions, one in which points were fully dependent on response time while the other gave a significant amount of points for correct responses independent of time and only a small bonus depended on response time. Both experiments showed significant effects of the points condition(s) on response time, with response time being lower (faster) when speed was rewarded with points. The effect sizes were low, however. The two conditions with different points reward schemes did not provide significantly different results. Neither experiment could show a significant difference in correct answers between the conditions. The authors postulate that the effect of the point conditions may have been masked by stronger motivating effects inherent in tests and that the effect of an individual gamification aspect may generally be small and only a combination actually useful.

Goehle & Wagaman (2015) tested the effect of badges in a homework platform on student performance in undergraduate math courses (n=115). They found no significant effect of the badge-enabled condition on the whole group but suggest that there may have been a slight positive effect on under-performing students. They conclude that “the notion that gamification elements like those presented here could have a consistent positive effect on student grades completely independent of the content of the course, the actions of the instructor, or the specific homework problems assigned to students, is probably unrealistic” but also point out that course results are not necessarily the only positive outcome that gamification can have in an educational context.

A study by da Rocha Seixas et al. (2016) used cluster analysis on engagement data collected from an experiment with the use of badges in the classroom (n=61). They found that students with more earned badges also had better results on the engagement indicators but unfortunately the authors did not provide correlation coefficients or measures of significance. They use this data to conclude that “gamification had positive effects on the engagement of students” (da Rocha Seixas et al., 2016, p.60). They do not discuss the possibility that causality might be reversed or that both measures may correlate with a third variable. It is entirely likely, that increased engagement leads to students earning more badges or that better performing students are more engaged and earn more badges. No control group was used in the experiment.

3.3 Summary of Previous Empirical Findings

Both our examinations of empirical studies of gamification began with an existing literature review that summarized the results found in studies to be mostly positive. Our further examination showed that neither studies on gamification in education, nor those on other areas of gamification have provided conclusive proof of such success, however. Most studies either rely on self-reported outcomes, use questionable experimental setups, or report mixed outcomes at best. Some results suggest that gamification can improve user activity, such as posts in forums or attendance to lectures. In the case of education, however, this increased activity does not seem to translate to better learning outcomes, some studies even showing a reduction in final grades for the experimental groups.

The results of Hamari et al. (2014) and Dicheva et al. (2015) show a clear focus on the gamification elements of points, badges, leaderboards, and levels and the additional empirical studies we reviewed showed a similar trend. This is congruent with the results from our analysis of definitions of gamification in section 4.1, where these elements dominated as well. Other elements are employed in a few experiments, but we have almost no empirical results on their effectiveness. Interestingly, social elements are often seen as separate from gamification even if they are employed within the same system. This can be seen in (Allam et al., 2015) and (de-Marcos et al., 2014), for example. Teamwork was implicitly included in (Barata et al., 2013) but not studied separately.

One effect that shows up in a variety of studies is the uneven distribution of the impact of gamification in many experiments, as we have discussed in (Broer, 2014). Hakulinen et al. (2013) report a significant change of behavior in a small group of their users, while most of them behaved similarly independent of their experimental conditions. A few users were greatly motivated by the introduction of badges in an e-learning environment, while most did not seem to be affected. Denny (2013) and Farzan et al. (2008) suggest such differences as well and they are one possible explanation for the effects measured in the heavy tail in (Anderson et al., 2014). While it is obvious that different users would react differently, the fact that there seems to be only a small group of users affected by a specific affordance seems important to both practitioners of gamification and to the interpretation of results of studies. If one is measuring participation, for example, a strong effect on a small group of participants might be lost in statistical variance. Within-subjects experimental setups might be useful in identifying such different types of users. A very rough calculation shows that any individual motivational affordance is likely to affect only a small subset of users. A study among US residents of the ages from 18–44 found that 55% of employed respondents “would be interested in working for a company that offers games as a way to increase productivity”

(Saatchi & Saatchi S, 2011). One conclusion from these data is that about half of a system's users can be expected to react positively to game related affordances.

Affecting half your users might still be a worthwhile goal and would likely give significant results. Unfortunately, it is probable that user bases are partitioned even further. As we have introduced in subsection 2.3.3, results from game studies (see e.g. Bartle, 1996; Yee, 2007) suggest that motivations among players of online games are quite heterogeneous. If one gamifies a system by taking individual game design elements from such games and then implementing them in a non-game context, one will not reach all members of the target audience that are motivated by games, but only those that are motivated by the chosen game elements. Arbitrarily assuming an even split between differently motivated users according to Bartle's player types, any single affordance would only reach about 15 percent of users. Percentages get even worse when implementing affordances that only make use of one of the sub-components mentioned by Yee. Gamification aimed purely at advancement, for example, would then affect an even smaller subset of users. If one considers different types of games as well, it is easy to imagine that only a single-digit percentage of users will be positively affected by any individual gamification effort. It is important to stress again that the above calculation is very rough and based on many arbitrary estimates. Furthermore, Yee (2007) notes that users are not limited to one motivation component. Players may be interested in both achievement and immersion, for example. Nevertheless, one can draw a few conclusions here:

- A narrow gamification strategy should not be expected to achieve large effect sizes across random populations. The implications for research are varied. For one, it would be very interesting to correlate users' game-playing habits and motivations (e.g. using Yee's inventory of items) to their response to gamification. For another, quantitative research designs should be adjusted to the expected effect size. Furthermore, qualitative research might be very useful in understanding the reasons for being affected by certain motivational affordances and not by others.

- The current focus in gamification on badges, levels, leaderboards, and points is too narrow. These speak to the achiever type and ignore other components of motivation.

- It is important to understand the target audience of a gamified system if one wants to gamify successfully. Saatchi & Saatchi S (2011) report major differences in interest in gamification, for example. While only 55% of their employed respondents indicated interest, 68% of males 18–34 responded that way as did 81% of respondents that spent more than $20 on app purchases in the 12 months preceding the study.

In the years since we started this work, empirical results about gamification have grown in number and in quality. While many studies still focus on self-reported measures or fail to properly control for external factors, researchers seem to see the need for large-scale, controlled, long-term studies of the effectiveness of gamification. Especially in education, such conditions are difficult to create as the number of students in an individual course is limited and a comparison across time or different courses introduces factors that are hard to control for. Massively Open Online Courses[4] provide further possibilities for such studies as they allow for a large number of participants and theoretically even randomized controlled experimental setups.

All statistical measurements of the effectiveness of gamification are complicated by the participants' different preferences in motivational affordances. Any effect that is only visible for a fraction of the population is lost when evaluating statistical averages and any experimental setup that employs many different affordances at once will make it difficult to pinpoint the effects of each individual affordance.

Armed with an understanding of empirical studies done in the past, we were able to identify issues in empirical research that our approach can attempt to alleviate and were able to derive a set of commonly used gamification elements as well as some information about the effects of employing those. Together with our theoretical work, this forms the basis for our research in the following chapters.

[4]We have briefly mentioned MOOCs in our summary of Anderson's work earlier in this chapter. A. McAuley et al. (2010, p.4) characterize MOOCs as integrating "the connectivity of social networking, the facilitation of an acknowledged expert in a field of study, and a collection of freely accessible online resources" and stress the importance of a high number (hundreds to thousands) of active students.

4 Analyzing and Categorizing Implementations of Gamification

The first major step towards an increased understanding of gamification is the creation of an instrument that helps to analyze existing systems for their use of gamification. There is, to our knowledge, no empirically validated instrument for the evaluation of gamification in an existing system. Schering (2014) and Chou (2015) both provide frameworks with eight dimensions for the use in such evaluations, but Chou's work is not scientifically evaluated and Schering's has been shown to be applicable and to have good inter-rater reliability, but does not seem to have an empirical basis for the choice of dimensions.

We suggest a different approach to the definition, and therefore analysis, of gamification. The similarity between the variations in definitions of games and those in the definitions of gamification suggest a similar issue: Just as Arjoranta (2014) describes for games, gamified applications are similar to each other, but do not all share the same properties. Arjoranta discusses this issue with nominal and real definitions for the term game and proposes an approach that focuses on Wittgensteinian family resemblances instead. This approach sees games (and, in our case, gamification) as members of a family where individual members are similar to others but not identical and not similar in the same degree. A system *A* (either a game or a gamified system) might share attributes with two systems *B* and *C* whilst *B* and *C* do not resemble each other much. A definition along such an approach would therefore have to be a list of properties that members of the family tend to have some of. If one wants to see if an individual system (say, *D*) is part of the same family, one would have to examine how many of the family's properties it shares (see figure 4.1 for an illustration).

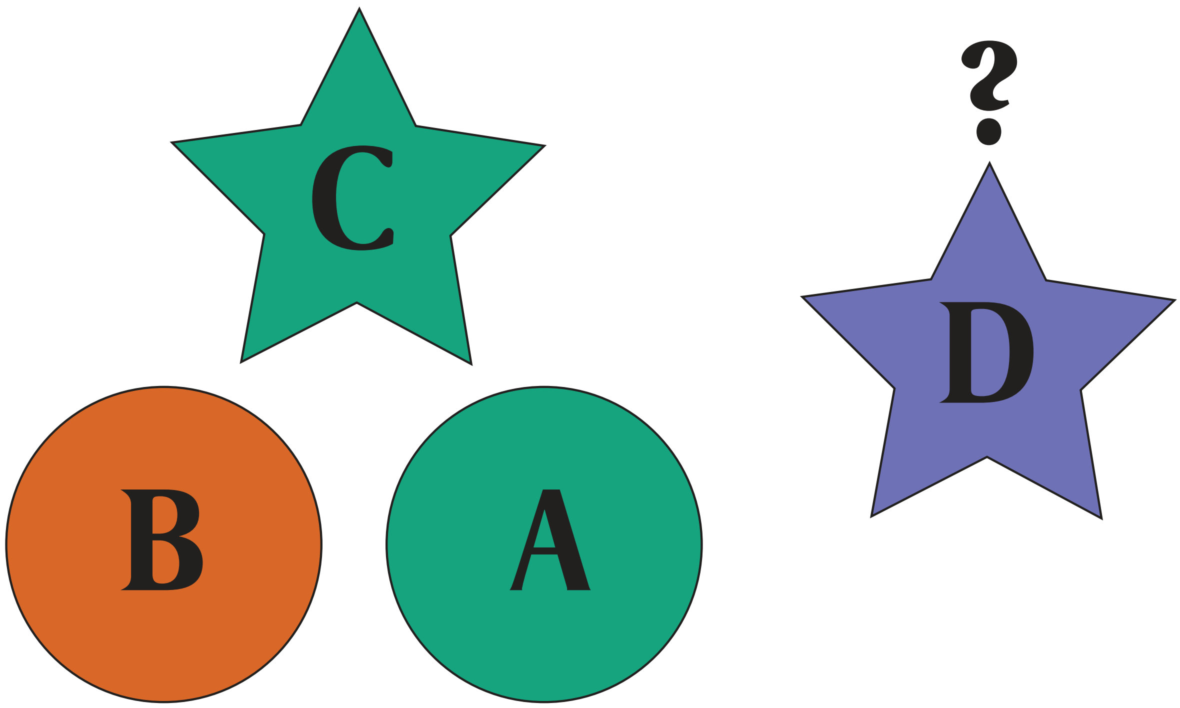


Figure 4.1: Simplified illustration of family resemblances. A, B, and C share properties that make them part of the same family. D is being compared to these properties to see whether it belongs as well. Own illustration.

Such a definition cannot be hard and fast, as it does not make sense to determine a fixed percentage of properties that a member of a family must share, especially because it is likely that not all properties are of equal importance. An examination using a family resemblances approach can therefore never make a definite judgment of family membership but rather only give a degree of resemblance.

In this research we applied a family resemblances approach to the definition (and identification) of gamification. As we have shown, definitions for gamification have many of the same issues that definitions of games have. This is partially based in the origins of the term — if you cannot define games properly, how can you define gamification? — but also in the fact that gamification, just like games, is a field that has naturally grown and has been interpreted differently by different people. Instead of embracing one given definition, or attempting to create yet another synthesized common core definition, we combined definitions and experts' opinions to derive properties shared among many, but not necessarily all, gamified systems.

This chapter describes the identification of those characteristics through literature and an expert survey as well as the resulting qualitative instrument, designed to guide evaluators (be they practitioners or researchers) through their analysis of a given gamified system. In the first step, we analyzed a variety of definitions of gamification for the goals that they claim gamification has, and the means that they say gamification employs to achieve those goals. A similar distinction can be seen in Landers' theory of gamified learning, which we discussed in subsection 2.2.1. In Landers' theory, game characteristics (our means) cause changes in behavior which can in turn cause learning outcomes (one possible goal). In our analysis of the means of gamification, we not only considered the means mentioned in the definitions themselves, as they tended to be very broad in nature, using terms such as *game mechanics*, but also more specific means (gamification elements) mentioned in the respective publications.

The following structured comparison allows for a better understanding of the commonalities and differences between the different authors' understandings of gamification. One important result is the lack of agreement on the means of gamification beyond the top four mentioned elements. With the broad spread of elements used by the individual authors, an identification of properties that signify membership in the family of gamified applications is impossible. Any such attempt would lead to a narrow definition that includes only the most basic forms of gamification that we see employed today.

To solve this, we turned to experts on gamification in our second step, looking for their input on which terms they consider to be relevant for the description of gamification. We used the terms extracted in step 1 as a basis for an expert survey and enriched it with terms from the other two major areas of theory — game studies and motivation. Since the resulting list of terms was not limited to game elements (the means in our analysis) but also included goals as well as broader concepts from game studies and psychology, we asked experts not only to rank the terms according to their importance but also to signify whether they consider them to be relevant to the description of gamification in general, or specific implementations of gamification that they were familiar with. The expectation being that broader concepts such as autonomy are relevant in the former category, while specific game elements (such as time pressure) would only be found in the latter. Given enough input, such a survey could furthermore help to identify types of gamification — clusters of applications within the family of gamified applications that share many of the same properties.

In step 3, we used the results of the expert survey to create an instrument for the qualitative evaluation of gamification in a given system, the Gamification Inventory. The inventory consists of 38 items in five categories, meant to guide experts through the evaluation process. In step 4, we evaluated the instrument by having a set of four evaluators apply it to the analysis of gamification in popular learning management systems and revised it based on that evaluation. Since each step builds upon the results of the previous one, we have included the results directly in this chapter instead of separating methods and results more strictly.

4.1 Step 1: Goals and Means Analysis

In (Broer & Poeppelbuss, 2013) we introduced the idea of defining gamification along the two dimensions of goals that a system (or a modification of a system) is aimed at and the means employed to reach that goal. This distinction was necessary in that research in order to identify written works on gamification that were not labeled as such. In a systematic literature review, we looked for all permutations of means and goals from sets of search terms identified from literature in order to find out, whether gamification was being written about in major information systems publications without using the term. We had previously determined that the term gamification was hardly mentioned at all among those outlets while both common goals of gamification (such as motivation and engagement) and common means of gamification (such as rewards) played a major role in information systems research. The review concluded that the use of gamification was rare among information systems researchers even though both fields have many goals, means, and even underlying theory in common.[1]

The same differentiation of goals and means can be applied to definitions of gamification. As the following examples will show, most definitions we analyzed include both ways to

[1]This has changed since and the term gamification is gaining traction in information systems research as well.

gamify a system (means) and reasons to do so (goals). Of our sample of definitions, only Deterding et al. (2011) attempt to define gamification independently of the goals it is aimed at. Huotari & Hamari (2012, 2016) criticize Deterding et al.'s definition as focusing on only the systemic but not the experiential aspects of gamification. The systemic perspective deals with the way a system is built, while the experiential perspective looks at the experience humans are getting from interacting with the system. In other words, a systemic view of gamification discusses the means with which systems are gamified (such as the addition of badges) while an experiential view only requires that the experience of the user becomes more gameful, no matter how that is accomplished. The difference between systemic and experiential perspectives can therefore also be described as the difference between a focus on the means of gamification and a focus on the goals of gamification. Most definitions discussed before include both means and goals, that of Deterding et al. being a notable exception.[2]

This distinction can also be found in game studies literature, especially the widely cited MDA framework (Hunicke et al., 2004). The authors name three distinct components of games — rules, system, and fun, and describe their "design counterparts" — mechanics, dynamics, and aesthetics (Hunicke et al., 2004, p.2). Essentially, mechanics are individual components the game is built of, dynamics the way they interact with each other and the player's input, and aesthetics the emotional responses of the player. They suggest using each of these elements as a lens to view games through, although there is a causal link (Hunicke et al., 2004, p.2). This corresponds nicely with the two perspectives discussed above. A systemic perspective of gamification looks at the individual elements of a gamified system, while an experiential perspective looks at the user's response. While dynamics are certainly relevant, we know of no research with a focus on such a perspective. One might argue that it is futile to research interactions without knowing individual elements first.

Interestingly, this conflict can often be seen in discussions as to whether a certain system can be considered to have been gamified. In a systemic approach, such as that of Deterding et al., it is relatively easy to make an objective decision as to whether something

[2]Huotari and Hamari point out that Deterding et al. do discuss experiential aspects in their paper but chose not to include those in their definition.

can be considered gamification. While there are still numerous gray areas (often founded in the lack of a clear definition of what constitutes a *game design element* (Deterding et al., 2011)), systemic definitions provide a set of objective attributes for identifying gamification. Experiential definitions, on the other hand, are by nature subjective. Both the individual using a system and the circumstances the system is used in can influence the experiences had through the use of the system. One user may have a gameful experience in a system that is nothing like a game to others. We will discuss the usefulness of such definitions in a later part of this chapter.

The following sections each introduce a definition of gamification and identify goals and means mentioned therein. In many cases, the authors included more detailed lists of means (often called *game elements* or *motivational affordances*) with their definition. In such cases, we have included those in the analysis as well. This concept of means being used to achieve goals is rather straightforward and deterministic. One employs a mean *A* in order to achieve goal *B*. All goals we found involve the user of the system, however, and users are not deterministic. The effect on the user experience does not only depend on properties of a system but also on the user him or herself. Huotari & Hamari (2012) address this with their focus on experiential factors. A system designer can therefore never determine the eventual outcome of a system. What he or she can do instead, is to include affordances for specific outcomes. “The term affordance refers to the actionable properties between an object and an actor.” (Zhang, 2008) In other words an object (in our case: a system) can have properties that enable an actor (the user) to do something. A common example is that of a doorknob that affords pulling the door open. Zhang combines the concept of affordances with motivational theory, introducing motivational affordances, “compris[ing] the properties of an object that determine whether and how it can support one’s motivational needs” (Zhang, 2008). A motivational affordance, in our case, therefore refers to a property of a system that enables users to be motivated. When discussing means of gamification, we are really discussing affordances: properties on a systemic level that may lead to certain experiential outcomes. Deterding (2011) has described this concept in further detail.

Table 4.1: Levels of abstractions in game design. Adapted from (Deterding et al., 2011).

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

4.1.1 Analysis of Definitions of Gamification

> Gamification is *"the use of game design elements in non-game contexts"*
>
> — (Deterding et al., 2011)

The first major attempt at defining gamification academically comes from Deterding et al. (2011). They describe gamification as "the use of game design elements in non-game context" (Deterding et al., 2011), stressing the importance of various terms in their definition. First of all, they discuss the meaning of the term "game", referencing Caillois (1979) differentiation of paidia (free-form play) and ludus (play bound by rules and aimed at a certain goal). Gamefulness (and therefore gamification) in Deterding et al.'s definition has clear ludic qualities, as opposed to playfulness on the side of paidia.

The second dimension they discuss is that of using elements as opposed to a full-fledged game. Deterding et al. therefore make a distinction between serious games — complete games that have a purpose other than entertainment — and gamification in which only elements of games are used. The authors further state that these elements cannot be ones that are exclusive to games nor ones that are common to all games. Instead, they suggest that gamification employs elements “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).

A further key distinction in this definition is the focus on *game design* and not “game-based technologies or practices of the wider game ecology” (Deterding et al., 2011, p.12). The authors differentiate between five different levels of abstraction in game design elements (see table 4.1). These levels of abstraction make it somewhat difficult to compare Deterding et al.’s game elements to those of the other authors discussed below. For the sake of comparison, we will only use the examples given for the first two levels of abstraction: Badges, leaderboards, levels, time constraints, limited resources, and turns. It should be noted that their other levels of abstraction include some examples of game design elements mentioned by other authors discussed here (e.g. clear goals). Unfortunately it is difficult to include these in the comparison without also including elements that are on a completely different scope than those mentioned by the other authors. The existence of different levels of abstraction shows in the definitions of others as well, but is not addressed.

The final important element of the definition is the *non-game context* that gamification takes place in, adding additional game elements to an existing game therefore does not constitute gamification, according to Deterding et al.[3]

> Gamification is *"the process of game-thinking and game mechanics to engage users and solve problems"*
>
> — (Zichermann & Cunningham, 2011, p.xiv)

Zichermann & Cunningham define gamification as “the process of game-thinking and game mechanics to engage users and solve problems” (Zichermann & Cunningham, 2011, p.xiv) . On one hand, Zichermann and Cunningham’s definition is rather broad. The term *game thinking*, aside from lacking a proper definition itself, allows for basically anything game-related to be used in gamification. On the other hand, their definition

[3]We have seen implementations of typical gamification elements on a meta level in game contexts for about a decade, starting before the term of gamification was even coined. All major platforms for the distribution of games (Microsoft’s Gamerscore, Valve’s Steam, Apple’s Game Center, to name a few) allow the user to collect badges from a variety of games. Valve even added an additional gaming element to the meta level with its addition of trading cards. The definition discussed here excludes these approaches from being gamification, but one would be remiss to not at least consider them as examples to learn from.

prescribes a purpose for gamification – engaging users and solving problems. The solving of problems is obviously not a very useful description, but user engagement is a common theme found in many texts on and implementations of gamification and was indeed one of the major goals for gamification we used in (Broer & Poeppelbuss, 2013). If one ignores the generic “solving problems” part of the definition, the declaration of a distinct purpose will limit the scope of the term *gamification*. Essentially, Zichermann and Cunningham put a larger focus on the goal one hopes to achieve when gamifying a system than on the means used to do so. Zichermann and Cunningham provide a set of seven game mechanics often used for gamification: points, levels, leaderboards, badges, challenges/quests, onboarding, and engagement loops. (Zichermann & Cunningham, 2011, p.36ff)

> Gamification is *"the process of engaging audiences by leveraging the best of loyalty programs, game design, and behavioral economics"*
>
> — (Zichermann & Linder, 2013, p.xii)

Zichermann & Linder describe gamification as “the process of engaging audiences by leveraging the best of loyalty programs, game design, and behavioral economics” (Zichermann & Linder, 2013, p.xii). This definition goes even further into the direction of emphasizing the goals instead of the means by putting the purpose of gamification first in the definition and broadening the potential tools used beyond the scope of games. While both loyalty programs and behavioral economics are often referenced in the context of gamification, most other definitions consider the origin in games (or game design) to be essential for something to be called gamification. Zichermann and Linder also drop the “solving problems” element of the Zichermann and Cunningham’s definition and focus solely on user (audience) engagement.

Zichermann and Linder name five game mechanics used in gamification: points, badges (achievements), levels, leaderboards, and rewards (Zichermann & Linder, 2013, p.18ff). Note that not only are three mechanics from Zichermann and Cunningham excluded here, Zichermann and Linder also specifically mention rewards. This definition, more than any other discussed here, shows the focus of practitioners on the value of what they are proposing as opposed to an analytical interest in the topic. Their definition is more useful for making readers understand the idea behind gamification than to provide a clear distinction between it and related concepts as Locke requires.

> Gamification is *"the use of game mechanics and rewards in a non-game setting to increase user engagement and drive desired user behaviors"*
>
> — (K. Duggan & Shoup, 2013, p.10)

Duggan & Shoup define gamification as “the use of game mechanics and rewards in a non-game setting to increase user engagement and drive desired user behaviors” (K. Duggan & Shoup, 2013, p.10). Duggan and Shoup’s definition — while very similar to that of Zichermann and Cunningham – notably mentions rewards as separate from game mechanics. Most likely this is simply done for emphasis, but one could also understand it as an attempt to declare rewards as disjunct from games as such. If one were to follow Juul’s classic game model, tangible rewards might be regarded as non-negotiable consequences (Juul, 2005, p.36) and therefore not typical to games.

Duggan and Shoup also mention the goal of increasing user engagement but add “driv[ing] of desired user behaviors” as well. The authors thereby go beyond the simple idea of motivating users to interact more with a system and toward making users do the right (as in: intended by the system’s designer) things. It seems that one can identify a trend towards goal oriented definitions in texts aimed at practitioners rather than scholars. Duggan and Shoup mention nine tools (game mechanics) that can be used for gamification: points, leaderboards, levels, missions, challenges, quests, achievements, rewards, and feedback (K. Duggan & Shoup, 2013, p.19).

> Gamification is *"a process of enhancing a service with affordances for gameful experiences in order to support user’s overall value creation"*
>
> — (Huotari & Hamari, 2012)

Huotari & Hamari define gamification as “a process of enhancing a service with affordances for gameful experiences in order to support user’s overall value creation” Huotari & Hamari (2012). It should be noted that the authors concentrate specifically on the perspective of service marketing and stress the importance of their distinctions for marketing in various places throughout their work. Similar to the practitioners discussed above, Huotari and Hamari focus on the goal of gamification – in this case supporting value creation – rather than on the means of doing so. They argue that Deterding et al.'s perspective is purely systemic and does not take into account that the same affordance might lead to a gameful experience for one user, but not another. It is noteworthy that Deterding did previously discuss the idea of situated motivational affordances (Deterding, 2011) as an addition to – not a contradiction of – Deterding et al.'s previous definition. While the authors disagree on what a definition should focus on, they seem to share a common understanding regarding the importance of creating a gameful experience.

Huotari and Hamari make an interesting distinction between services that are gamified and services that are only used to enhance the gamified process. Foursquare[4] is a common example in literature on gamification that would not be considered a gamified service according to Huotari and Hamari. Foursquare, according to the authors, is rather a service that enhances a core service (such as a café) and in doing so, can create a gamified service. In this context it is important to note that Huotari and Hamari explicitely state that the core service being gamified can be a game itself, thereby creating a meta-game. This is in contrast to Deterding et al.'s definition which specifically mentions non-game contexts as a requisite for gamification.

Huotari & Hamari (2012) do not offer a list of means of gamification nor a categorization of gamified systems. In a later work, Hamari et al. performed a systematic literature analysis of empirical studies on gamification, however, and listed the motivational affordances used in each of those. The most common variants they found were: points, leaderboards, achievements/badges, levels, story/theme, clear goals, feedback, rewards, progress, and challenge (Hamari et al., 2014). For the purpose of this analysis we will use the definition proposed in (Huotari & Hamari, 2012) and referenced in (Hamari et al., 2014) and the motivational affordances described in (Hamari et al., 2014). This will provide a more robust basis for the analysis and should not lead to contradictions as both papers share a definition (not to mention an author.) Unlike Huotari & Hamari (2012), Hamari et al. (2014) do not focus purely on a marketing perspective.

[4]www.foursquare.com

Table 4.2: Game design elements and motives following Blohm & Leimeister (2013).

game-design elements		motives
game mechanics	game dynamics	
documentation of behavior	exploration	intellectual curiosity
scoring systems, badges, trophies	collection	achievement
rankings	competition	social recognition
ranks, levels, reputation points	acquisition of status	
group tasks	collaboration	social exchange
time pressure, tasks, quests	challenge	cognitive stimulation
avatars, virtual worlds, virtual trade	Development/organization	self-determination

According to Blohm & Leimeister (2013), gamification is "enriching products, services, and information systems with game design elements in order to positively influence motivation, productivity, and behavior of users". Blohm and Leimeister attempt to combine the definitions of Deterding et al. and Huotari and Hamari while replacing Deterding et al.'s *non-game context* with *products*, *services*, and *information systems* and Huotari and Hamari's *value creation* with *motivation*, *productivity*, and *behavior*. Their replacement of Deterding's *non-game context* with the three somewhat specific terms mentioned above is easily understood when looking at the authors' background and the field their paper was published in — information systems.

> Gamification is *"enriching products, services, and information systems with game design elements in order to positively influence motivation, productivity, and behavior of users"*
>
> — (Blohm & Leimeister, 2013)

Blohm and Leimeister refer to neither gamefulness nor motivational affordances in their paper. They do provide a set of game mechanics that they link to game dynamics and motives. Those mechanics are: documentation of behavior, scoring systems/badges/trophies, rankings, ranks/levels/reputation points, group tasks, time pressure/tasks/quests, and avatars /virtual worlds/virtual trade. The connection of those mechanics to what they call *game dynamics* and *motives* is interesting, but – similar to Deterding's levels of abstraction – it would make for a difficult comparison later on. Notably, one of their game dynamics is challenge, which others mentioned as a game mechanic or motivational affordance. Blohm and Leimeister do not really explain those three categories, but one could assume that game mechanics and game dynamics are categories from the MDA framework of game design (Hunicke et al., 2004). If that is so, the third category of the MDA framework — aesthetics — was left out of their model on purpose. That would suggest that Blohm and Leimeister explicitly exclude game aesthetics from the definition of gamification.

> Gamification is *"using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems"*
>
> — (Kapp, 2012, p.10)

Kapp defines gamification as "using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems" (Kapp, 2012, p.10). Focusing on learning and instruction, Kapp attempts a synthesis of various definitions of gamification. One can see clear similarities to Zichermann and Cunningham's definition in both the means for gamification (game[-based] mechanic, game thinking) and the goals it is used for (engagement, problem solving). Kapp adds some aspects not previously discussed here as well: The promotion of learning is a rather obvious desired outcome when discussing the use of gamification for learning, the use of game-based aesthetics, however, is something that has gotten little attention elsewhere. As noted in the previous section, one could construe Blohm and Leimeister's definition to exclude aesthetics. Similarly, Deterding et al.'s focus on game design mechanics equally excludes the use of game aesthetics. Kapp, however, considers aesthetics essential for successful gamification, but does not specify that these have to be taken from or inspired by games.

Also somewhat contrary to the previous authors' opinions, Kapp mentions that "[game] mechanics alone are insufficient to turn a boring experience into a game-like engaging experience" (Kapp, 2012, p.11). This is in stark contrast to the idea of game elements as motivational affordances that promote gameful experiences. Kapp nevertheless lists a set of game mechanics used for gamification: levels, badges, point systems, scores, and time constraints. He also describes a gamified system as including "an abstract challenge, defined by rules, interactivity, and feedback" (Kapp, 2012, p.10f), building on existing definitions for the term *game* itself.

4.1.2 Comparison of Definitions

We have briefly discussed seven definitions for gamification and extracted goals and means for gamification mentioned therein. We further split the means into two categories — broad descriptions of means as used in the definitions themselves and specific means such as game elements or motivational affordances mentioned in the further discussion of those definitions. Out of the definitions discussed here, only Deterding et al. did not include a reference to both means and goals directly in the definition. Deterding et al. nevertheless provided a list of game elements for use in further analysis. As mentioned previously, Deterding's simultaneous publication of a paper on motivational affordances (Deterding, 2011) makes it unlikely that the authors did not consider the goals of gamification to be important, they simply did not consider them important to the question as to whether something can be considered gamification or not. There is therefore little argument that the goals of gamifying the system (and by extension the users of that system) play an important role in designing and examining a gamified system. It is, however, questionable whether it should be part of a general definition of the term. While most authors stress the importance of goals, none mention why only specific goals are valid uses of gamification.

Goals of Gamification

In the definitions discussed here, the actual goals mentioned vary significantly. Table 4.3 shows the distribution of goals mentioned in the six definitions that mentioned goals at all. Similar terms have been collapsed into one (such as *productivity* and *value creation*) for simplicity's sake, even if their meanings are slightly different. None of the goals can be found in all definitions and no definition includes all of the goals. The most frequently mentioned goal is *engagement*, appearing in four of the six definitions, followed by *motivation*, *problem solving*, *user behavior*, and *value creation* with two mentions each. The terms *learning* and *gameful experience* are only mentioned once. Due to the non-exhaustive selection of definitions, one should not put too much weight into these frequencies. Nevertheless, engagement seems to be an important goal to most authors and the only one to command a majority in this sample. This exemplifies a problem with including specific goals within the

Table 4.3: Goals of gamification as mentioned in each definition. Similar terms have been aggregated for better comparison (such as productivity and value creation), ordered by frequency of occurrence.

	ZC[a]	ZL[b]	DS[c]	HH[d]	BL[e]	KA[f]
engagement	x	x	x			x
motivation					x	x
problem solving	x					x
user behavior			x		x	
value creation				x	x	
gameful experience				x		
learning						x

[a] Zichermann & Cunningham (2011) [b] Zichermann & Linder (2013)
[c] K. Duggan & Shoup (2013) [d] Huotari & Hamari (2012) [e] Blohm & Leimeister (2013) [f] Kapp (2012)

definition of gamification: Certainly engagement can be instrumental to the value creation mentioned by Huotari & Hamari (2012) but might not be the only influencing factor. The divergent use of goals seems to be related to the field that the authors work in. Marketing-oriented authors (Zichermann & Cunningham, 2011; Zichermann & Linder, 2013; K. Duggan & Shoup, 2013) seem to favor engagement, learning-focused Kapp (2012) is the only one to mention learning, and information systems/economics oriented authors Blohm & Leimeister (2013) and Huotari & Hamari (2012) focus on value creation.

While it is logical that authors would adapt definitions to suit the needs of their field of work, such adaptations are unsuitable to an attempt of finding an overall definition for the term *gamification*. If goals are dependent on the field observed they therefore cannot be part of an overall definition of gamification. That would exclude the use of gamification in any field that was not explicitly considered in the definition. If goals are to be included in an overall definition, they would have to be general and broad. *Engagement* has the potential to be such a term, as both Huotari and Hamari's *value creation* and Blohm and Leimeister's *motivation* could be linked to engagement in some form or another. Even within the definitions discussed here, however, one can already find examples in which engagement is not necessarily the goal of gamification: Duggan and Shoup mention user engagement but also the steering of user behavior. It is not difficult to imagine a scenario where users are already engaged and gamification is introduced to simply steer their behavior in certain directions or provide other benefits.

Huotari and Hamari's definition is special insofar as it could be considered to consist only of goals, without specifying means at all. They acknowledge any affordance that leads to a gameful experience (which in turn supports value creation) a viable mean for gamification (Huotari & Hamari, 2012). According to them, the creation of a gameful experience is a necessary intermediate goal for gamification that supports value creation. No other authors mention such an intermediate goal. Overall it seems difficult to identify specific goals to include in a definition for gamification. It is important to always keep goals in mind when creating gamification and it seems similarly important to understand the goals of any gamified system one analyzes, but this preliminary research does not show great promise for the inclusion of goals in a definition. Goals could, however, be very helpful when attempting to categorize different forms of gamification if they are kept broad enough.

Means of gamification

All but one of the definitions analyzed here provide a set of general means for gamification. Deterding et al.'s original definition even focuses only on the way something is gamified – through the use of *game design elements* in their case. The same terminology is employed by (Blohm & Leimeister, 2013), but they add goals as well. Another term used by three definitions seems very similar at first glance: *game mechanics*. Deterding et al. explicitly state the importance of the term *design* in their definition — the use of other game elements does not constitute gamification in their opinion. The three authors mentioning *game mechanics* leave the definition of that term rather open, allowing for a broader definition than that of Deterding et al. Even broader is the term *game thinking* employed by Zichermann and Cunningham as well as by Kapp. Neither source further specifies what game thinking actually means, but one can assume that this term is a broad envelope for anything pertaining to games. This could be compared to Huotari and Hamari's approach of considering everything that creates a gameful experience to be gamification — maybe game thinking is supposed to comprise anything that does so. Kapp is the only one to mention aesthetics in

Table 4.4: Means of gamification as mentioned in each definition. Similar terms have been aggregated for better comparison, ordered by frequency of occurrence.

	ZC[a]	ZL[b]	DS[c]	BL[d]	KA[e]	DE[f]
game mechanics	x		x		x	
game design elements				x		x
game thinking	x				x	
aesthetics					x	
behavioral economics		x				
game design		x				
loyalty programs		x				

[a] Zichermann & Cunningham (2011) [b] Zichermann & Linder (2013)
[c] K. Duggan & Shoup (2013) [d] Blohm & Leimeister (2013) [e] Kapp (2012)
[f] Deterding et al. (2011)

his definition, while others often explicitly or implicitly exclude it. If one analyzes common examples of gamification, one will see that game aesthetics are often used in combination with other game elements. It remains to be seen whether aesthetics alone can gamify a system.

It should be noted here, that the term *aesthetics* is used in two different ways in the discussion of games and gamification. Kapp refers to aesthetics as “art, beauty, and visual elements” (Kapp, 2012, p.46), which is close to a colloquial understanding of the term. Often, people discussing aesthetics in game studies refer back to the MDA framework (Hunicke et al., 2004), however, which describes aesthetics as an integral part of game design (alongside mechanics and dynamics) and very experiential in nature. They name sensation, fantasy, narrative, challenge, fellowship, discovery, expression, and submission as examples thereof. While art and beauty can certainly be part of these experiences, Hunicke et al. subsume much more under the term. MDA’s prominence within game studies can make the colloquial use of the term confusing.

Zichermann and Linder use the very broad description of “the best of loyalty programs, game design, and behavioral economics” (Zichermann & Linder, 2013) in their definition, allowing for basically any mean to be used for gamification — as long as it stems from one of those three areas. While less useful as a definition due to its broadness, it does show that gamification is not simply based in games, but has important foundations in more established fields of study as well. Table 4.4 shows the distribution of mentions of general means for gamification among the analyzed definitions. The distribution is even more disjunct than that of the goals for gamification, with no term commanding a majority. Game mechanics are most frequently mentioned, followed by game design elements and game thinking. All authors except for Blohm and Leimeister employ at least one of those terms. One could consider their use of "the best of game design" to be somewhat similar, however. All authors that mention means (meaning all but Huotari and Hamari) thereby agree that gamification requires the extraction of something from games (design elements, mechanics).

Specific Means of Gamification

We managed to extract a list of specific means for gamification from all seven works analyzed here (referring to (Hamari et al., 2014) instead of (Huotari & Hamari, 2012)). In many cases these are only lists of common examples, however. Please note, therefore, that the means in table 4.5 are not necessarily exhaustive. The means in the table are ordered by frequency of mention, showing an interesting trend. The table reveals a few means that are commonly referenced in most texts, followed by a steep drop in the frequency of mentions. All authors name badges and levels as common means for gamification and all but one author mention leaderboards and points, respectively. Even though the seven texts cannot be considered a representative sample of literature on gamification, we can see a very clear focus on these four common elements. The seven texts have been shown to be very heterogeneous in their understanding of gamification before, giving additional credit to their homogeneity in this point. These four elements therefore have to be considered to be at the core of what is commonly understood as gamification. All four follow the premise of giving users extrinsic virtual rewards for their participation in the gamified system but are not identical.

Points, leaderboards, and badges are also the elements most commonly found in a meta analysis by Hamari et al. (2014). Levels, however, are a lot less common in their analysis and at the same level as *story/theme* and *feedback* and even less frequent than *challenge*,

Table 4.5: Detailed means for gamification as mentioned in each of the referenced publications, ordered by frequency of occurrence.

	ZC[a]	ZL[b]	DS[c]	HA[d]	BL[e]	KA[f]	DE[g]
badges	x	x	x	x	x	x	x
levels	x	x	x	x	x	x	x
leaderboards	x	x	x	x	x		x
points	x	x	x	x	x	x	
quests	x		x		x		
feedback			x	x		x	
rewards		x	x	x			
time pressure					x	x	x
challenge				x		x	
avatars					x		
clear goals				x			
doc. of behavior					x		
engagement loops	x						
group tasks				x	x		
interactivity						x	
limited resources							x
onboarding	x						
progress				x			
reputation points					x		
rules						x	
story				x			
turns							x
virtual trade					x		

[a] Zichermann & Cunningham (2011) [b] Zichermann & Linder (2013)
[c] K. Duggan & Shoup (2013) [d] Hamari et al. (2014) [e] Blohm & Leimeister (2013)
[f] Kapp (2012) [g] Deterding et al. (2011)

which was only mentioned by two authors in our analysis. Dicheva et al.'s meta study of gamification in education shows a similar trend. Points, badges, and leaderboards are found most often, followed by levels (Dicheva et al., 2015). Anecdotally, these four elements seem to be the go-to choice for the early stages of gamification, showing up in almost any example, be it research or marketing-oriented, as we have shown in our examination of existing empirical studies. Gamification has even been referred to as "pointsification" by opponents of the concept in the past (Robertson, 2010). Table 4.5 shows quite a few other elements as well, however, which are only mentioned by a few authors or even only in a single source. Possible reasons for this divergence are many. For one, the terms are not completely disjunct. Time pressure, for example, always includes some form of challenge and quests usually provide clear goals. Any further summary of the terms by us is likely to lead to misleading conclusions about the authors' opinions, however. The expert survey and interviews in the following chapters will be used to attempt additional summarization of the terms. For another, the understanding of what gamification actually is seems to differ from author to author or at least from discipline to discipline. There are a number of tools acknowledged by the majority, but after that every author seems to have his or her own ideas of how else one could gamify systems. There also seems to be a large amount of possibilities for gamification and it is unlikely that any one list could cover them all.

Another problem with a list such as the one in table 4.5 is the inequality of terms used. Some (such as badges or reputation points) are very specific game elements that do not vary greatly in their implementation while others (such as interactivity or onboarding) are rather broad concepts and it can even be difficult to confirm their presence, let alone to pinpoint a specific implementation. Deterding et al. addressed a part of this problem with their levels of abstraction. The presence of elements of the same list showing up in very different layers of Deterding et al.'s categorization of game design elements clearly shows this problem. In order to do a proper comparison along that line, one would have to break down (or summarize, respectively) each term so that all the terms in the comparison would fit onto the same level of abstraction.

We have introduced the distinction between goals and means of gamification and analyzed how a variety of authors use these in their definitions of the term. We have shown a clear focus on a small subset of means common among all authors but also identified a wide variety of other gals and means that individual authors consider to be important. This list, together with our theoretical results, is the basis for our expert survey described in the following section.

4.2 Step 2: Expert Survey

There is a number of viable approaches to the problem of identifying properties that signify membership in the family of gamified systems. One such approach would be to examine the properties of a large number of systems that are considered to be gamified, looking for overlap. Such a purely inductive approach would be very interesting, but would also run the risk to find many common properties that are not connected to gamification at all. To give a very basic example, virtually all gamified systems described in literature make use of digital technologies and, more specifically, a client-server architecture.[5] If we were to perform the aforementioned study, we would find these two properties as core to the gamification family, even though they are irrelevant to most definitions of gamification (see section 4.1.2). It seems therefore prudent to use a deductive approach that makes use of existing literature and expert knowledge to narrow down relevant properties, followed by an inductive refinement. As with all the following steps, we will describe our methods in the next subsection, followed by the results of the respective step.

4.2.1 Methods: Expert Survey

We extracted a preliminary list of potential properties of gamified systems from popular literature on the topic. We included literature directly concerned with the nature of gamification (e.g. Kapp, 2012; Zichermann & Cunningham, 2011; Blohm & Leimeister, 2013) but also sources from related areas, such as game studies (e.g. Juul, 2005) and psychology (e.g. Deci et al., 1991; Csíkszentmihályi, 2008), if they were frequently mentioned in gamification literature. We collected a set of 60 terms that are used to describe games or gamification.[6] Such a list is subjective, as the selection of literature is not exhaustive and no objective criteria for the inclusion of specific terms exist. A full list of all terms with example references can be found in appendix A. Note that many of these terms have been adopted by a variety of authors and that the examples listed do not necessarily point to their first mention.

[5] In most cases, user data is collected locally, through a mobile application or a browser, and transferred to a server for storage and communication.

[6] We conflated similar terms in order to keep the number of terms manageable. E.g. badges and achievements, tasks and missions, and groups and group tasks.

In order to validate the selection, we compiled the terms into a survey and distributed it to the authors of the top 100 hits for a Google Scholar search on the term *gamification* on 07.08.2014, given that an e-mail address was available. This resulted in a total list of 120 experts. The survey was also circulated in two social media groups concerned with the topic of gamification, comprised of about 2000 members each. Finally, roughly 50 participants[7] of a workshop on gamification at the Mensch und Computer 2014 conference were also asked to participate. The experts were asked to rate each term as a) relevant to a description of a gamified system of their choice, b) relevant to a description of gamification in general, or c) not relevant for a description of gamification. The experts were further asked to name any other terms that they considered to be relevant. There was no limit to the amount of items the experts could mark as relevant. Unlike nominal definitions, this approach embraces the fact that not all gamified systems have identical properties, but rather properties that resemble each other. By asking experts what the most important properties for describing gamified systems are, we hoped to identify areas in which gamified systems may overlap. The distinction between a specific gamified system and gamification in general was introduced in order to distinguish between properties that are common to most gamified systems and those that signify specific forms of gamification.

While the survey included open ended answers to enable the experts to add terms that we missed in the creation of the initial set, this approach was still biased towards our previously selected terms. This is due to the difference in response behavior between open-ended and closed questions, as shown for example by Griffith et al. (1999). We discuss our mitigation of this bias below, but it is important to keep in mind that this approach is much more effective at proving the validity of a term on the list than at proving the validity of a terms absence. In other words, we can show that a selection of terms describes relevant properties of gamification, but it is likely that other relevant properties exist.

Proper statistical analysis of the resulting data is difficult with a low number of respondents, as we cannot generalize from the results. As such, we can only attempt to extract the opinions of the experts that did participate. We employ two major tools to do so:

[7] We do not have an exact count of the participants that were on the workshop's mailing list. We use the number of participants reported by the organizers of the workshop here (Schmidt, 2014).

For one, simple aggregate statistics help us to identify majority opinions — terms that a majority of experts considered important in either category and terms that the majority did not consider relevant. They also help to identify terms that differ greatly in importance between the categories. Simple K-means clustering[8] (see Hartigan & Wong, 1979) of the responses additionally allows us to identify groups within the experts that followed similar voting patterns. This is especially helpful, since the responses are high dimensional in nature — effectively 180 binary dimensions, or three sets of 60 binary dimensions (with the sets split along the response options of relevant to *gamification in general*, relevant to the *specific chosen implementation of gamification*, and *not relevant*, with *yes* or *no* responses for each.) As we will see, some dimensionality reduction is simple, as certain combinations of responses do not make sense (i.e. no term can be both relevant in one of the categories and not relevant at the same time). Still, the data is far too complex for simple manual analysis, even with few responses.

In K-means clustering, the choice of the number of clusters K is left to the expert evaluator and "the partition that appears the most meaningful to the domain expert is selected" (Jain, 2010). While this may seem arbitrary, one must keep in mind that the analysis is exploratory in nature and intended to help the expert identify patterns in the data. We followed this approach and ran K-means clustering on the three sets of 60 dimensions with a varying number of clusters, looking for a useful partition. One tool that helps with the identification of clusters is to compute the within groups sum of squares[9] for various values of K and to look for an abnormally large change therein. K-means clustering attempts to minimize the within groups sum of squares (see Jain, 2010), which is easier for higher values of K. Obviously, too high a number of clusters is useless for analysis, as it will not reduce the data enough to be useful for manual analysis. (The extreme would be to have K equal to the amount of points in your dataset, in which case clustering would result in no change whatsoever.) It is therefore necessary to find a value of K that finds all meaningfully different groups of points but does not separate them further. A large dip

[8]While many other clustering algorithms exist, none is clearly superior and the dated K-means still considered to be acceptable (Jain, 2010). As our exploratory analysis does not require more than a simple measure of similarities between groups, K-means will suffice here.

[9]The within groups sum of squares is a measure for the differences between points in a single cluster.

in the within groups sum of squares between two values of K indicates that a meaningful partition has taken place, while low changes indicate additional separations that are not meaningful. We have therefore calculated these values for our datasets and used the dips identified as a guideline for our choice of K in the following analysis (see figure 4.2). As one can see, the only consistent dip among all four lines is at two clusters. After that, only the "not relevant" set of responses (and through it the set containing all responses) shows possible additional breaking points at four and six clusters respectively. Since six clusters are hardly useful when analyzing 13 responses and the change at four clusters is minimal, the following analysis will mainly use two clusters. We have, however, run an additional analysis with 3 clusters for each set and will refer to them where they provide additional insight.

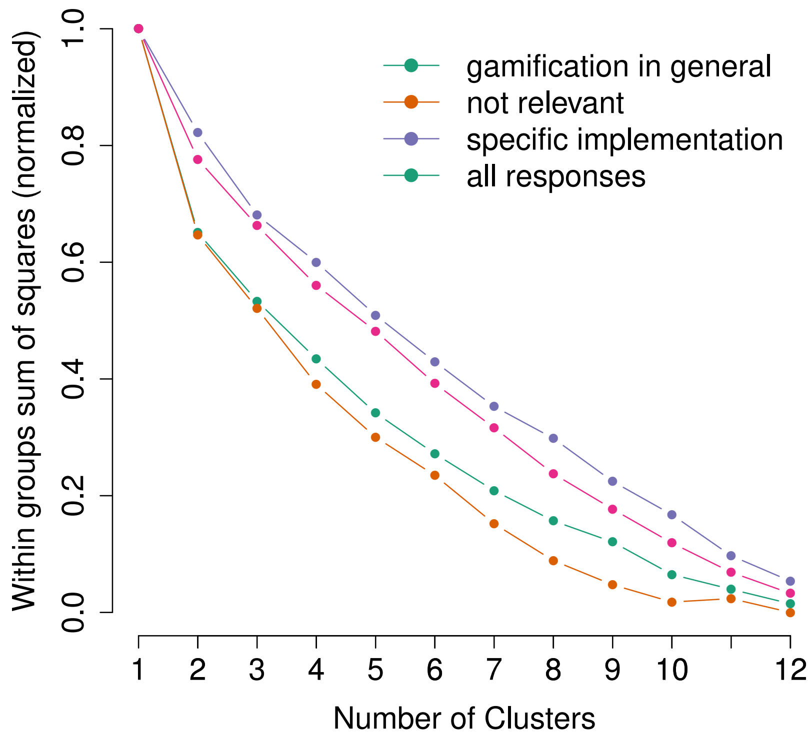


Figure 4.2: Within groups sums of squares for different values of K in K-means clustering of expert survey results. The all responses line shows clustering over all 180 dimensions, the other three represent 60 dimensions each.

4.2.2 Results: Expert Survey

Thirteen experts completed the survey, an unfortunately low number when compared to the number of experts asked to participate. This amount of participation does not allow for statistical analysis, such as the identification of differences in opinion between researchers and practitioners or even different schools of thought among those groups. Of the participating experts, one identified as a practitioner, two as "other" and the remaining 10 as researchers. It is important to consider this bias towards research when working with the results. All thirteen participants completed the questionnaire. When asked for a gamified system to consider throughout the questionnaire, 10 different systems were named, with Duolingo[10] being named twice and Wordrobe[11] thrice. All but one of these systems were digital, the analog one was a teddy bear. The main results of the survey were to be found in the section that dealt with the rating of terms for their importance in describing gamification. We will deal with the responses for gamification in general and the respondent's chosen system in particular in the following sections, followed by a section comparing the results of the two.

Due to the low number of responses to the first survey, statistical evaluation can hardly be relied upon. Similarly to the goals and means analysis described in section 4.1, one can identify certain trends, however. There were 37 terms that were mentioned by more than two thirds of participants as being relevant for a description of gamification in general. As one can see in appendix B, these include the elements agreed upon by most authors in the literature review (points, badges, leaderboards, levels) but also a variety of other terms, including many that were not mentioned by the other authors at all, or only mentioned by a few. Documentation of behavior, for example, was only mentioned by Blohm and Leimeister in the previous analysis, but 10 out of 13 experts in the survey considered it to be important for a description of gamification in general.

It is important to note here that the methodological differences had to be expected to lead to a variety of results. For one, authors may have tried to formulate their definitions on a single level of abstraction while the survey allowed the participants to choose terms on a variety of such levels. For another, authors may have only included the terms they consider

[10] https://www.duolingo.com
[11] http://www.wordrobe.org/

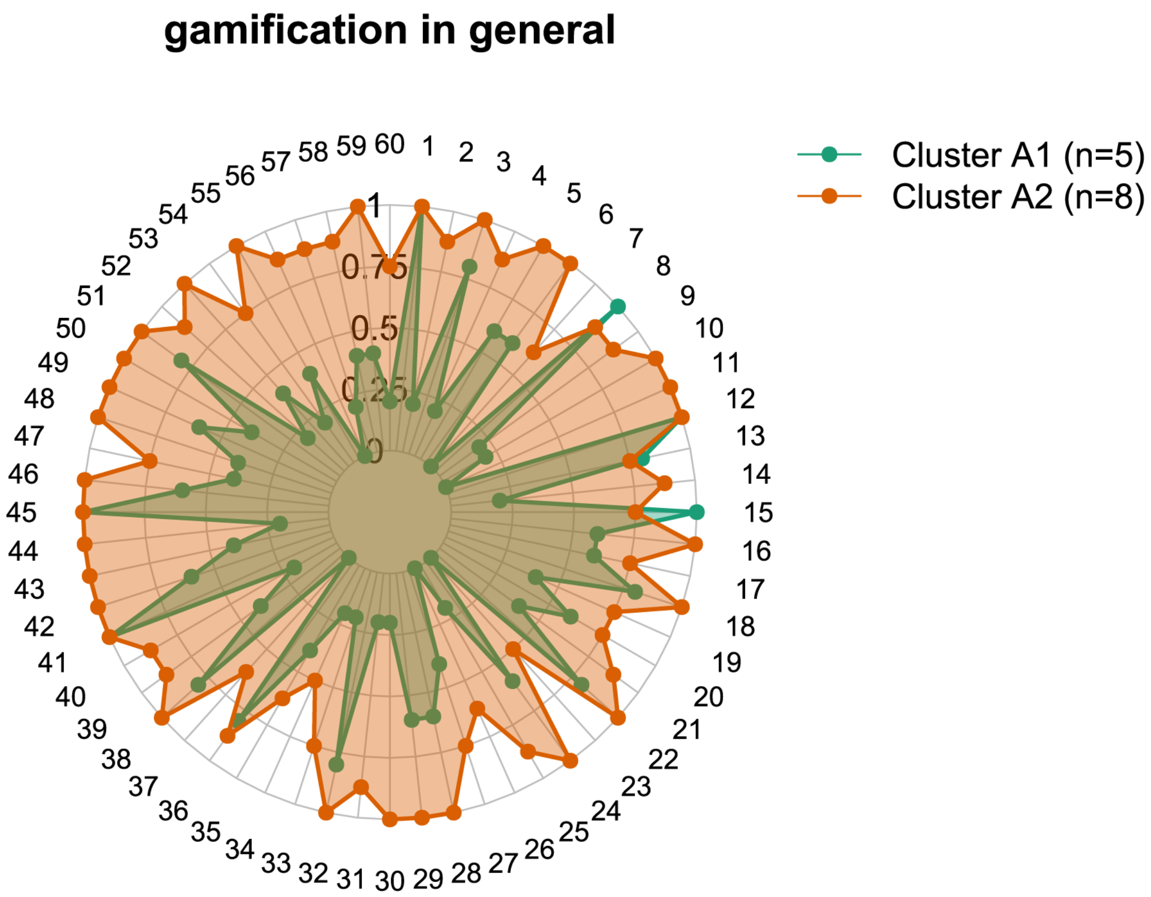


Figure 4.3: Average response for each item in the gamification in general category, divided into 2 clusters. Each number on the spider diagram corresponds to one of the 60 terms from our survey, see appendix A for an index.

most important in their own definitions, while the survey allowed the participants to rank any amount of terms as important. Finally, the fact that the survey gave predefined options to the experts may have led some to mention terms that they would not have used on their own. It is also important to note that while some of the sixty terms were not found in our goals and means analysis, all of them originated in literature either on gamification or often cited in gamification contexts.

The terms that were unanimously rated as being relevant were: achievable tasks, accomplishment, goals, and challenge. Even the item with the lowest amount of mentions, conflict, still got the approval of four experts. On average, experts rated 72% of items as being relevant for the description of gamification in general (562 out of 780 possible ratings.)

Cluster analysis shows two distinct response patterns in the general category (see figure 4.3): one group (cluster A1, n=8) in which experts considered most of the terms in the survey to be relevant for gamification in general, and one (cluster A2, n=5) with a more distinguished response. All data points with fractions show that there is disagreement within the clusters. The points that were rated low by both groups are of less interest here, as they logically also score low on the simple analysis of the number of mentions above and are clearly not terms that the experts considered to of much relevance for describing gamification in general: negotiable consequences (7), conflict (23), leading others (26), effortless involvement (34), and organizing and creating order (37). Similarly, the terms that both groups ranked highly are logically at the top of the simple analysis (e.g. achievable tasks (1), player effort (8), challenge (12), accomplishment (41), and goals (45), to name the extremes.)

Much more interesting from a cluster analysis point of view are the terms on which the two groups differ. Cluster A2 considered the following terms to be unimportant, while cluster A1 found them to be relevant: being the hero (2), curve of interest (4), relationships (9), surprise and unexpected delight (10), aesthetics (11), customization (14), time pressure (30), avatars/virtual worlds (31), valorization of outcome (40), fame and getting attention (44), competence (52), meaning (56), and player attachment to outcome (57). These terms can be split roughly into two groups, those that are rather used to describe the content of a game rather than its mechanics (e.g. being the hero, aesthetics) and some high level concepts (e.g. meaning, player attachment to outcome). Especially the former grouping is a clear distinction between the two clusters, as no content-related term was rated highly by cluster A2. This is a split in opinion that we have also seen in literature, where some authors considered elements like narrative (here: storytelling (5) with a less pronounced but still quite strong difference between the clusters) to be game elements useful for gamification, while others stress that gamification focuses on mechanics only. In literature, this split is roughly along the practitioner/academic line. Our results cannot confirm that, however, due to the low representation of self-identified practitioners. It is noteworthy that this distribution of responses gives cluster A2 a very large impact on the overall results in the general category, leading to an over-representation of the mechanical/researcher view in the end result.

Not surprisingly, the terms selected as important for the description experts' chosen gamified

Table 4.6: Comparison of mentions of terms for gamification in general and gamification of a specific implementation. The rank columns show the position in an ordered list of terms, with rank one being the most mentioned item. The last two columns show the difference in mentions and ranks respectively. Items with a difference of less than 5 mentions in favor of the general category were excluded.

	spec. imp.		gam. in general		Δ	
term	#	rank	#	rank	#	rank
group tasks	2	55	10	23	−8	32
intangible rewards	5	36	11	12	−6	24
social incentives	5	36	11	12	−6	24
storytelling	4	49	10	23	−6	26
gifting	1	58	7	51	−6	7
levels	6	29	11	12	−5	17
variable outcome	5	36	10	23	−5	13
virtual trade	2	55	7	51	−5	4
...						
reward structures	12	2	11	12	1	−10
competence	9	17	8	38	1	−21
player attachment to outcome	10	11	8	38	2	−27

system are less homogeneous. Only 20 terms were mentioned by more than two-thirds of the participants, about 54% of the amount seen for the general description. Participants only gave 27% less votes for relevance in the specific category (409 out of 780 ratings, 52.4%) as opposed to the general category, however, suggesting that the way applications were gamified varied quite significantly. Notably, levels and leaderboards did not appear in this list of often mentioned terms for specific applications, being used only 46% and 54% of the time, respectively. Points and badges were still extremely common among the examples selected by the experts, however.

These sorts of differences can be seen in table 4.6. It shows the terms with the highest difference in mentions between the specific and the general categories. Group tasks, for example, were seen as important for gamification in general by a large majority of participants, while only two marked it as relevant to their specific application. Similar, if less pronounced, differences can be seen for the terms intangible rewards, social incentives, and storytelling. Interestingly, gifting shows a high discrepancy as well. While it was mentioned to be relevant for gamification in general by seven participants, only one said it was also relevant in his or her own work. Levels and leaderboards can also be found on this list, with slightly less pronounced differences (11 to 6 and 11 to 7 mentions, respectively).

While hard to prove from such a small sample, terms on this side of the list are interesting because experts acknowledge their importance for gamification but do not use them themselves (or do not see them in the systems they study.) Obviously, not all systems can have all attributes, but if some are often mentioned in the discussion of gamification but rarely used, there might be an underlying reason. One might be that they would be nice to have but are hard to implement, another might be that experts mostly agree that they are not very helpful but are still something other people use. Leaderboards are an example that is sometimes mentioned in literature as having been excluded on purpose due to the possible negative effects competition can have on users. We have discussed this effect earlier in section 2.2.3.

Another major takeaway from the list of terms that are highly ranked in the *general* category but less often mentioned in the *specific* category is that these are features that may be used to distinguish gamified applications from each other. While the sample size is too low for proper clustering of applications, the terms at the top of the list in table 4.6 are the ones most likely to matter for such a categorization. A term that is used in most applications (high rating for both *specific* and *general*) is not useful for such a distinction. One that is rarely used at all is likely not very useful for describing a gamified system at all. The ones with a high *general* rating but a low *specific* one, however, designate those terms that are commonly understood by experts for their value in describing gamification but separate a subset of gamified systems from the rest. The terms with a specific rating of less than 40% and over 70% recognition by the experts are, in order of rating difference: group tasks, intangible rewards, social incentives, storytelling, and variable outcome. Other terms show a strong potential for cluster separation as well, such as gifting, or virtual trade, but they have less recognition among the experts in the general category. It might be valuable to rate a large set of gamified applications using these terms in order to identify clusters and thereby types of gamified systems.

The reverse comparison shows less pronounced differences. Only three terms were selected more often as being important for the specific application than for gamification in general (player attachment, reward structures, and competence) and in each case the

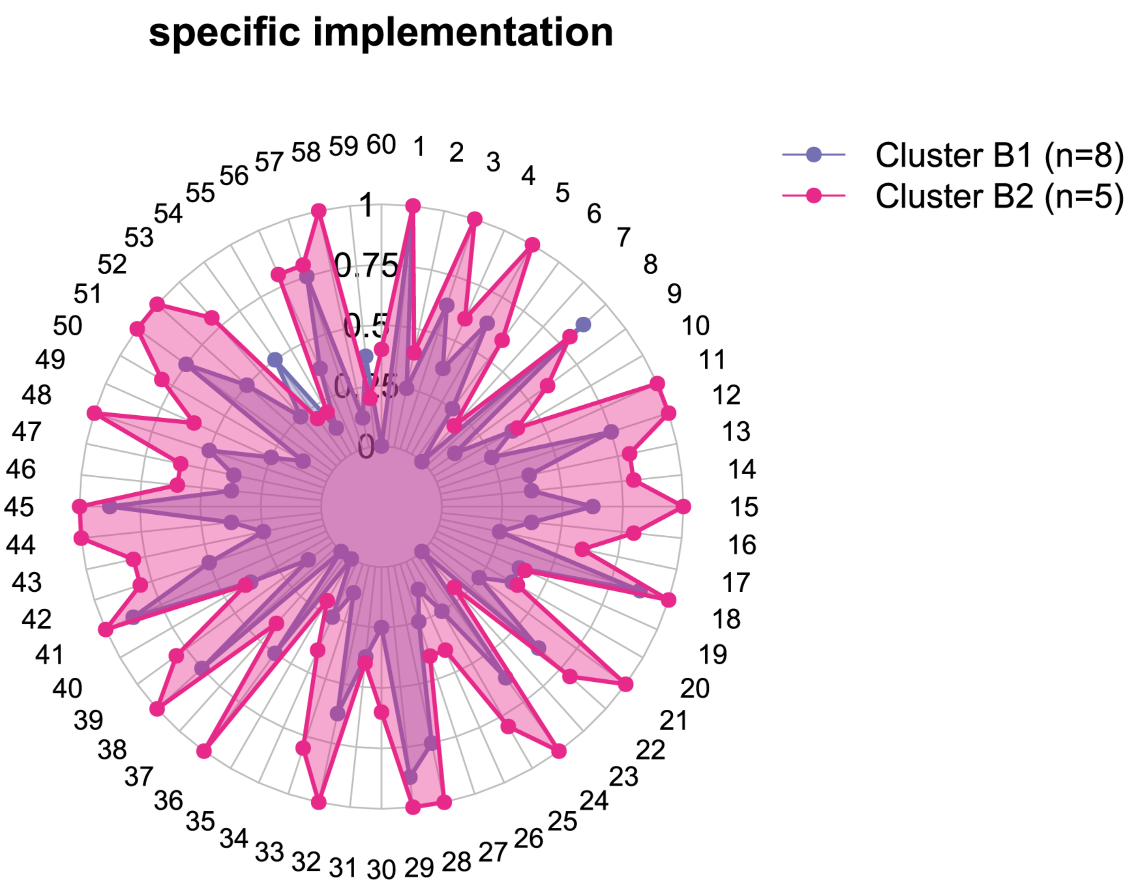


Figure 4.4: Average response for each item in the specific implementation of gamification category, divided into 2 clusters. Each number on the spider diagram corresponds to one of the 60 terms from our survey, see appendix A for an index.

difference was slight. These small aberrations can easily be attributed to the relatively complicated structure of the survey or, potentially, to different interpretations of the categories by the authors. Clearly, most authors assumed that any term relevant to describing their specific system would also be relevant to a description of gamification in general. A few participants might have assumed that the category "gamification in general" might mean "all gamification". In case of a repetition of the survey, this needs to be made more clear.

Unlike the general category, there is no cluster in the specific responses that selected most terms to be important (see figure 4.4). The identified clusters are of the same size as in the previous analysis. While they are not comprised of the same members, they are similar. Three members of the previous cluster A1 are now in cluster B2. Once again, the most interesting result of the cluster analysis are the terms on which the two clusters disagreed. Since the overall agreement within the clusters was lower (expected, as the experts chose different implementations of gamification to discuss), the maximum disagreements are lower here as well. Terms with high disagreement between clusters are: competition (3), feedback (5), relationships (9), surprise and unexpected delight (10), customization (14), autonomy (21), documentation of behavior (24), time pressure (30), social engagement loops (39), cooperation (43), fame and getting attention (44), social incentives (49), and variable outcome (50). In all these cases, cluster B2 had an equal or higher average response than cluster B1 and only five items in total were rated higher on average by cluster B1. This is the opposite behavior observed in the general category. This shows us that the experts who were more specific in which terms they considered to be relevant to gamification in general and who excluded many of the content related terms had a stronger agreement on which terms where important to their specific implementations and considered more terms to be important for their description than the other group.

Interestingly, this list has a non-trivial overlap with the list of terms that that group considered to be of little importance for describing gamification in general: relationships (9), surprise and unexpected delight (10), customization (14), time pressure (30), and fame and getting attention (44). This may be a case of the experts highlighting properties of gamified systems that they consider important (and therefore include in their own applications and/or chose to discuss a specific application in the survey that they consider to be a good example of gamification) but that is not often used in gamification and therefore not selected in the general category. We consider this to be a sign that these experts expressed the same conclusion we came to in our analysis of empirical studies and in our own evaluation of gamification in LMS: implementations of gamification are far more limited in scope than the theory behind gamification would suggest.

Other interesting facts can be gleamed from looking at the terms last on each list. Conflict, for example, has the least mentions in both categories (no term was not selected by at least one person in each category). This is especially interesting as conflict is seen as a major element of most definitions of the term *game* (see section 2.2.3). Now, if conflict is integral to games but decidedly not relevant for gamification, that could be seen as an indication that the distinction between full games and gamification drawn by Deterding et al. is not artificial. Closely related is the term "negotiable consequences", directly adapted from Juul's classic game model (Juul, 2003). According to Juul, it is integral to a game that any consequences of the game (such as losing or gaining money, for example) are optional. A game of poker, to use a classic example, can have real-life consequences, even severe ones. That does not make poker less of a game, however, as these consequences are not an integral part of the game, but negotiable. You can play poker without any real-world consequences at all.

Unlike conflict, which is likely to be excluded by the experts on purpose (since not all users would react well to it), negotiable consequences are more likely to have been excluded either due to the experts not seeing the relevance of the term (potentially due to unfamiliarity with Juul's classic game model) or due to the rarity of gamification ever having real-world consequences. While there are examples of the latter (such as coffee shops offering their *mayor* in Foursquare free coffee), these are rare and mostly positive. It seems likely that most experts would agree that adding binding real-world consequences to gamification (e.g. tieing wages to a score in the company's knowledge management system) would be a bad idea if asked directly.

Leading others is a final example of a term very infrequently chosen by the experts to be important to either form of gamification (specific or general). Unlike the previous two examples, leadership is not a feature common to games, but nonetheless found in a variety of games (mostly those for multiple players, though some video games put the player in the role of the leader of a group). Massively multi-player online roleplaying games (MMORPGs), for example, tend to include group content that is so difficult that an unorganized group cannot hope to succeed. Leaders quickly emerge (either ad-hoc in random grouping situations, or designated when players organize in guilds). One factor that might lead to the less frequent use of leadership in gamification might be that only a small subset of the population of a given game or system can be in a leading role. Since gamification tends to be targeted at a broad majority, leadership might not be seen as a useful tool.

Our survey provided helpful insights into the different understandings of the term gamification among our experts, especially highlighting a split between those who focus on

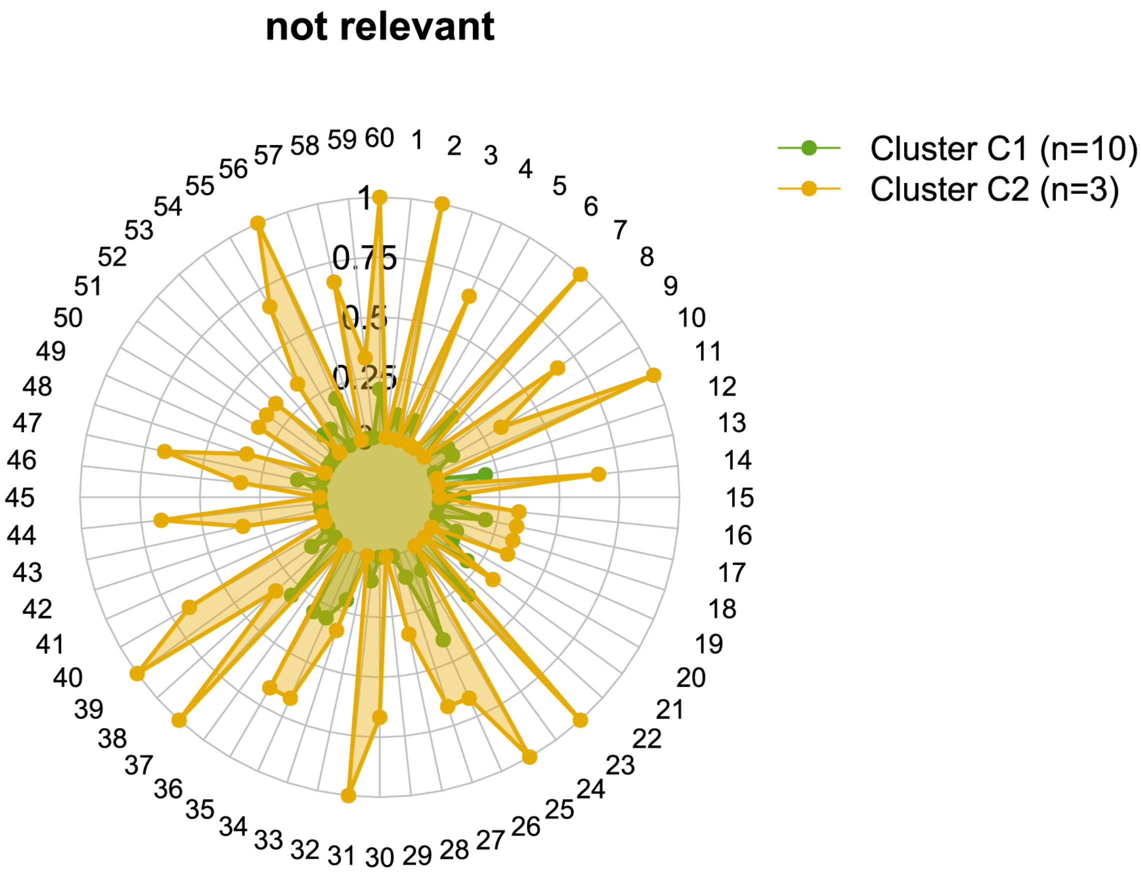


Figure 4.5: Average response for each item in the *not relevant* category, divided into 2 clusters. The majority of experts made little use of this option. Each number on the spider diagram corresponds to one of the 60 terms from our survey, see appendix A for an index.

mechanics and those that consider game-like content to be gamification as well. The main reason for the survey was a narrowing down and a confirmation of the list of properties of gamification that we extracted from literature for further use in our work, however. The next section describes the process and results of said narrowing down. As that process is mostly based on majority opinions gathered from the expert survey, one should keep the previously mentioned split in the experts in mind. While the disagreements in the specific category had very little influence on our finally chosen terms (most terms on that list still made it to the narrowed list), the mechanics-oriented cluster of experts had enough influence in the general category that eight terms that the other group considered important for the description of gamification in general were not included: being the hero (2), curve of interest (4), relationships (9), aesthetics (11), customization (14), avatars/virtual worlds (31), valorization of outcome (40), and meaning (56). The fact that many experts considered most terms to be important makes such an exclusion unavoidable, however, if we wish to actually narrow down the list of terms. A quick look at the clusters of respondents according to their choice of *not relevant* (see figure 4.5) shows that without minority opinions, no exclusion of terms could happen. On the positive side, this high agreement of experts with our chosen terms and the very low number of manually added terms support the results of our previous work: We have extracted a list of terms that is consistent with what a number of experts consider to be relevant for gamification and do not seem to have missed highly relevant terms. We are therefore confident that the resulting narrowed down list is a good starting point for more structured analysis of gamification.

4.3 Step 3: Creating the Gamification Inventory

One major result of the expert survey was a rating for each of the sixty terms, indicating how important experts considered each term to be for the description of gamification. Following the previously outlined idea of a family resemblances approach to defining gamification, we used this to identify properties of gamified applications that were likely to overlap between different gamified systems. This approach was intended to help us reduce the number of items to use as a basis for the creation of an instrument for the qualitative evaluation of gamification, the Gamification Inventory.

4.3.1 Methods: Creating the Gamification Inventory

In order to reduce the original list of terms to a more manageable set of items for the inventory, we had to remove those terms from the list that the experts did not agree upon and add additional terms, if enough of the experts mentioned them. We chose to include all terms that were selected to be relevant in either category by at least two thirds of the experts for further consideration, as well as all terms that at least a quarter of experts added manually. This assumes that a term that experts choose to add manually must be very important to them, allowing for a lower overlap and thereby mitigating (but not removing) the aforementioned bias introduced with the survey structure.

This resulted in 38 terms, which we grouped into five categories. Two categories contain most of the items, split according to an experiential/systemic dichotomy (see Broer & Breiter, 2015). The other three categories (rewards, goals, social) are mostly systemic in nature, but address specific issues often discussed in gamification literature and therefore deserve highlighting. A detailed description of the final instrument can be found in the results chapter. For the use of experts, we have expanded the terms from the survey to questions that can be answered in the positive or the negative and asked for descriptions of how each term was expressed in the analyzed system. The instrument also allowed for an in-between response of “maybe” to cover cases in which the response is not as clear-cut. The main reasoning behind this third option were cases in which a system was customizable in such a way that it could have certain properties if properly configured, but did not express them otherwise. This response was especially necessary in systems with user created content.[12]

It is important to note, that the standardized part of the instrument (the yes-no-maybe response) was intended to be used as a summary of the qualitative findings for each item, allowing for better statistical evaluation and comparison. One could just as well employ the instrument without the standardized part and then later on code the responses according to any other scheme one deems interesting.[13] The resulting inventory allows for a structured investigation of gamification aspects in a given system and a structured comparison of such systems, but remains a qualitative instrument at its core. The main function of the inventory is to guide an evaluator through the evaluation process and to ensure that the major properties of gamified systems have been considered.

4.3.2 Results: Creating the Inventory

The resulting Gamification Inventory can be found in appendix D. The instrument contains a total of 38 items, consisting of an identifier comprised of a letter that denotes the category and a number that references the order of the items, the original term, an expanded question,

[12] E.g. our analysis of learning management systems (LMSs) showed that challenge was difficult to assess, as it depended on instructor input. The evaluated systems themselves did not provide challenge, but they afforded the teacher to add challenging content.

[13] Which is what we changed towards in version two of the inventory.

System mechanics		
collecting		Does the system provide opportunities for the user to collect things?
Y/N/M		
documentation of behaviour		Does the system visibly document the user's behavior?
Y/N/M		

Figure 4.6: Items 1 and 2 from the mechanics category of the first version of the Gamification Inventory. Own illustration.

a standardized field for a yes/no/maybe response, and a free-form response field. See figure 4.6 for an example.

The experiential category of the instrument contains terms dealing with the user's experience, or, more specifically, with affordances for specific experiences. Since the instrument is intended to be used for the evaluation of a system, we cannot use the actual user experience in the evaluation but only the affordances therefor. The category includes items such as accomplishment, curiosity, positive emotions, and relatedness.

The second category, system mechanics, deals purely with the functionality of the system, rather than the experience of the user. Example items are feedback, storytelling, time pressure, and variable outcome. Unlike the experiential category, most of these are rather straightforward to identify in a system, as they do not require an evaluator to anticipate the reaction of the user to specific properties of the system. E.g. a system either uses time pressure or it does not, that does not usually depend on the user.[14]

The rewards category highlights some mechanics that are very common in gamification, those dealing with extrinsic rewards – such as points, badges, or progress. As such it is likely to be the category most relevant to identifying classic gamification — or pointsification (see section 4.1.2). The rewards category is essentially a subset of the mechanics category, but the correlation to the most common gamification techniques makes it stand out.

The goals category includes those items that stress the importance of goals and structured tasks in games. It is the smallest of the categories with only three items — clear goals, challenges/quests/missions/tasks, and achievable tasks, and it could be argued that

[14]Systems can still employ a variety of forms of time pressure, of course.

the items therein could be included in other categories (such as mechanics) instead. Goals are integral to games, however, which is why we gave them a more prominent position in the instrument. The original survey actually included both *goals* and *clear goals* as terms. We combined these for simplicity's sake, as the two terms were very similar.

Finally, the social category encompasses social functionalities that one might adopt from games. The properties discussed in the items in this category all require other users either directly or indirectly. Examples are leaderboards, cooperation, and social engagement loops.

While the selection of terms was based on expert opinion, the split into the five categories as well as the design of the actual instrument were not. Additionally, the experts were not asked to specifically select terms for such an instrument, which might have affected their choice. This made testing the instrument a vital next step, described in the following section.

4.4 Step 4: Test for Reliability and Refinement

Once the first version of the Gamification Inventory was complete, it had to be tested for reliability. While the instrument is generally intended for the use of a single expert, it is important that different raters have a similar understanding of the items. Otherwise, the instrument would likely fail in guiding evaluators through all important properties of gamified systems and make the comparison between systems more difficult. The following sections describe our methods for testing and refinement as well as the results of both.

4.4.1 Methods: Test for Reliability and Refinement

In order to test the validity and reliability of the instrument, the following setup was used: Five LMSs (Moodle, Edmodo, Blackboard Learn, Schoology, and Canvas) were selected from a list of popular LMSs provided by Capterra (2015). Capterra provides a ranking of LMSs according to their user base and social media presence. We selected the top five LMSs from that list, that were targeted at a K–12 education context (rather than, for example, training on the job), and set up a demonstration system for each one. We further collected external information available online for each LMS, such as lists of features, and introductory videos.

Access to the demonstration systems and additional information were provided to each of four evaluators. Three of the evaluators had basic training in gamification while the fourth was an expert on the topic. The use of non-expert evaluators was meant to highlight issues with the understanding of the individual items. This setup was likely to result in a higher disagreement between raters, as they could not be expected to have the same understanding of the nuances of gamification as an expert. At the same time, issues with the way the items were formulated might have been hidden if only expert evaluators were employed. The expert evaluator was used as an authoritative vote in cases were the evaluators were split in their results as well as a measure to compare the non-expert evaluators results to. Each evaluator was asked to familiarize him- or herself with each system and then complete the Gamification Inventory for it. This resulted in a total of 19 filled out inventories, four for each of five LMSs, except for Canvas, which one evaluator failed to rate.

Evaluation of the reliability of the instrument was focused on its standardized part — the yes/no/maybe responses to each question. While the instrument itself is mostly qualitative in nature, our validation thereof employed quantitative measures to see if different evaluators could use it to arrive at similar results. We applied two different measures to the results, one aimed at determining the overall inter-rater reliability (IRR), that is the rate of agreement between the raters (see Gwet, 2014, p.4), the other at the rate of disagreement for each item. IRR is important to consider when creating such an instrument, as it is a sign of the instruments success at guiding evaluators through the evaluation process. While different evaluators will have different opinions, especially in qualitative evaluation, low IRR can be an indicator of differences in understanding among the evaluators. To measure inter-rater-reliability, we considered all data to be nominal, thereby considering a *maybe* to be just as different from a *yes* as a *no* would be. With nominal data and more than two raters, Fleiss' kappa is a common measure for inter-rater reliability (Gwet, 2014, p.52). We therefore employed Fleiss' kappa as a measure of overall agreement between raters.

While a useful statistic, Fleiss' kappa does not provide us with all relevant information. For one, improvement of the instrument requires not only the measurement of inter-rater reliability, but also the identification of items with low agreement as those are likely to be the ones that need improvement. For another, treating the data as nominal may measure a higher disagreement than there actually was because of the use of the *maybe* response. In order to measure the disagreement for each item, we therefore re-coded the data to be on an interval scale, with all three responses being equidistant (the distance between *yes* and *maybe* is the same as between *maybe* and *no*) (see Stevens, 1946; Labovitz, 1967): yes = 1, maybe = 2, no = 3. The choice of numbers here is arbitrary and should not influence the outcome as long as equidistance is observed. This means that the disagreement between a *yes* and a *no* is considered to be twice as high as between either one and a *maybe*. Having access to an interval scale, we were able to employ variance (see Wonnacott & Wonnacott, 1985, p.38) as a measure of disagreement between raters for a single item and a single system, and the sum of variances over all systems for a single item as the measure of disagreement on that specific item.[15] For the purposes of our study, we considered every item with a sum of variances higher than 0.5 to be an item with a high disagreement among raters. We then analyzed the open answers to those specific items qualitatively in order to identify the reasons for the low agreement. An alternative approach to measuring the agreement would be the use of weighted kappa (see e.g. Gwet, 2014, p.62) to indicate that the difference between a *yes* and a *maybe* is less than that between a *yes* and a *no*. As Gwet (2014, p.98) points out, however, the choice of the proper weighting scale is complicated and would likely be of similar arbitrariness as our approach using the sum of variances. Both methods will rank items according to the agreement between raters and those rankings will be similar, independent of individual numbers.

The testing in step 4 provided us with data on items that were difficult to apply, signified by low agreement between raters, as well as qualitative data detailing the reasons for the disagreement and data on the applicability of the instrument as a whole. We used these data as a basis for refining the instrument. At this point in time, refinement was not about changing the properties selected through the expert survey, but rather the phrasing

[15] In theory, we could employ Fleiss' kappa as a measure of IRR for each individual item and look for items with a low score. Since we are only interested in finding items with a high disagreement among raters however, and since the usefulness of the adjustment for randomness inherent in the kappa statistic (see McHugh, 2012) with a low number of raters and observations is questionable at best, we employ the purely descriptive measure of variance instead. As all items use the same scale, normalization is not necessary. Similarly, the mean variance or the composite standard deviation could be used for this comparison, but that would not change the ranking of results, just create less legible numbers.

of the questions and the description of the response options in the standardized part of the instrument.

The issues with individual questions were addressed as follows: All items with a sum of variances of 0.5 or higher in the test in step three were examined individually and the free-form responses analyzed for signs of issues with the question. For example, item M9 - time pressure was originally phrased "Does the system limit the amount of time a user has to complete a task?". We found that some evaluators considered the ability to set deadlines for tasks as time pressure, while others did not. We rephrased the item to be more specific as "How are spontaneous decisions or reactions or a sense of urgency…". The other problematic items were rephrased as well, but in many cases proper detailed specifications would have complicated the questions too much. Instead of packing all details about the intended understanding of each term into the instrument, we chose to create an evaluation guideline describing the meaning of each item in more detail. In practice, evaluators would work with the inventory as designed, but be able to refer to the guideline when unsure about an item. Additionally, the guideline can be used during the coding process to properly categorize evaluator responses. The guideline can be found in full in appendix D

4.4.2 Results: Test for Reliability and Refinement

Once the Gamification Inventory was complete, it was tested for reliability with four evaluators. When considering all four evaluators, Fleiss' kappa returns a relatively low inter-rater reliability (IRR) of $\kappa = 0.23$ for the standardized part of the instrument. Disregarding the responses of one evaluator who showed a rather low understanding of the topic, as visible in the qualitative responses, improves IRR slightly to $\kappa = 0.31$. Comparing all additional evaluators' results with those of the main evaluator using equally weighted Cohen's kappa,[16] we arrive at $\kappa = 0.451$, $\kappa = 0.321$, and $\kappa = 0.14$ respectively. The most obvious overarching issue, as identified from comparing free-text responses by different evaluators, was the use of the standardized part of the instrument. Evaluators had a tendency to find similar properties in the systems but to code them differently. Academic responses to the instrument

[16] The classic measure of non-weighted IRR between two rater introduced by (Cohen, 1960). Here, we are following the procedure described in (Gwet, 2014, p.32ff).

E2 autonomy	How is a feeling of autonomy on the side of the user…	
… built into the system?	… possible with proper user input?	I'm not sure, because:

Figure 4.7: Item 2 from the experiential category of the revised Gamification Inventory. Own illustration.

during and after publication were also concerned with the apparent quantitative nature of the instrument when we described it as a qualitative instrument. As the standardized part of the instrument was only intended for ease of evaluation anyway, we have removed it in the refined version. The refined version now has three free-form response fields for each item, asking the evaluator to describe how the respective property was built into the system, how it could be part of the user experience with proper user input, and what reasons the evaluator might have to be unsure about the response. Evaluators are further instructed to leave empty all fields that do not apply. (E.g. if an evaluator is sure about a response, no reasons for insecurity need to be listed.) Proper analysis with the refined instrument now requires coding of the free-form responses, but thereby should eliminate the differences introduced through coding by the evaluators themselves and any impression of the Gamification Inventory as a quantitative instrument. In their essence, these questions still reflect those on the original instrument. The main two reasons for *maybe* responses — insecurity and dependence on user input — are now more directly reflected and the difference of creating gamification elements through user input and gamification elements that are built into the system should be more obvious. Figure 4.7 shows an example item from the refined instrument.

Analyzing each item individually, we identified 13 items with a high disagreement between raters as indicated by the sum of variances for that item. We consider a sum of 0.5 as the lowest acceptable agreement (see table 4.7). Figure 4.8 shows the mean and median variances for items by category. The goals category had the highest average disagreement among raters with two out of the three items in the category having a sum of variances higher than 0.5. In all other categories the mean was much higher than the median, indicating the existence of outliers (see Wonnacott & Wonnacott, 1985, p.204). You will find the

Table 4.7: Items with a high disagreement among raters.

#	Item	Σ Variance	#	Item	Σ Variance
E6	learning /mastery	1.03	M9	time pressure	0.75
S6	social engagem. loops	1.03	G3	clear goals	0.66
R4	points	0.92	E10	relatedness	0.61
M5	nurturing. growing	0.91	S1	competition	0.59
E8	positive emotions	0.81	S2	cooperation	0.58
G2	challenges/quests/missions/tasks	0.81	M4	fixed rules	0.52
			R5	progress	0.52

results of the qualitative analysis of the free-form answers for each of the questions with a high disagreement below. A second version of the inventory, refined with the results from this analysis, can be found in appendix D.

E6 – learning /mastery: Learning and mastery was one of the items with the highest disagreement in our evaluation. With two exceptions (both by the same rater), learning and mastery were always coded as either being there or not there and most raters gave the same rating for all five systems. This split indicates an issue with the interpretation of what exactly constitutes learning and mastery. The qualitative data even shows that the two raters who answered the question in the positive had a diverging understanding of the item. One rater considered the presence of quizzes and assignments as sufficient for promoting learning while the other looked for the option to retake quizzes and assignments and for dedicated mastery systems. This is, in part, an issue with the field of application. Since we analyzed learning management systems, evaluators found it easy to identify features connected to learning. With a focus on mechanics, as suggested by the result of our expert survey, learning rather refers to gamification design principles such as freedom to fail and feedback, rather than content. We have clarified this difference in the evaluation guideline.

S6 - social engagement loops: This item was similar to item E6 in that there was a high disagreement between raters and in the predominant use of “yes” or “no” responses. Unlike for *learning/mastery*, individual raters’ ratings differed between systems, however. Three of the raters found social engagement loops in some, but not all systems. They disagreed on which systems that should be, indicating a different understanding of the term. Further evaluation underlines this hypothesis. Some raters considered the presence of social

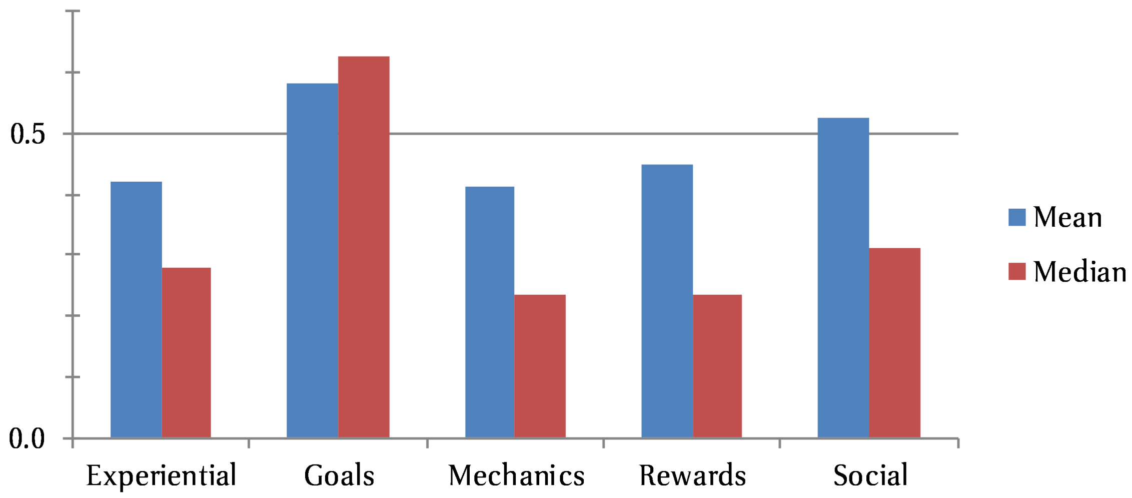


Figure 4.8: Mean and median variances of item responses by category. Own illustration.

networking functions to be sufficient to speak of social engagement loops, while others looked for features such as SMS notification on user activity. Social engagement loops are one specific gamification element in the gamification design principle *social elements*. We have clarified this difference in the evaluation guideline.

R4 - points: It was surprising to find points near the top of the list of disagreed upon items. Again, there are almost no *maybe* ratings and the qualitative evaluation points towards differences in understanding once more. Some raters considered the fact that assignments were graded with points (e.g. 7 out of 10) as enough of an indicator for R4. Others looked for cumulative point gain, such as experience points. The broad general meaning of the term is what complicates the response to this question, although it was meant to specifically represent the use of the gamification element *points* as part of the gamification design principle of *rewards*. We have clarified the focus on accumulation of *points* as a *reward* in the evaluation guideline.

M5 - nurturing, growing: The high disagreement level on this item is mostly due to one rater considering all forms of accumulation (e.g. points, badges) to be nurturing or growing. Other raters were clearly unsure about the meaning of the item in their *maybe* responses. We consider *nurturing/growing* to be part of the game design principle of *goals*, with an argument to be made for *autonomy*, and have clarified the meaning based on the work by Zichermann & Cunningham (2011).

E8 - positive emotions: Similar to E6, there was a high intra-rater agreement where positive emotions were concerned. With the exception of a single *maybe*, each rater rated all systems identically, but differed greatly from other raters in his or her evaluation. When positive emotions were identified, these were mostly attributed to the possibility of fulfilling tasks and getting rewarded for accomplishment, e.g. through badges. In addition to a different understanding of what would create positive emotions, this item also highlights a difference in rating for the same observation. Two raters found the same affordances for positive emotions in each system, yet one rated it as *maybe* while the other chose *yes*. Unfortunately, the open responses did not explain the reasoning behind either choice. *Positive emotions* have to be considered a goal of gamification, rather than a mean,[17] and as such cannot be as easily linked to specific features of a system. We have included a description of the expected answers with examples in the evaluation guideline.

G2 - challenges/quests/missions/tasks: This item highlights an issue with user created content. While most raters saw the possibility for instructors to create challenges/quests/missions/tasks, some answered the question in the negative (because the system did not provide them on its own), some rated the item as *maybe* (as was intended for such cases), and some answered with *yes* because they saw the possibility for such structures. The presence of this gamification element is relatively easy to detect, the issue was with the standardized coding part of the instrument, which has been removed for version two.

M9 - time pressure: Disagreements in the *time pressure* category mostly stem from a disagreement over whether deadlines already constitute time pressure or whether one would need more immediate time pressure to consider this gamification element to be

[17]One could argue that they are an intermediate goal, as one may want to cause positive emotions in order to increase user engagement, for example. Our goals/means dichotomy does not provide for this distinction, however.

present. We have clarified the difference between long-term and short-term time pressure in the evaluation guideline.

G3 - clear goals: Similar to G2, the potential for clear goals to be set was sometimes rated as *maybe*, sometimes as *yes*. One rater understood the question as one of user interface design with a resulting divergence in ratings. We have clarified the meaning of this gamification design principle in the evaluation guideline.

E10 – relatedness: Relatedness was mostly interpreted by the raters to be dependent on the presence of social networking functions in the system. The amount of such functions needed for a positive rating differed between raters. Additionally, this item shows that different research efforts can have significant influence on ratings. In one system, communication between students was limited on purpose, but some raters did not realize this. Evaluators mostly correctly identified gamification elements that are part of the game design principle of *social elements* as relevant to *relatedness*. As such, the issue is mostly solved through the removal of the closed part of the inventory. We have nonetheless clarified the expected response in the evaluation guideline.

S1 – competition: Some raters already considered the gamification design principle of *competition* to be present or potentially present when they found the possibility to gain status — e.g. badges. Other raters looked for more direct ways of competition, such as leaderboards. Additionally, all but one rater missed the possibility to include a leaderboard in the Moodle LMS through a plug-in. We have highlighted the focus on competition between users in the evaluation guideline.

S2 – cooperation: Cooperation was seen as possible or present by some, when basic group communication functions were in place, while others required dedicated affordances for cooperative work. We have previously subsumed cooperation under the gamification design principle of *social elements*, as the two can be difficult to separate, but we do not consider all *social elements* to be indicative of cooperation. We have clarified the focus on dedicated affordances for cooperation in the evaluation guideline.

M4 - fixed rules: The interpretation of the fixed rules item was mixed. One rater looked for rules such as a definition of when a specific badge will be handed out, another for the option of assigning tasks to users. The other two mostly did not find rules to be present at all. We have already discussed the issue with the term rules in detail in subsection 2.1.1 and this will remain a problematic term. We have stressed the importance of explicit rules in the evaluation guideline.

R5 – progress: Progress, the final item over the variance threshold, sees mixed responses mostly due to different interpretations of the term. Some saw the attaining of rewards (e.g. badges) as visible progress, while others looked for a dedicated display of progress, such as a progress bar. Measures of progress are gamification elements that can be attributed to both the rewards and the goals design principles. We have emphasized the importance of dedicated elements that display progress in the evaluation guideline.

Overall, the variance in closed responses can be explained well through the analysis of the free-form responses and we have been able to find promising solutions for all problematic items. These results indicate that the low inter-rater reliability in the closed questions was not due to an innate weakness of the instrument or an incorrect choice of items but rather due to issues in the design of the instrument that could be fixed in a second version. Some variance also was expected to occur from the use of evaluators with only basic training in gamification and the resulting differences in the understanding of terms. The evaluation guideline we created should help reduce such misunderstandings. This part in our research both resulted in a valuable instrument for the qualitative analysis of gamification and further insights into the differences in expert opinions about the definition of gamification as well as providing the tools for and informing our further work towards measuring the effectiveness of gamification.

5 The Effectiveness of Gamification in Learning Management Systems

A large variety of learning management systems (LMSs) is available today. We used a ranked list of LMSs by Capterra in our research focusing on the top five LMSs that were aimed at education at school (Moodle, Edmodo, Blackboard Learn, Schoology, and Canvas) (Capterra, 2015). Functionally, these systems are similar, as seen for example in the comparison of Blackboard and Moodle by Machado & Tao (2007). While the choice of LMS is not an easy one and cost and benefit have to be assessed, Weaver et al. (2008) have shown that learners are less concerned with the system than with the way it is used, while teachers focused more on technical and administrative aspects. Similarly, de Smet et al. (2012) show perceived ease of use to be the strongest predictor for adoption among teachers. Schulmeister (2005) discusses the importance of didactics when using LMSs and McGill & Klobas (2009) shows the importance of task-technology fit in their adoption. It seems therefore important for an LMS to have a high degree of usability in order to be used by teachers, but that alone will not make the LMS an effective teaching tool. The use of the LMS has to be adapted to the content being taught and the students that it is being taught to. Our focus, however, is on the experience of the users of those systems, specifically gamification.

Whereas the previous chapter dealt with understanding gamification and identifying forms of gamification in existing systems, this chapter describes our evaluation of the effectiveness of specific implementations of gamification in LMSs. As we have shown in section 3.3, there is little scientific evidence of gamification being effective, and most claims of effectiveness come from practitioners with a vested interest. In order to be able to test the effectiveness of gamification as it is being used, we first need to understand how gamification is employed in LMSs. Simply comparing LMSs that are gamified to LMSs that are not would not be a sound research design for two reasons: For one, without proper analysis we would have to rely on either manufacturer or third-party claims to make that distinction, for another, different LMSs are sure to differ in more than just their implementation of gamification, introducing a large source of error. Therefore, the first step towards analyzing the effectiveness of gamification in LMSs was an evaluation of gamification in existing LMSs using our Gamification Inventory (see chapter 4). The results of that evaluation combined with the information gathered from our study of related work and knowledge of experimental design gave us a list of functional requirements for the design of our software.

Aside from the popularity of gamification in education, the main reason for choosing LMSs for our experiment is the large number of users under similar conditions. As such, we needed to find an LMS with a large user base that we could modify. Creating a new LMS would both exceed the scope of this work and be of little use, as it would be very difficult to recruit enough users for a new system. We found a partner in the German federal state of Baden-Württemberg, where the LMS Moodle is available to all interested schools. The LMS was extended with a competence grid and a mobile application for use with that grid was newly developed. This gave us the opportunity to add our own changes to the new addition and to have them centrally deployed. This resulted in additional requirements for our software development and experimental design. Even with such a partner, field access in schools is highly complicated (see subsection 5.2.6) and useful results would require long-term measurement. This chapter therefore describes the requirements for, and the development of our addition to the Moodle system used in Baden-Württemberg, as well as a 2x2 experimental setup for testing the effectiveness of the commonly used gamification elements *points* and *badges* in preparation for a large-scale experiment. We will successively develop the requirements for our application throughout the following sections, with a summarizing list of requirements and brief summaries of their origin in appendix E.1, followed by a description of our implementation and the suggested experimental setup.[1]

[1]As in the previous chapter, the individual steps of the process build on the results of the previous step, which is why we have refrained from splitting results into a separate chapter.

5.1 Qualitative Evaluation of Gamification in Learning Management Systems

The basic setup for the qualitative evaluation of gamification in LMSs was already described above in section 4.4 — four raters were asked to apply the Gamification Inventory to each of five allegedly gamified LMSs. In this section, we focus on the analysis of their responses concerning the actual gamification used in these systems, as opposed to the validation of the instrument discussed in section 4.4. Version 1 of the Gamification Inventory, the one used here, asked not only for free-form responses but also for evaluators to rate their response as either indicating the presence of each respective property of gamified systems (yes), its absence (no), or as being more complicated (maybe). We used these categorizations as a guideline to the analysis of the free form responses. If an item was marked as *yes* for a system by the majority of experts, we then analyzed the responses to see what property of the analyzed system lead to the experts' responses. Similarly, qualitative responses to items coded as *maybe* gave insight into the reasoning behind that categorization. In order to properly analyze the qualitative responses, with a focus on finding similarities and differences in evaluations, we coded those responses according to the mentioned functions of the LMS.[2]

The application of the Gamification Inventory did not only provide us with information about the quality of the instrument, but also about the use of gamification in popular LMSs as a prerequisite to setting up a proper evaluation of the effectiveness of that gamification. The following sections summarize the results from our findings in the analysis of five LMSs with the use of the Gamification Inventory. These results were originally published in (Broer & Breiter, 2015). The results from all four evaluators can be found in appendix C. Our description of the results below is organized according to the five categories of the inventory.

[2]One repeating code does not directly reflect functionality, however. *Depends on instructor input* could often be found in *maybe* responses, showing the overall trend that the presence of many items of the inventory can depend on the content presented through the system rather than the system itself.

Experiential. The experiential items show a very high overlap of gamification in the five analyzed LMSs. Accomplishment, autonomy, learning/mastery and positive emotions were found in all five systems. All systems had the potential to provide the user with a form of accomplishment through the successful completion of tasks and courses and through the awarding of badges and grades. These factors are also the causes of positive emotions — none of the systems was found to include additional elements specifically designed to afford positive emotions. Mechanics often found in games that give players a feeling of accomplishment even when they are not successful are not obvious in the systems analyzed here. Challenge and competence were extreme examples of this, as they were uniformly coded as being potentially available but depending on instructor input.

Autonomy was reflected in the student's ability to choose when and where to complete their tasks. Depending on instructor input, they may still be bound to a strict syllabus, however. Learning/mastery was usually present in one of two ways: Users were allowed to re-take quizzes and assignments, or there was a dedicated mastery system. Such systems were usually based on formative assessment and the alignment of tasks to gaols. Affordances for curiosity, purpose, and concentration were not found in either platform. Figure 5.1 shows the results of the evaluation in this category.

Mechanics. The major gamification mechanics found in the different LMSs in the expert review were collecting, documentation of behavior, and player effort. Collecting can mostly be seen as an incidental result of the ubiquitous use of badges in these LMSs. Documentation of behavior was generally available in different forms. Some systems provided activity streams for the users themselves, teachers, or parents. Player effort was a difficult item to apply as the system needs to provide a measure of success that is distinct from outside measures of the same if it is to be coded in this way. This is directly related to the item of *variable outcome*. All systems provided variable outcomes in so far as users could get different grades and potentially fail to complete the course. These are hardly different from the same mechanic in traditional courses, however, and therefore hard to ascribe to gamification. Fixed rules were also rare beyond the norm provided by the frameworks of a school course and an information system. Blackboard Learn provided the ability to define rules for obtaining badges, which can be seen as similar to games. In all other systems, badges were

Figure 5.1: Results in the experiential category, showing the number of LMSs coded as expressing each subcategory (Yes), potentially expressing it (Maybe) and not expressing it (No). Results for all categories can be found in appendix C

handed out by the instructor, allowing for arbitrariness. Schoology and Blackboard Learn were the only systems found to offer time pressure. Here, instructors can set time limits for specific questions/tasks. A common element in many games, time pressure seems to be rare in LMSs. Nurturing/growing and storytelling are completely absent from the analyzed systems, even though both were thought to be highly relevant for gamification by the experts in the preceding survey.

Rewards. The rewards section is essentially identical for all five surveyed systems. All of them offer intangible rewards in the form of badges and some measure of progress. None made use of points or levels, which are equally commonly found in gamification literature. It should be noted that LMSs with such features exist; they were simply not part of the sample analyzed here. Progress was shown through a mastery system in some cases, through general course completion in others. Some LMSs allowed the sharing of badges outside the application through Mozilla's Open Badges program.

Goals. Goals were almost identically rated among all systems as well. In each case, instructors had the ability to subdivide coursework in such a way that the goals for the users are clear and potentially achievable. All systems show users in an understandable way, which next steps they should take, provided that the teacher has set up such steps. Some systems, such as Blackboard Learn, specifically encourage teachers to set up tasks that align with course goals, facilitating clear and achievable goals.

Social. The social category shows actual conceptual differences between the systems. Tools for collaboration were explicitly built into some systems (Blackboard Learn and Canvas), and possible in others through communication tools. Some systems, such as Edmodo, explicitly excluded communication between students from their features, while others even provided chat rooms and discussion boards (e.g. Blackboard Learn). Only Moodle allows for competition in the form of a leaderboard (through a third party plug-in), underlining an interesting result from the expert survey: Experts considered competition very important for a description of gamification in general, but most did not consider it important for their own chosen implementation of gamification. Finally, most systems did not make use of social incentives in general or social engagement loops in particular. Exceptions are Blackboard Learn and Canvas that provide notifications to users, when other users show activity.

Overall, we found the examined systems to be highly similar in their use of gamification. This is shown in the comparison of results from the standardized part of the instrument as seen in figure 5.1 as well as the qualitative evaluation that followed. With points and badges being ubiquitous and some form of numerical outcome common (be it points or grades), these two gamification elements are most relevant for testing for effectiveness. While it would be very interesting to test new experimental elements, the major issue we seek to address is the lack knowledge about the effectiveness of what is already being used. Those elements that were evaluated as depending on instructor input on the other hand are not useful for a large-scale experiment, as such an experiment needs to avoid uncontrolled variables wherever possible and varying instructors are such a variable. Testing these elements in an educational setting is not impossible, but would be much more promising in a massively open online course (MOOC) setting, were many users can be presented with the same content. Finally, as we have discussed in (Broer, 2015), many of the elements that we identified in the analysis are properties that our experts considered to be relevant for the gamification family (hence the inclusion in the inventory) but that have a clear non-game origin in LMSs and in fact are commonly used in education. This is a both a downside of and a reason for the family resemblances approach, as the properties shared among many gamified applications do not have to be exclusive to gamification. In fact, the failure to identify properties exclusive to gamification is what makes it so difficult to find a real definition of the term. For our experiment, elements, such as grades, that did not arise from the process of gamifying an LMS, are of little interest. Our evaluation of existing LMSs therefore informed our further experimental design in so far as that it should put a clear focus on the two gamification elements points and badges. Additionally, we found that the presence of gamification elements often depended on instructor input, which is detrimental to a proper experimental design with as few uncontrolled variables as possible. It is therefore important that the gamification elements added for our experiment do not depend on instructor input. Instead, they need to be usable without instructor intervention and should therefore focus on automated assessment of user input, also facilitating rapid feedback.

5.2 Measuring the Effectiveness of Gamification

As noted by Hamari et al. (2014) and underlined in our analysis of empirical studies of the effectiveness of gamification, many of those studies suffer from similar problems that make it difficult to ascertain, whether the gamification elements employed actually had the desired effects. For one, many studies did not employ control groups for proper comparison and/or relied only on self-reported measures such as the satisfaction with a new system. For another, the sample size in many studies is not sufficient to draw accurate conclusions. Finally, gamification was often employed in combination with other factors or multiple game elements were employed at the same time. Either makes it difficult to link results to specific changes. As a result, there are very few studies that actually help us to ascertain, whether gamification is has the desired effects or not. In order to help alleviate this lack of knowledge, we have designed and prepared a large-scale, controlled experiment to measure the effectiveness of certain approaches to gamification in LMSs.

5.2.1 Target Audience and System

As we have shown in our literature review, education is one of the major areas that gamification is employed in. At the same time, most studies of gamification in education focus on gamifying specific courses, which makes it difficult to reach a large target audience. LMSs, however, are usually employed across subjects and age ranges and are used by many different users. We therefore suggest an experiment with different forms of gamification in such a system. For this example, we will focus gamification in Moodle, according to Capterra (Capterra, 2015) the most used LMS worldwide in 2015.[3] While any LMS could be used for these purposes, Moodle has a few advantages. For one, Moodle is a modular open-source system, greatly reducing the difficulty of modifying it. For another, Moodle is centrally deployed in all schools in the German state of Baden-Württemberg, allowing for large number of participants with relatively little administrative effort. Expanding an already existing system removes a number of parameters that can otherwise complicate results. If users are already used to the basic system, responses to experimental changes should be much more pronounced than in a situation were users have to learn to use the system itself. New experimental features might not even be noticed by novice users. Using a new system would also add the potential parameter of the speed and intensity of adoption at the individual schools that could further muddle results. Finally, introducing new software to schools comes with large administrative and, more importantly, social barriers. Baek et al. (2008) discuss a number of international studies of barriers to technology use in classrooms and Breiter et al. (2013, p.260) state that a skeptical attitude of teachers[4] towards new media is not limited to older generations but rather inherent in their training. Any experiment that attempts to introduce new software to large-scale classroom use will likely run into many issues unrelated to gamification, significantly reducing chance of results.

As with any good experimental design, we aim to reduce the number of variable factors to a minimum. Technology adoption among teachers in one such factor, the intensity of usage another. LMSs provide many functions and it is usually up to the teacher to decide which of those functions to employ in his or her classroom. Whitmer et al. (2016) describes

[3]Moodle has been overtaken by Edmodo in Capterra's 2017 rankings.
[4]Their study dealt with elementary school teachers only, however.

five patterns of LMS usage, among them the supplemental archetype that is “high in content but with very little student interaction” and the evaluative archetype that makes “heavy use of assessments to facilitate content mastery”. The results of gamifying an LMS are likely to be quite divergent, depending on these usage patterns. In order to avoid this variable parameter, experiments should focus on a specific function of an LMS. In our example, we chose to implement gamification elements in a special feature that was added to Moodle for use in schools in Baden-Württemberg. A plug-in developed by the Austrian firm GTN Solutions integrates a competence grid[5] into Moodle. This grid is set-up to include all competencies that students are meant to acquire within a subject and allows students to rate their own competencies and submit proof therefor. Teachers can then confirm those competencies. Furthermore, Baden-Württemberg is now employing a mobile application called DAKORA[6] for easy access to these features. The competency grid introduces a level of standardization that allows for a good compromise between creating very generic gamification that does not refer to individual conditions and tailored gamification that would usually require the input of the respective teacher. Any form of gamification that requires teacher input in order to function (especially one that requires the teacher to have an understanding of gamification) would immediately run into the issue with differences between teachers described above.

Educational Standards and Competency-Based Teaching

As part of our chosen target audience and system, the competency grid plays an important role in our research. We will briefly describe the background behind using such a system here. The results of international educational comparisons, namely TIMMS, PISA, and IGLU at the beginning of the century have led to a shift in the German[7] educational system from a primarily input-oriented system to one that focuses more on the output as well (see Drieschner, 2009, p.10, 25ff; Kultusminister Konferenz, n.d.). The conference of the German ministers for education and culture (Kultusministerkonferenz, KMK) now publishes

[5]We briefly discuss the notion of competence based teaching below.

[6]Das Arbeiten mit dem KOmpetenz RAster

[7]While the aforementioned comparisons are international, our focus is on the german education system and especially schools in Baden-Württemberg. Therefore our discussion of competency-based teaching is limited to German schools and references German literature.

standards for education that list the competencies that students should have at certain stages of their education (Kultusminister Konferenz, 2016, p.9). The KMK publishes detailed documents that list and describe these competencies in addition to example tasks and solutions (e.g. Kultusministerkonferenz, 2012). Obviously, education is not changed by simply testing according to different standards, but educational institutions had to adopt in order to be able to fulfill the new standards (and to improve Germany's rather negative results in international comparisons) (Drieschner, 2009, p.63ff). German states have individual strategies of helping their educational institutions adopt. Baden-Württemberg, for example, publishes[8] educational plans (Bildungspläne) that include two-tiered grids of competencies (see e.g. Landesinstitut für Schulentwicklung Baden-Württemberg, 2016). Figure 5.2 shows an excerpt from such a competency grid, in this case for the subject German in grades five and six in Baden-Württemberg. The grid includes competencies (on the vertical axis) and different steps in achieving those on the horizontal axis. These steps (Lernfortschritte, LFS) build upon each other, e.g. creating and reciting meaningful texts (3b, LFS 1) is usually a requirement for clearly structuring one's own speech (3b, LFS 3). To further integrate the grid into teaching and especially autonomous learning, the LMS Moodle has been extended with a competency grid module that allows students to judge their own competencies, as well as to find and submit tasks associated with those and teachers to grade their students' progress in those competencies.

		LFS 1	LFS 2	LFS 3
Sprechen und zuhören	**1 Spielerisch sprechen**	Ich kann Standbilder bauen und besprechen.	Ich kann Erlebnisse und Haltungen szenisch darstellen.	Ich kann eine Gesprächssituation im Spiel dialogisch ausgestalten.
	2 Verknüpfung	Ich kann aufmerksam zuhören und mich auf gemeinsame Gesprächsregeln verständigen.		
	3a Miteinander sprechen	Ich kann die wesentlichen Aussagen eines Gesprächs erkennen.	Ich kann Sprachvarianten unterscheiden.	Ich kann situationsangemessen und addressatenbezogen kommunizieren.
	3b Zu anderen sprechen	Ich kann Texte sinngebend gestalten und vortragen.	Ich kann verschiedene Formen mündlicher Darstellung verwenden.	Ich kann meine Redebeiträge klar strukturieren.

Figure 5.2: Excerpt from the competency grid for German, grades 5 and 6 for Baden-Württemberg. (Landesinstitut für Schulentwicklung Baden-Württemberg, 2016)

[8] at http://www.bildungsplaene-bw.de

5.2.2 Experimental Design

The goal of our proposed experiment[9] is to measure the effectiveness of two common gamification elements (badges and points) on user activity. The effectiveness of each element individually could be measured with a simple one-factor experimental design, relying on an analysis of variance in order to estimate these effects (c.f. Montgomery, 2009, p.60ff). It is not unlikely, however, that an interaction between the two factors exists (suggested e.g. by the common use of both elements in gamification and by the interaction between gamification design principles shown in subsection 2.2.3). The common approach in such cases is to use a factorial design. Such designs do not only have the potential to measure the impact of individual factors, but also their interaction (c.f. Montgomery, 2009, p.162ff). Factorial designs also have the advantage of requiring fewer observations for the same precision of measurement where the measurement of main effects is concerned (c.f. Montgomery, 2009, p.166).

In our case, the factors in the experiment are binary — either badges or points are present, or they are not. This leads to a total of four different types of observations, or experimental versions, as described below. As we are dealing with the reactions of human users, a large number of variables can influence the reaction of each individual participant. Differences in the attitude of the users (e.g. towards school and technology), demographic differences, differences in access to technology, differences in school policy and teacher behavior, and difference in subject matter are only a few possible sources for variance between measurements. While attempts at eliminating sources of variance should be made, it is to be expected that there will be a large amount of uncontrolled variance in measurements. In theory we could record as many of these sources of variance as possible and use regression in order to eliminate their impact (see Wonnacott & Wonnacott, 1985, p.16). This clashes with our goal of an non-invasive experiment however, that is one in which no additional effort on the side of participating schools and teachers is required.

[9]It could be argued that we are describing a quasi-experiment here. Lazar et al. (2010, p.42) describe that the difference between quasi-experiments and experiments is the random assignment of participants. We are using random assignment, but on a per-school basis, not per individual. All other aspects of Lazar et al.'s requirements are fulfilled. We will use the term experiment throughout this research with the caveat of limited random assignment in mind.

It is therefore necessary to include a large number of subjects to reduce the chance for uncontrolled factors to influence the results significantly.

Large sample sizes should help reduce the chance of erroneous conclusions, but as we do not have access to an infinite number of subjects, measures need to be taken to control as many external factors as possible. Differences between individuals could theoretically be controlled with a within-subject design — that is, each participant would be subjected to all four experimental conditions and the differences in measurement calculated for each individual. Unfortunately, this would introduce bias as experiences from earlier trials would impact the outcome of later trials and the timing of the experiment would also be introduced as a potential systematic bias. Especially in the case of schools, different intervals of the school year are likely to show different behavior from students.[10] We have already discussed the overjustification effect above and it is quite possible for the use of a system to be lower after the removal of a feature than it would have been without its introduction. We will therefore have to rely on one experimental condition per participant and compensate for individual differences through sample size. (see Lazar et al., 2010, p.46ff)

With such a within-group design unavailable, the assignment of participants to experimental conditions becomes relevant. Ideally the composition of all groups should be identical for all possible factors. Since this is impossible, random assignments in combination with a large sample size can be used to reduce the chance of error. Theoretically, group membership could be assigned randomly on a per-person basis. In our case it is likely that this will impact experimental results, however, as the participants are likely to communicate with one another. Seeing that other students have access to features that they themselves do not get may reduce the motivation for use in some participants, for example. Additionally, the status effects of points and badge conditions have very little meaning if students are unable to compare themselves to others because they do not share the same experimental condition. Instead of a per-person basis, assignment could also be made on the basis of classes, grades, subjects, or schools. Each of these introduces its own potential bias (e.g. age as a factor, or interest in certain subjects). We chose the option with the lowest chance of contact between students of different experimental conditions, a per-school assignment,

[10] e.g. the last weeks before a school year ends tend to be filled with exams.

accepting the included potential for bias. (see Lazar et al., 2010, p.57ff)

The experiment is therefore set-up in a 2x2 between groups factorial design. We have developed two approaches to gamification, meant to be applied to a quarter of the participants each. Both approaches should be activated for a third quarter and the final quarter is meant as a control group with no visible changes to the standard application. Assignment to these experimental groups should be done randomly, on a per-school basis.

Experimental version A extends the mobile application DAKORA with the capability to earn badges for actions within the system. There are two major variations of badges available — those that reward activity and those that reward competence. Badges that reward activity are awarded, for example, for uploading proof for an acquired competence or adding tasks to the calendar. Badges that reward competence require confirmation of competence by the teacher. One badge, for example, is awarded for correctly assessing one's own competence — that is, a student claims a competence through the application that is then confirmed by the teacher. Other examples are badges for completing an area of competence in the grid or achieving all competences within a course. Whenever a badge is acquired, a pop-up will be shown to the student with the badge's name, picture, and description. Badges are also available for review in a designated area of the application. This experimental version is designed to give infrequent, but strong positive reinforcement as well as giving the students goals to work towards. Each individual badge should provide meaning to the student and motivate him or her to continue his or her efforts.

Experimental version B extends DAKORA with the ability to earn points and levels. Most actions within DAKORA (e.g. rating a competence or adding a task to the calendar) as well as the confirmation of competences by the teacher will grant a small number of points. Point gains are immediately shown in the application with an animation, but do not require confirmation and are not shown prominently. Once a certain amount of points is reached, the user will gain a level. This event is shown in a pop-up similar to the one used for badges in experimental version A. The current level and the progress towards the next level can be checked in a designated area of the application. This experimental version is designed to give frequent and immediate feedback for desired actions within the application as well as a measurement for progress not directly connected to the attaining of competencies.

Table 5.1: A 2x2 experimental design

	no badges	badges
no points	D	A
points	B	C

Experimental version C is a combination of versions A and B — both versions are active at the same time and students gain both points and badges. The gamification page of the application shows both the acquired badges and the progress towards the next level.

Experimental version D is the control group, neither modification is active and students see no difference to the regular application. We suggest that each experimental version should be in use at at least two schools to reduce the potential impact of the school as a factor, for a total of at least eight participating schools in the experiment. Obviously, a higher number of schools will lead to a better protection against any bias introduced through the differences between schools. Standardized curricula and competency grids should go a long way towards reducing such bias, but factors such as the promotion of the system at individual schools will still impact results. The total number of students in each school varies, therefore the experimental groups will not have the same number of participants, but they should be similar. The high number of classes (and therefore teachers) involved should decrease the chances of a strong skew in either direction, but significant variations in user numbers are still possible with this setup.

This experimental design comes from a strictly systemic perspective (see section 4.1) investigating the effectiveness of individual means on system usage in conditions A and B and a potential synergistic effect in condition C. Condition C can be seen as a first step towards measuring the effectiveness of dynamics and not just mechanics from an MDA perspective, but only in a very limited way. (For one, users are not actually a controlled condition and we can therefore not infer any results about the impact of mechanic-user interaction.)

The timing of such an experiment will naturally vary slightly. Literature suggests that short-term studies are not sufficient (see e.g. Hamari et al., 2014; Mekler et al., 2013). The shortest time frame that seems sensible for such an experiment in a school context is be the semester. Anything less can hardly be considered long-term and the nature of a school semester also makes it imperative to include it in full.[11] Longer periods are likely to improve results, although it seems sensible to use increments of full semesters. The following section describes how the data for our proposed experiment can be collected and which data that should be. We discuss the analysis of those data further in subsection 5.2.4.

5.2.3 Data Collection

Previous empirical studies of gamification have measured a large variety of different outcomes. These measures can be roughly distinguished into two categories — measuring changes in behavior and measuring target outcomes. We have discussed this in more detail in subsection 2.2.1 for gamification in education, describing the model of gamified learning by Landers (2014) in which learning was the target outcome, influenced (among others) by gamification elements. Learning is a common target outcome in education, but other gamified systems are targeted at other outcomes — such as increased user fitness in a gamified sports application. While measuring the target outcome is certainly the most useful measurement when creating a system aimed at that outcome (c.f. Mayer, 2014, for a discussion of measurement in game-based learning), it is less useful when attempting to measure the effectiveness of gamification (or certain gamification elements).

Mayer (2014, p.37ff) takes a strong stance on appropriate measures in the research of games for learning.[12] To him, academic learning outcome is the most important outcome of employing an educational game Mayer (2014, p.38). His arguments are strong and he describes an issue that can also be found in gamification research in many cases. There is a strong reliance on subjective, self-reported measures of factors such as enjoyment or motivation in both game and gamification research that does not help answer the question of whether the object of the research is actually effective at reaching its goal. That said, Mayer's strict focus on learning outcomes can be problematic. For one, education has more goals than just academic learning. But even if we focus strictly on academic learning, other

[11] Holidays and exam periods for example are likely to have an impact on LMS usage.

[12] While Mayer writes about games for learning instead of gamification, we consider his work to be pertinent here, especially as similar work on gamification in education does not exist yet.

measures can certainly be helpful and interesting. One reason for this is the difficulty of measuring learning outcome appropriately. For example, Liu et al. (2012) have shown that motivational conditions have a substantial effect on performance evaluations. Worse, the validity of many instruments for assessing learning outcomes is questionable at best, especially in the case of low-stakes assessment (see e.g. Wise & DeMars, 2005), such as a post-intervention questionnaire. Furthermore, game-based or gamified learning approaches have a tendency to foster practical learning as opposed to the improvement of conceptual knowledge (see e.g. de-Marcos et al., 2016). While Mayer is certainly correct that learning outcomes are a primary objective in games for learning, their assessment is difficult and any chosen form of assessment has the potential to hide benefits or to exaggerate them. One should therefore certainly assess games for learning for their learning outcomes, but be careful with construct validity and include additional measurements as well.

This research, however, is not so much concerned with game-based learning but rather with the gamification of parts of the learning process. From a purely scientific perspective, the main goal of this study is to evaluate the effectiveness of various approaches to gamification and learning is just one possible context of many. A focus on learning outcomes would not only assume that these are the primary goal of gamification in an educational context, but also remove any possibility for a transfer of results to other domains.

In his theory of gamified learning, Landers (2014) proposes that the effect of gamification on learning is less direct than that of game based learning anyway. Whereas in the context of game based learning[13] instructional content and game characteristics combine to form the game cycle that produces learning outcomes, the impact of game characteristics on learning outcomes is only indirect in gamification. According to Landers, game characteristics mediate behavior/attitude which, in turn, mediates learning outcomes and moderates the mediation of learning outcomes by instructional content. This assumes that there are behaviors that can impact learning outcomes and that these behaviors in turn can be influenced through game characteristics. Even if learning outcomes are to be considered the primary objective of gamification in learning, any measurement thereof would only indirectly measure the impact of gamification as such. Gamification may, for example,

[13]Landers speaks of serious games.

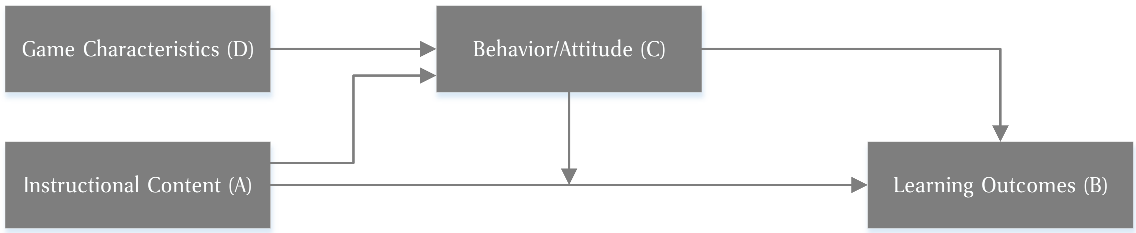


Figure 5.3: Landers' theory of gamified learning, after (Landers, 2014). Behavior and attitude serve as both a moderating and a mediating intermediary between game characteristics and learning outcomes.

be successful in promoting a certain behavior but that behavior may then not result in an actual improvement of learning outcomes.

With a focus on gamification, and not learning, it is therefore crucial to measure not learning outcomes but changes in behavior. This avoids the pitfalls of learning outcome measurement and at the same time provides more direct information about the effectiveness of gamification. We do not know if causing students to increase their activity in the DAKORA application will cause increased learning, that is for others to find out.[14] We concern ourselves with the question, whether gamification can increase said activity and whether different gamification elements differ in that effect. Data collection in the experiment therefore needs to focus on activity data, concentrated on those activities that are actually targeted by the gamification intervention. Other behavioral data could certainly enhance the results of the experiments, such as self reports of intrinsic motivation (using, for example, the intrinsic motivation inventory (see e.g. Ryan et al., 1991)) or assessments of the quality of student submissions, as employed by Denny (2013). Either would greatly increase the difficulty of a large-scale experiment, however. Any non-automatic data collection (such as questionnaires that students have to fill out) must be considered an intrusion into the natural use of the system, skewing results and reducing the willingness to participate. The large-scale analysis of the quality of student contributions would require a high amount of expertise

[14] For example, Mabed & Köhler (2012), in a study of the LMS OLAT, could not find a correlation between LMS usage and academic achievement or attitude towards learning, citing, among other factors, the content of the course as a potential reason. When studying the impact of LMSs from a task-technology fit perspective, McGill & Klobas (2009) found a moderate effect on students' perceived learning but only a weak impact on student's grades.

and manpower, further complicated by different levels of education. We suggest that a large-scale quantitative experimental setup such as ours should eschew such measures. Results could be complemented with additional, small-scale studies focusing on qualitative aspects.

Although rarely found in the analysis of gamification and serious games, there are other outcomes that are worth evaluating. Conati & Gutica (2016) analyzed the emotions of players of an educational mathematics game as coded by expert observers. The authors found many occurrences of engaged concentration and confidence, followed by confusion. They point out that they found only few occurrences of pride, shame, and curiosity even though those are supposedly "important for learning in educational research" (Conati & Gutica, 2016, p.29). This could be seen as a confirmation of Mayer's claim that measuring learning outcomes is more helpful than measuring intermediate outcomes. On the other hand, an understanding of the emotions caused by games or game elements may give us more immediate information about their effect than a raw measure of learning outcomes.

All data collection for the purposes of our proposed experiment can be done through our plug-in for Moodle. Every action that has the potential to lead to either the awarding of points or badges is stored in a database table (gamification_log). Each entry consists of the type of action taken, the id of the user that took it, the date and time that it was registered at, and a data structure that contains additional information depending on the action.

Additionally, all data generated by the Moodle plug-in through its usage is available for evaluation as well. Database tables of relevance here are the one that stores user badges (gamification_badge) as well as the one that stores the points for each user (gamification_points). In the case of Baden-Württemberg, each school uses its own database with Moodle. All data is delivered on a per-school basis and can easily be attributed to each experimental group. All data stored is pseudonymized through the user id, which is not publicly available (according to requirement 9 below). While Moodle's database contains information that allows linking user ids to natural persons, the database tables used for evaluation here do not and they are the only ones that need to be made available to the researchers, alleviating privacy concerns. The following section provides a suggestion for the analysis of the quantitative data that our experimental setup provides.

5.2.4 Data Analysis

The data collected in the suggested experiment are purely quantitative in nature and thus require quantitative analysis. We hypothesize that the various forms of gamification employed could increase both the amount of user participation as well as its quality. The amount of user participation is measured by the amount of events triggered by each individual user within the experimental time frame.[15] The quality of user participation is difficult to assess automatically. We therefore suggest the use of the teachers' evaluation of student competencies as a measurement for the quality of submissions. This seems appropriate, as the intended use of the system is that the students submit material through DAKORA as proof of their competence. Submission quality is a combination of the amount of competencies gained and their level. We suggest the hypotheses shown in table 5.2 for testing in the experiment. One should test whether each gamification element separately causes an increase in quantity or quality of submissions compared to the control group as well as compared to each other and to a combination of both.

Table 5.2: Hypotheses to be tested in the experiment. Each hypothesis has the following form: Awarding [condition] increases user participation [outcome] compared to the [comparison] group.

#	condition	outcome	comparison
A1	badges (A)	quantity	control (D)
A2	badges (A)	quality	control (D)
B1	points (B)	quantity	control (D)
B2	points (B)	quality	control (D)
B3	points (B)	quantity	badges (A)
B4	points (B)	quality	badges (A)
C1	both (C)	quantity	control (D)
C2	both (C)	quality	control (D)
C3	both (C)	quantity	badges (A)
C4	both (C)	quality	badges (A)
C5	both (C)	quantity	points (B)
C6	both (C)	quality	points (B)

[15] The analysis described here does not take changes in behavior over time into account as it is designed for the minimum time frame of a single semester and the minimum number of particiating schools. Under these conditions, data will not be robust enough for a proper analysis of changes in system usage over time as individual factors are highly likley to skew those results. Under different circumstances, especially a much higher number of participating schools, a time series anaylsis would become very interesting. The software is already set up for such analysis, as all data are recorded with timestamps.

The chosen measurements should allow relatively simple statistical testing of the aforementioned hypotheses. With an unknown distribution of data, we suggest using the Mann-Whitney U test (see Mann & Whitney, 1947), a non-parametric test of the equivalence of the distribution of two independent sets of data, as employed for example by (Denny, 2013). If a significant difference is found, a simple calculation of the median of both sets of data should suffice to test for the hypothesized increase of quality or quantity. Depending on which of the 12 hypotheses above are confirmed, this will help us understand whether the employed gamification elements increase desired user behavior compared to the control group (hypotheses A1 and B1), which one has a higher impact on that behavior (B3), and whether combination of the two leads to an even higher impact (C1, C3, C5). The corresponding even-numbered hypothesis deal with the quality of student contributions depending on the gamification elements employed, as measured by teacher ratings. For a pure measurement of the impact of gamification, all measures of quality could be dropped, as discussed in subsection 5.2.3. From a practical point of view, however, it is interesting to see whether adding gamification elements to an LMS can actually help increase learning[16] (whether mediated by behavioral factors as suggested by Landers (2014) or not) and the system already provides basic data for submission quality. Past results would indicate that we are likely to see an increase in quantity rather than quality of submissions with these basic gamification elements, but the reverse is theoretically possible. In the unlikely case that the only significant effects are those in quality of submissions, a focus on the odd-numbered hypotheses would hide that fact. Similarly, a potential decrease in submission quality caused by students merely trying to gain rewards through submission quantity would be hidden by such a design.

The analysis suggested here is purely summative in nature and does not take the time distribution of activities into account. Given a relatively short experimental time frame of one semester of study and the minimum number of participating schools, other analyses are unlikely to provide significant results as they would introduce additional biases. If one is able to increase either the number of participating schools or the time frame that data

[16]Again, actual measurements of learning outcomes would be much more complicated. An increase in submission quality could also be seen an indicator of another behavioral factor, such as the amount of time spent on producing high quality submissions. Nevertheless, better teacher ratings of student submissions would suggest that learning may have been increased.

is collected in — or ideally both — large numbers should help to average out individual differences in timing in different schools and classes. Snow et al. (2015) describe three types of methods for dynamic analysis of log data in educational games — random walks, entropy analysis, and hurst exponents — that could be adapted for the analysis of our log data as well. Access to a sufficiently large number of participants would also help in answering the question of different types of users of gamified systems that we have discussed in section 3.3. Cornforth & Adam (2015, p.152) used cluster analysis to find different types of profiles in the game Minecraft, claiming that such analysis allows "(1) to systematically investigate how learning depends on individual characteristics and (2) to adjust the learning approach accordingly". In the case of gamification of LMSs, an understanding of how different types of users react to different gamification elements could be very helpful.

Following these more theoretical preparations of our proposed experiment, the following two sections deal with the practical implementation thereof: The next section describes the application we developed, followed by a description of the issues we encountered with field access in education.

5.2.5 Gamification Plug-in for DAKORA and Moodle

Our previous research resulted in a list of requirements for a software system to use in our proposed experiment. Appendix E.1 lists those requirements and explanations for them as extracted from our previous work. We have developed a full-functioning plug-in for experimental use with a combination of the Moodle LMS with the Exabis Competencies plug-in and the DAKORA mobile application based on these requirements. Exabis Competencies integrates the competence grid functionality (see section 5.2.1) into Moodle, while DAKORA provides a mobile application that interfaces with it. Exabis Competencies is used for most administrative work, such as setting competence data and tasks, while DAKORA is meant for the daily use of choosing and completing tasks and rating competencies. Moodle (and by extension its plug-ins) is built in PHP and uses a Mysql database for data storage. Moodle has a modular design with an application programming interface (API) that allows for extension through plug-ins. Plug-ins are separate pieces of software that can be installed on any

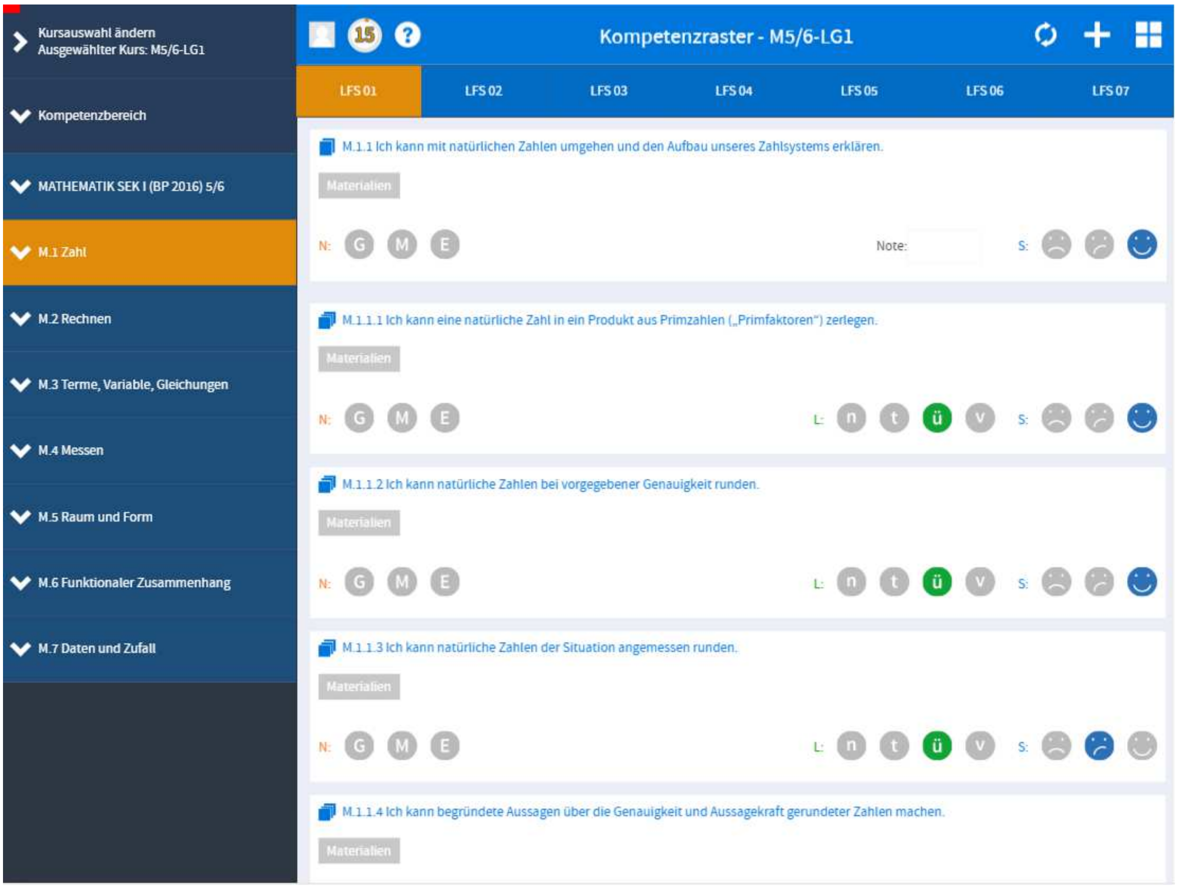


Figure 5.4: Screenshot of the competence grid for DAKORA with gamification enabled.

Moodle installation, as long as the respective versions are compatible. Moodle employs an events system that allows for communication between plug-ins and the core of the software.

DAKORA is a mobile application built in Javascript that uses Apache Cordova to deploy the application to various platforms. At its core, DAKORA is still a server-based application, but a lot of the computation is done in the client instead. The lapstone framework that DAKORA is built on is modular in nature and easily extendable. Like Moodle, DAKORA uses an events system that allows plug-ins to integrate seamlessly with the core of the application.

Our software is designed to provide basic gamification elements to DAKORA that do not require any customization on the side of instructors or schools. In order to facilitate factorial experimental designs, individual gamification elements can be separately enabled on a per-installation basis. The current version supports the use of badges and points (with levels), but a flexible events system allows for an extension with other gamification elements as

Figure 5.5: Screenshot of the progress screen that our plug-in adds to DAKORA if badges are enabled.

desired. A set of predefined badges is included as well as triggers that define when badges or points are earned. Changes to these rewards are intentionally limited to the source code of the application, as instructor intervention would defeat the purpose of the experiment. An eventual productive version of the software would benefit from a graphical user interface that allows for changes in the employed gamification elements without the need to edit source code.

System architecture

The additional experimental versions of the Moodle/DAKORA ecosystem required modifications on both sides. We developed a plug-in for Moodle that is responsible for the storage of data as well as all computations that need to be tamper proof. The plug-in has no display functionality and almost no server-side configuration. Upon installation, the plug-in adds the experiment specific badges to Moodle. These are then accessible through Moodle's configuration and could be modified by the teacher, although this is not intended behavior. If the plug-in is not installed, DAKORA users will not see any gamified functionality in the application. The main functionalities of the Moodle plug-in are shown in table 5.3

Table 5.3: Main functionalities of our Moodle plug-in and references to the corresponding requirements (see appendix E.1).

#	Functionality	Req. #
1	Storing a log of all experiment-related events in a database table (gamification_log)	8.4
2	Registering all experiment-related events sent from DAKORA and taking appropriate action	8.2
2.1	Awarding badges if their conditions are met. (E.g. a specific action has been done a specific amount of times.)	1
2.2	Awarding points if their conditions are met (E.g. a student has graded him or herself.)	2
2.3	Counting actions (mostly relevant for the awarding of badges .)	3
3	Returning results of actions to DAKORA (e.g. points or badges gained.)	8.3
4	Registering experiment-related events in Moodle (e.g. a competence being confirmed by a teacher.)	8.1
5	Storing events that happened in a user's absence (e.g. a badge being awarded due to teacher action.)	8.1

All visualization and direct interaction with the modifications is done through a plug-in for DAKORA. The event system along with DOM manipulation with jQuery allows the plug-in to modify all pages within the application. Events can be registered and responded to independent of the area the user is currently in. Table 5.4 shows the main functionalities of the DAKORA plug-in.

Table 5.4: Main functionalities of our Moodle plug-in and references to the corresponding requirements (see appendix E.1).

#	Functionality	Req. #
1	Sending of experiment-related events to Moodle	8.3
2	Receiving and displaying responses from Moodle	8.3
2.1	Displaying badge and level gains as a pop-up	1.1
2.2	Displaying point gains in an unobtrusive animation	2.1
3	Periodically checking Moodle for new events	8.3
4	Displaying badges and level progress on a separate page	3

The basic data flow can be seen in figure 5.6. Moodle serves as an intermediate for all communication between the server and the client as well as that between the plug-in and the database. The former is handled through JSON encoded data. Most actual computation happens in the trigger-response loop within the gamification plug-in for Moodle, the DAKORA plug-in is mainly responsible for informing the server of client side event and displaying responses appropriately.

The differentiation between experimental groups is done through a simple flag in Moodle (requirement 4). All users will be using the same software, but DAKORA will request the experimental setting from Moodle and alter its behavior accordingly. This means that the same data is available for all four experimental groups, they will simply see different parts of it (requirement 5). All data that determine the rewards awarded to the users are stored in the Moodle plug-in. In addition to the badge information that is transferred to Moodle upon installing the plug-in, all logic that determines the awarding of badges and points is embedded in the plug-in itself. All awards are triggered through either listeners or watchers. Listeners are methods that can be registered to be called when a certain event is triggered — this can be either a specific Moodle event or an event from DAKORA. Watchers function similarly, only that they are not bound to a specific triggering event but instead check for a counter to reach a certain number. A watcher could, for example, look for a certain action having been taken in DAKORA 30 times and then award a badge for that. Watchers and listeners can be as simple as awarding points or a badge whenever they are called, but they can also include arbitrary additional code. This can be used to create more

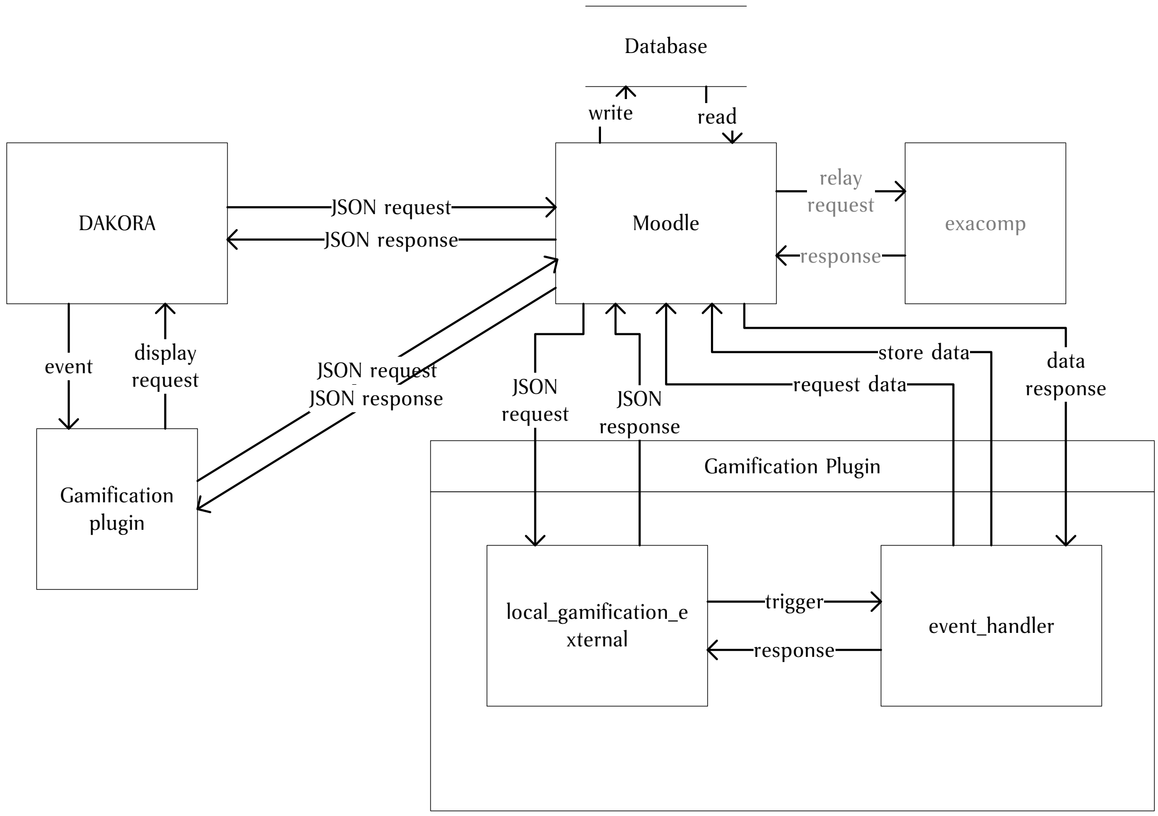


Figure 5.6: Diagram of the data flow between Moodle and DAKORA. Exacomp, the competence plug-in used for DAKORA, is only shown for reference. It is not directly used for our plug-ins, but provides the competence grid functionality they are based on.

complicated conditions, such as a badge awarded for attaining all competencies in three different courses or for completing 5 tasks in a single day (requirement 11). The watcher/listener setup uses dynamic method calls for easy extension. Watchers are only passed the id of the user in question, while listeners are passed the triggering event (which includes additional information, see figure 5.9). We have included an explanation of how to edit these gamification elements in appendix E.3.

Figure 5.7 shows a sequence diagram reflecting the process triggered by a student rating him or herself in DAKORA. As one can see, most computation happens in the Moodle plug-in, although the parse_return() function of the DAKORA plug-in is relatively complicated, handling the proper display of responses to the user.

Figure 5.9 shows the class diagram for the Moodle plug-in. The two major classes are local_gamification_external, which handles the communication with Moodle and, through

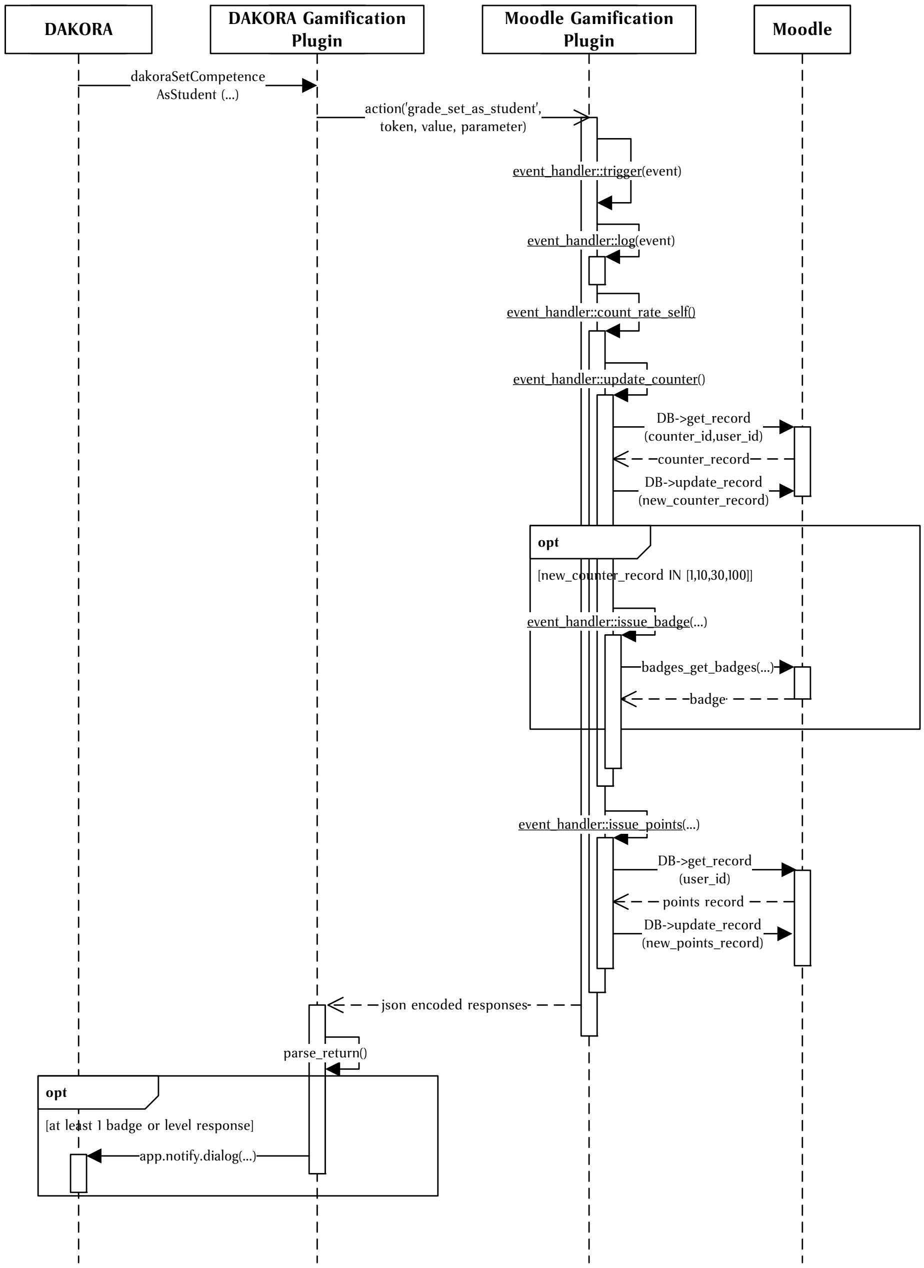


Figure 5.7: Sequence diagram of an example interaction with DAKORA. In this case, a student rated him or herself. Only the plug-in functionality is shown in detail and some simplifications have been made for readability's sake. Own illustration.

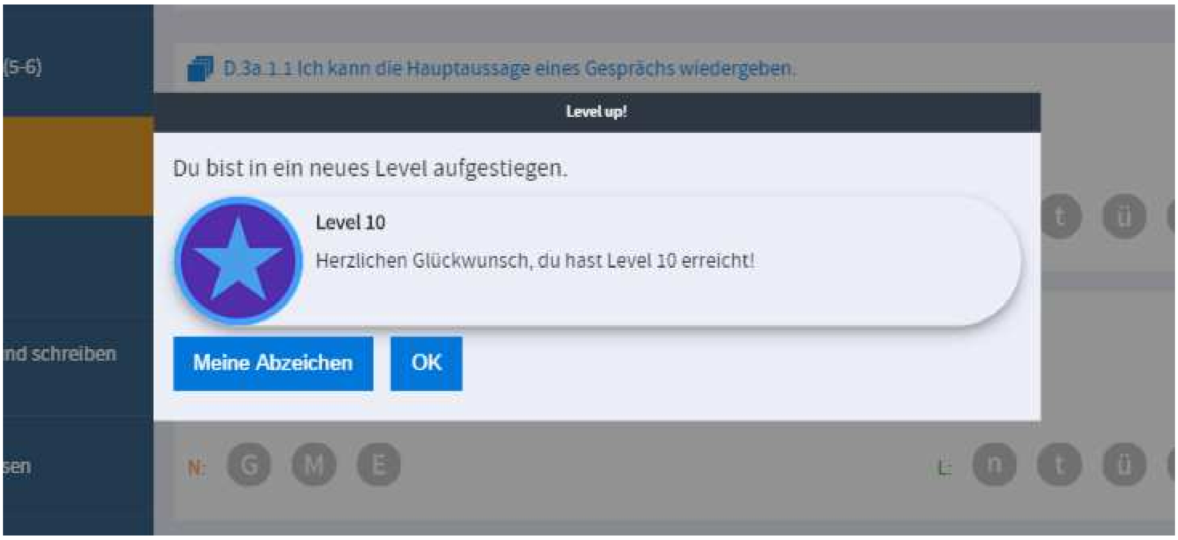


Figure 5.8: Screenshot a pop-up in DAKORA, announcing that the student has reached a new level.

it, DAKORA, and \local\gamification\event_handler. Note that most classes reside in the local\gamification namespace and will be referred to simply with their class names from here on out. The event_handler encapsulates the watcher/listener system describe above. The responses, response, and event classes are used within that system. The badger class is unique in that it is not used in the normal process of running the plug-in at all, but only invoked on installation or upgrade to insert the required badges into Moodle and to store a mapping of Moodle badges to DAKORA badges (requirements 8, 8.3). Badges are referenced by an id number in Moodle and these numbers can be different depending on the system that the plug-in is used in. In order to still identify the correct Moodle badge for each DAKORA badge, this mapping table is needed. The event_handler's get_badge() function accesses this mapping to provide the correct Moodle badge for each plug-in-internal id.

The plug-in for DAKORA is written in JavaScript and comprises two main components — plugin.gamification.js and page.gamification.js — as shown in figure 5.10. This follows the general structure of DAKORA, which uses pages as a representation of the individual screens within the application and plug-ins to handle all other functionality. Plug-ins can manipulate pages through DOM (document object model) manipulations using the jQuery library. Most of our added functionality is handled by plugin.gamification.js, allowing us

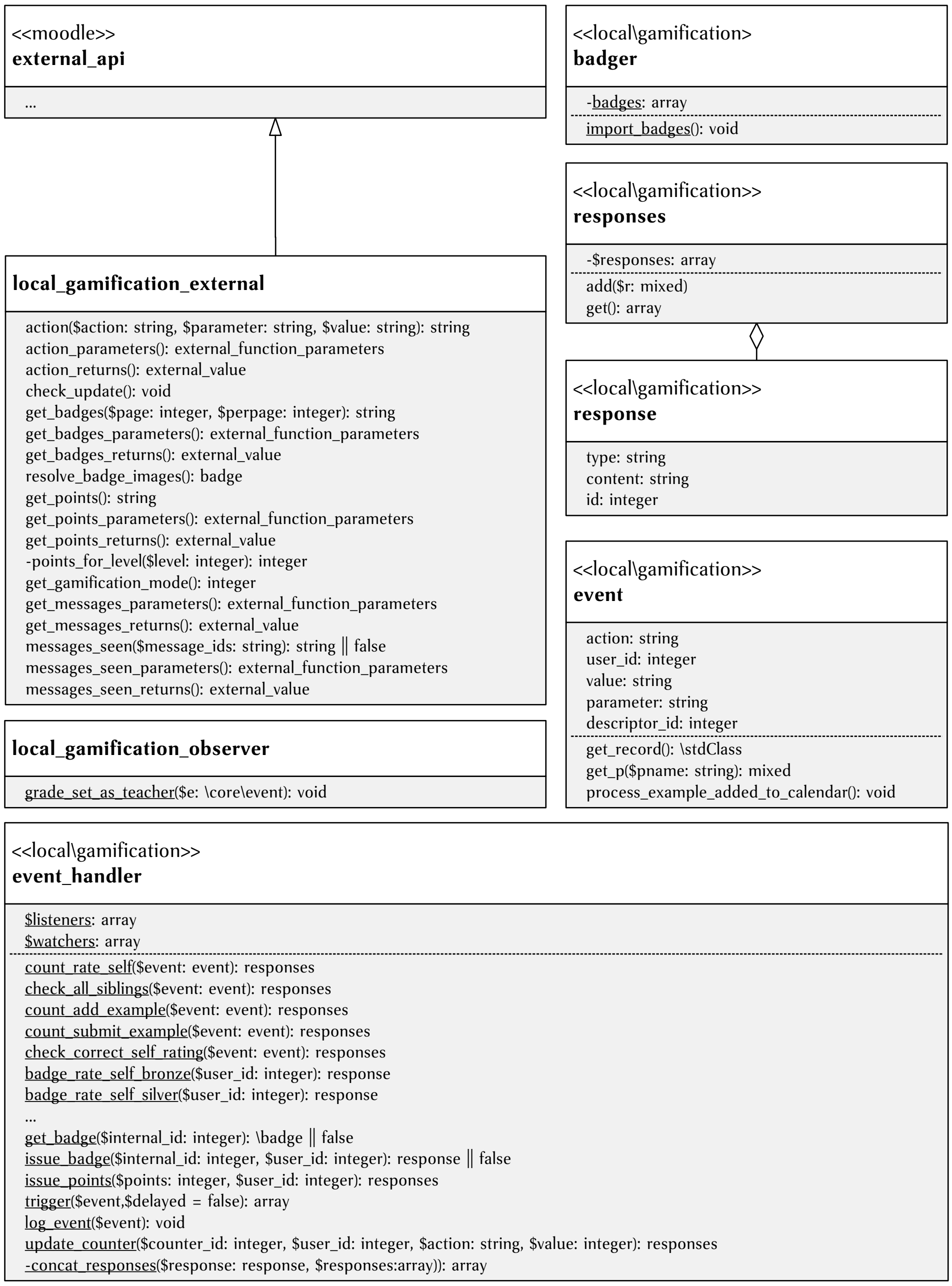


Figure 5.9: Class diagram of the Moodle plug-in. Includes some simplification for readability's sake. Own illustration.

to track actions throughout the application and to alter the display of arbitrary pages (requirement 10). We mainly use this to display the level and progress button at the top of pages with an upper menu bar, to show point gains in the header and to show message pop-ups if certain responses are returned by Moodle. page.gamification.js is only accessible if the badge condition is enabled and simply displays the badges a user has attained and, if points are enabled, a progress bar towards the next level. No gamification-related data is available to other users, both to protect privacy and to prevent competition (requirement 7).

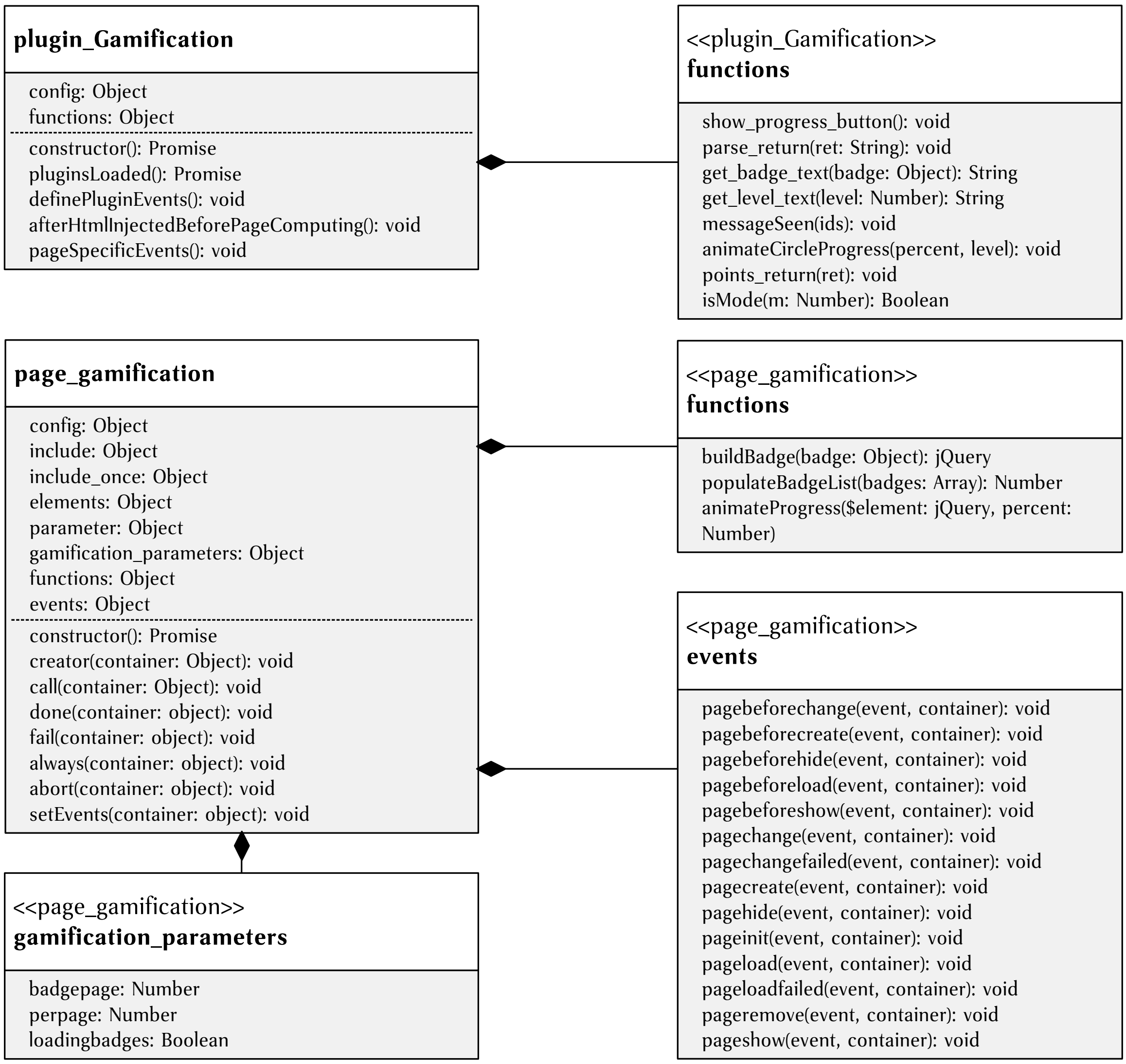


Figure 5.10: Class diagram of the DAKORA plug-in. Own illustration.

5.2.6 Field Access

One big issue in the measurement of the effectiveness of gamification is the number of participants. Our proposed experimental design attempts to alleviate this by using an established system with a large user base as a basis for our implementation of gamification. The German state of Baden-Württemberg offers the LMS Moodle as a centralized service for all public schools[17] in the state, 3665 schools with slightly more than one million students (Statistisches Landesamt Baden-Württemberg, 2016). At the time of our inquiries, the competency grid plug-in for Moodle that we extended for our experiment had already been requested by and installed at 64 of those schools. While we do not have individual data for those schools, extracting an average number of students per school from the numbers above yields an estimate of almost 18,000 students at those schools. This provides a large potential user-base, greatly exceeding the minimum numbers required for our proposed experiment.

In the case of schools, regulations — especially those regarding the protection of privacy — suggest a top-down approach to recruitment of participants. Instead of directly approaching schools, it is advisable to acquire permission for the experiment from the governing authorities first, before approaching the schools. Additionally, a top-down approach with support from the state government is likely to increase the chances of successful recruitment. We sought such support in preparation of our experiment through the ministry of education and culture in Baden-Württemberg and, more directly, through a group of educators tasked with the further development of Moodle in the state[18]. This allowed us direct access to the developers of the software as well as information about requirements for the software and schools that had agreed to employ it. The issues with field access do not stop there, however. Even with the support of the ministry and their acceptance of our setup for privacy protection, recruiting schools proved problematic. Once we had assembled the support and a list of contacts at schools that had requested the competency grid add-on for their installation, we attempted to recruit those schools directly. A description of the research project and the proposed changes to Moodle was sent to these schools via e-mail and fax, including a statement that the schools and teachers would not have to invest

[17]Private schools can also request participation.
[18]Called “Konzeptgruppe Moodle”.

any additional time into participating in the experiment. We considered this to be very important in oder to motivate teachers to participate and designed our software accordingly.

Unfortunately, response rates to our inquiries were very low with only 12 schools responding at all. Worse, the answers painted a rather bleak picture of the reality of software usage at schools: None of the schools that responded to our enquiry actually used the competency grid add-on for Moodle and some even indicated that Moodle itself was barely used at their schools at all. This is an example of a large problem that occurs when using distributed organizations, such as schools, as a means of field access for experiments with technology. The decision to adopt a technology at the highest level of the hierarchy does not immediately lead to adoption at the individual level. In our case we have the decision to adopt Moodle and the competency grid at the ministerial level which then has to be followed by the decision of individual schools to adopt the technology, followed by the decision of individual teachers to do so. The 64 schools that had decided to adopt the competency grid add-on are therefore not identical to a list of 64 schools in which teachers actually use said technology. Straub (2009) describes different models for the adoption of technology and specifically mentions the reluctance of teachers to do so. He discusses "deep-seated beliefs and identity structures [that] can lead to resistance to change and to acceptance of innovation" (Straub, 2009, p.633).

While our example is a problem very specific to schools and the particular software we chose to extend, it does indicate a more general issue. On one hand, top-down access is almost required if one wants to do a large-scale experiment, on the other hand the top-down perspective can often be very different from what is actually happening at the individual level. This problem is especially pronounced in a system with a large number of relatively autonomous individual deciders, such as schools.

6 Conclusion

We presented the results of a series of successive steps of research that build upon each other, with the goal of improving our ability to analyze gamification and its effectiveness. The final results are an instrument for the qualitative evaluation of implementations of gamification, the Gamification Inventory, and an experimental design including the software necessary to perform an experiment for assessing the effectiveness of gamification in learning management systems. The intermediate steps did not only inform our further steps, but also provided insights that can be useful on their own. As we have shown that education in gamification is a highly discussed topic in academia, an area that has a potential to work with a large number of users and yet one in which little significant results have been published, education was a natural fit for an application context for further study of gamification.

The analysis of empirical studies of gamification had two major results that shaped the rest of our studies: For one, academia has yet to show that gamification actually has the positive effects that many practitioners claim it has. We have seen individual studies that reported partial success, but in most cases the reported outcomes are either inconclusive or not pertinent to the question of the effectiveness of gamification. In a few cases, especially in studies concerning education, negative outcomes were also reported. For another, many existing empirical studies of gamification suffer from flawed experimental design. The reasons are many and varied, but prominent among them are a low number of participants, changes of many variables at once, and the lack of a proper control group. Additionally, studies often relied on self-reported measures instead of measuring actual effects.

Beside issues of experimental design, gamification research is plagued by a problem that is already known from game studies — a disagreement over the nature, the definition, of the subject at hand. We have shown that definitions of gamification vary greatly with a few common denominators: experts both in literature and in our survey agree that points, badges, levels and, to a lesser degree, leaderboards are integral gamification elements. Beyond that, we have identified a dividing line at the inclusion of content-related elements as part of gamification. While some experts focus only on game mechanics (such as points or time pressure) as gamification elements, others also consider the inclusion of game-like content (such as narrative and aesthetics) to be gamification. This division suggested by literature was confirmed through cluster analysis in our expert survey.

The lack of an agreement on a nominal or real definition of gamification led us to the adoption of a different approach of defining, and thereby analyzing, gamification. Following Arjoranta's Wittgensteinian approach to the definition of the term game, we propose a family resemblances definition of gamification along properties that are often found among the family of gamified applications, but not shared among all of them. Analysis of gamification and game studies literature resulted in a list of 60 terms that represent properties of gamified systems on different levels of abstraction. Through an expert survey we narrowed these down to 38 terms that the experts found to be most relevant, while at the same time confirming our initial selection of terms. The terms that were excluded were mostly those that the aforementioned mechanics-oriented group of experts did not consider to be relevant. Each of our chosen terms had at least some supporters among the experts and no single term was added by a large enough number of experts to warrant inclusion. While this is certainly not proof of the list being exhaustive or final, it is strong support for using the list as a basis for future work, both in our research and in that of others.

The variety of terms on both lists (initial and final) and the variance in responses among experts also suggests that these experts see potential in gamification much beyond the currently employed elements of points, badges, levels, and leaderboards. Especially when asked about a specific system of their choice, experts' selections of important terms differed greatly. We even identified a set of terms that the experts considered more relevant to their chosen specific implementation than to gamification in general, suggesting an intention to separate one's own work from run-of-the-mill gamification.

The analysis of gamification in a given system is more nuanced than the deceptively simple question of whether something should be considered gamification or not. We used the aforementioned list of terms to create an instrument for the qualitative analysis of gamification, the Gamification Inventory, and refined it through application to gamified systems in a specific context, in our case education. The Gamification Inventory is intended to guide researchers with basic background knowledge of gamification through the process of analyzing a system, much as a guideline in heuristic usability evaluation would. The Gamification Inventory informs the evaluator about the properties to look for in a system that could signify gamification and prompts him or her to describe how these properties are expressed in the system. The results can then be used for a variety of qualitative analyses, depending on the question at hand.

A practitioner can use the instrument to inform him- or herself about gamification already in use in competing products, while a researcher might be interested in a variety of topics, such as the proliferation of specific types of gamification, or the analysis of a specific system in preparation of an experiment. The refinement process for the instrument suggested some limitations to its use which we will discuss below. It also showed that in many cases, at least in the context of learning management systems (LMSs), the presence of certain properties of gamification can depend strongly on user (in our case: instructor) input. This is especially important to keep in mind when attempting large-scale deployment of gamification, both in an experimental and in a practical setting. Relying on a large number of untrained users to use new features of a system in such a way that it would be considered gamification and to do so with hopes for success is difficult. While literature stresses the importance of tailoring gamification to your users, it may not be the right choice to leave that up to your users.

For this research, we employed the Gamification Inventory to inform our further work in studying the effectiveness of gamification in education, specifically in LMSs. Four evaluators used the inventory to analyze five supposedly gamified LMSs in order to find out whether and how these were gamified. Our study found that all systems were highly similar in their use of gamification, supporting the same trend of employing mostly points, badges, and levels that we have seen elsewhere. Competition, for example in the form of leaderboards, was rare. We have attributed this to the phenomenon seen in some empirical studies that competition can have a negative effect on a subset of users, which is especially problematic in an educational setting. We have also seen the high reliance on instructor input mentioned above — many gamified features in LMSs are optional to use and have to be properly setup by the instructor to have any potential for a positive effect. Our analysis informed our experimental design in a variety of ways, mainly defining the gamification elements to be tested and mandating that all introduced gamification elements should be independent of user input to avoid bias. It also confirmed LMSs as a valid target system for gamification for our experiment and influenced our choice of field access.

We presented a 2x2 factorial design to test a variety of hypotheses on the effectiveness of gamification on the quantity and quality of user (here: student) participation. We chose LMSs, specifically the LMS Moodle with competency grid functionality, in the German federal state of Baden-Württemberg as the application context for our proposed experiment. Baden-Württemberg fulfilled many of the requirements for our setup, as it has a centrally deployed LMS that can be easily modified and has newly developed competency grid functionality that allows the use of gamified features that are somewhat tuned to the specific use case without actually requiring instructor input. After compiling a list of requirements that the software to be used in the experiment needed to fulfill, we designed and implemented it fully. The result is a gamified version of the application DAKORA which provides access to the competency grid functionality on mobile devices, adding points and badges based on user activity and providing expansion options for further gamification elements to be used in future research.

Unfortunately field access has proved to be highly difficult, even with support and permission from the political and administrative parts of the educational system. Additionally, a proper experiment would need to be run long-term, exceeding the scope of this research. Issues with individual schools' and teachers' adoption of new technology further complicate field access. We have provided a summary of the issues and are positive that these can be overcome given enough time. Currently, adoption of the competency grid in schools in Baden-Württemberg is insufficient for proper large scale testing, even though it is politically desired. Since all required applications including the one developed for this research are freely available to interested parties, an adoption of our experimental design to other educational contexts is possible, however.

Much gamification research has been performed with a priority on the design of novel gamified systems, leading to interesting applications but leaving us with a lack of underlying data. This research provides both an argument for structured basic research as well as instruments intended to help further that goal. Firmly rooted in the human computer interaction section of computer science research, the nature of the research question required a broadening of the horizon towards psychology and game studies. Employing established theories from those fields, we were able to contribute to the growing body of knowledge about gamification by providing a better understanding of what relevant properties of gamified systems are and leveraged that to create an instrument that both researchers and practitioners can use in their analysis of gamified systems. It seems obvious, that no gamification research can be successful without such a multidisciplinary approach, but the Gamification Inventory we created in combination with the accompanying evaluation guideline should reduce the need for both researchers and practitioners to delve deeply into other disciplines. HCI experts should be able to adopt the Gamification Inventory similar to their use of heuristics in usability evaluation in order to formatively and summatively assess systems, while experts from other domains should be able to employ the HCI knowledge integrated into the inventory to better understand gamification in their own domain. Finally, with gamification increasing greatly in popularity, many researchers and practitioners who are not well versed in the field of game studies are confronted with it. The Gamification Inventory is designed to help alleviate that gap in knowledge.

6.1 Limitations

We have summarized a variety of positive results above, but these do not come without limitations. The major limitation of our work on the Gamification Inventory is the limited amount of experts that participated in our expert survey. With only 13 experts participating and most of those identifying as researchers, generalization is impossible. While the congruence between literature and expert opinions supports our results, a larger sample size would be needed to be able to draw definitive conclusions. Related to this limitation is the relative shallowness of the data acquired in the expert survey. Given such a low number of participants, qualitative interviews would have added much more detail to our results as we would have been able to discuss the understanding of the individual terms in the survey and the reasoning behind the experts' responses. As it is, our research relies on experts having a similar understanding of the terms in the survey and we have to rely on our own assumptions for the reasoning behind expert responses.

When analyzing the results of the survey in more depth, we identified two clusters of different response patterns and choosing terms for our further work by majority vote meant that the cluster of experts focusing mostly on mechanics had a larger influence on the result than the other. This leads to a bias towards terms that describe game mechanics in the Gamification Inventory following a stricter understanding of gamification. With such a low number of experts, this can very easily be a statistical fluke. Quite possibly, a different sample of experts would skew in the other direction instead. That said, with our methods groups that focus on fewer terms will always have a larger influence on the results than those considering most terms to be of relevance.

The first iteration of the Gamification Inventory was sometimes mistaken for a quantitative instrument due to the inclusion of a closed yes-no-maybe response for each item. This was meant for ease of coding and comparison of evaluator responses, but turned out to be confusing. This has been removed in the second iteration of the inventory, which should remove all appearances of a quantitative instrument. The second iteration has not been tested with external evaluators yet, however, so we can only postulate that the changes were positive. Similarly, we attributed the relatively low inter-rater reliability achieved in the application of the first iteration of the Gamification Inventory to different interpretations of that closed response, especially where yes and maybe are concerned. While this is suggested in the qualitative responses, we have not yet shown that a higher agreement can be reached with the new instrument (and coding of the responses). Coding the qualitative responses from that first evaluation shows a decent overlap in identified gamification elements, however, especially considering that three of the evaluators had only basic training in gamification.

The basic premise of our work, as extracted from literature, is that researchers have yet to confirm the positive effects attributed to gamification. This premise is confirmed by our analysis of empirical studies, but that analysis is not exhaustive. The rate at which publications about gamification are created is so high that it is possible that we have missed new breakthrough results in the field. Our sample of studies uses major meta studies as a base and expands upon that, giving us a reasonably high amount of confidence in the result. Yet, a new thorough examination of studies published since Hamari's and Dicheva et al.'s work as well as our own publication discussing gamification in its trough of disillusionment would help alleviate concerns in this area.

The most impactful limitation of our work on the effectiveness of gamification is clearly that we did not perform the experiment we designed and prepared. Instead of furthering the body of knowledge in that regard — an absolute priority in gamification research following the premise discussed above — we only provided the tools to do so. The requirements of a large-scale, long-term experiment in combination with the difficulties of field access unfortunately leave the execution of the experiment out of the scope of this work. Not only does that deprive us of the experimental results, but also of a validation of our experimental setup and large-scale testing of the software we developed. That said, we have thoroughly prepared the experiment based on literature and the results of our previous work and are confident that this brings gamification research a step closer to answering the important questions at hand.

While most of our preparatory work for the experiment can be useful for other experiments, the final results are tailored towards measuring the effectiveness of specific gamification elements in learning management systems, specifically in Moodle with the competency grid plug-in as it is used in Baden-Württemberg. All software necessary — Moodle, the competency grid plug-in, DAKORA, and our own plug-ins is available as open source software, enabling other researchers to adapt them to their own circumstances. Adaptation is likely necessary, however, when switching circumstances. Our own software is designed to be expandable for the inclusion of other gamification elements and for easy modification of those that are already implemented. This requires some programming knowledge and some elements will be more difficult to add than others. It is therefore quite possible to use our work outside the specific context we have prepared, but the further one sways from the original context, the more adaptation will be needed.

Overall there are a variety of areas of improvement for both major parts of our work and one should keep these limitations in mind for further work. If one does so, these results should be very useful for further studies of gamification — both its analysis and the evaluation of its effectiveness.

6.2 Future Work

This research did not so much provide answers but rather discovered important questions in gamification research and presented instruments that can help answer those questions. One important aspect of future work is therefore definitely to do large-scale, long-term tests of the effectiveness of different gamification elements as motivational affordances for system usage. We have chosen points and badges as the most used gamification elements for testing, but other elements will also have to be tested following that. The Gamification Inventory can support the identification of such elements. As soon as a sufficient number of schools in Baden-Württemberg have adopted the competency grid and DAKORA, our proposed experiment should be performed. The results will give some insight into the effectiveness of points and badges, but additional research will be required to form a full picture. For one, the experiment bears repetition in other LMS contexts and school systems. For another, similar experiments in other contexts will be needed to be able to abstract from the context to gamification in general. The context and the target audience may have a large impact on the effectiveness, even if preliminary research suggests that gamification may be effective in a variety of contexts. Since research suggests that user experience plays an important role in the adoption of LMSs by teachers, it may be beneficial to think about improving the teacher side of the LMS environment in addition to the previous focus on student interaction. Since teachers act as gatekeepers in this case, low adoption among teachers makes improvements on the student side rather futile.

Further research is also needed to improve the Gamification Inventory we presented. An application to a variety of fields by a variety of evaluators will help identify both problematic and redundant items as well as discover new items that have not yet been included in the inventory. It would be especially interesting to use the Gamification Inventory in an attempt at classifying gamified systems and to compare the results of such a classification to that of other approaches. With such a classification and successful experiments for the effectiveness of various gamification elements, one could also look for the influence of player types on the creation of gamified systems and seek to further validate player type models. Any holistic understanding of gamification needs to include not only the systems itself — the area we have mostly dealt with in this research — but also the contexts that they are used in and especially the people that use them. The Gamification Inventory could also be validated and improved by directly confronting experts with the instrument in combination with an interview.

Some areas of our research should be repeated for validation: As we have discussed in the limitations section, it would be good to survey a larger number of experts to enable generalization or at least support our findings. This would also allow for a more detailed cluster analysis of the responses, as one would no longer be limited by the small size of clusters. Repetition of our analysis of empirical studies including the newest work would also be very informative. Our research has helped to provide a foundation for this future work and should take us one step closer to understanding the effects and different expressions of gamification as it is being used and support the development of new forms of gamification.

References

Aarseth, E. J. (1997). *Cybertext: Perspectives on Ergodic Literature*. Baltimore: JHU Press.

Aarseth, E. J. (2012). A Narrative Theory of Games. In *Proceedings of the International Conference on the Foundations of Digital Games* (pp. 129–133). ACM. doi: 10.1145/2282338.2282365

Abt, C. C. (1987). *Serious Games*. Lanham, MD: University Press of America.

Agmon, M., Perry, C. K., Phelan, E., Demiris, G., & Nguyen, H. Q. (2011). A Pilot Study of Wii Fit Exergames to Improve Balance in Older Adults. *Journal of Geriatric Physical Therapy*, *34*(4), 161–167. doi: 10.1519/JPT.0b013e3182191d98

Akpolat, B., & Slany, W. (2014). Enhancing Software Engineering Student Team Engagement in a High-intensity Extreme Programming Course using Gamification. In *2014 IEEE 27th Conference on Software Engineering Education and Training (CSEE T)* (pp. 149–153). doi: 10.1109/CSEET.2014.6816792

Alha, K., Koskinen, E., Paavilainen, J., Hamari, J., & Kinnunen, J. (2014). Free-to-play Games: Professionals' Perspectives. In *Proceedings of Nordic Digra 2014*. Gotland. Retrieved 2017-01-09, from http://www.digra.org/wp-content/uploads/digital-library/nordicdigra2014_submission_8.pdf

Allam, A., Kostova, Z., Nakamoto, K., & Schulz, P. J. (2015). The Effect of Social Support Features and Gamification on a Web-Based Intervention for Rheumatoid Arthritis Patients: Randomized Controlled Trial. *Journal of Medical Internet Research*, *17*(1). doi: 10.2196/jmir.3510

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with Massive Online Courses. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 687–698). New York, NY, USA: ACM. doi: 10.1145/2566486.2568042

Anderson-Hanley, C., Arciero, P. J., Brickman, A. M., Nimon, J. P., Okuma, N., Westen, S. C., . . . Zimmerman, E. A. (2012). Exergaming and Older Adult Cognition: A Cluster Randomized Clinical Trial. *American Journal of Preventive Medicine*, *42*(2), 109–119. doi: 10.1016/j.amepre.2011.10.016

Andriotis, N. (2014). *Gamification Survey Results.* Retrieved 2016-09-08, from http://www.talentlms.com/blog/gamification-survey-results/

Ang, C. S., & Zaphiris, P. (2010). Social Roles of Players in Mmorpg Guilds. *Information, Communication & Society*, *13*(4), 592–614. doi: 10.1080/13691180903266952

Arjoranta, J. (2014). Game Definitions: A Wittgensteinian Approach. *Game Studies*, *14*(1). Retrieved 2014-09-17, from http://gamestudies.org/1401/articles/arjoranta

Ashmore, C., & Nitsche, M. (2007). The Quest in a Generated World. In *Proc. 2007 Digital Games Research Assoc.(DiGRA) Conference: Situated Play* (pp. 503–509).

Attali, Y., & Arieli-Attali, M. (2015). Gamification in Assessment: Do Points Affect Test Performance? *Computers & Education*, *83*, 57–63. doi: 10.1016/j.compedu.2014.12.012

Attwell, G. (2007). Personal Learning Environments - the Future of eLearning? *Elearning Papers*, *2*(1), 1–8. Retrieved 2016-09-07, from http://digtechitalia.pbworks.com/w/file/fetch/88358195/Atwell%202007.pdf

Baek, Y., Jung, J., & Kim, B. (2008). What Makes Teachers use Technology in the Classroom? Exploring the Factors Affecting Facilitation of Technology with a Korean Sample. *Computers & Education*, *50*(1), 224–234. doi: 10.1016/j.compedu.2006.05.002

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Improving Participation and Learning with Gamification. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 10–17). New York, NY, USA: ACM. doi: 10.1145/2583008.2583010

Barik, T., Murphy-Hill, E., & Zimmermann, T. (2016). A Perspective on Blending Programming Environments and Games: Beyond Points, Badges, and Leaderboards. In *2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 134–142). doi: 10.1109/VLHCC.2016.7739676

Bartle, R. (1996). *Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs.* Retrieved 2012-

07-12, from http://www.mud.co.uk/richard/hcds.htm

Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., & Salas, E. (2012). Toward a Taxonomy Linking Game Attributes to Learning An Empirical Study. *Simulation & Gaming*, *43*(6), 729–760. doi: 10.1177/1046878112439444

Blair, S. (1985). Professionalization of Attitude toward Play in Children and Adults. *Research Quarterly for Exercise and Sport*, *56*(1), 82–83. doi: 10.1080/02701367.1985.10608436

Blohm, D. I., & Leimeister, P. D. J. M. (2013). Gamification. *Business & Information Systems Engineering*, 1–4. doi: 10.1007/s12599-013-0273-5

Bogost, I. (2007). *Persuasive Games. The Expressive Power of Videogames.* Cambridge, Mass.: MIT Press.

Braungardt, K. (2014). *Zentrale Lernplattformen an Deutschen Universitäten / Hochschulen > 10.000 Studierende.* Retrieved 2016-09-07, from http://de.slideshare.net/brillux/lms-2014

Breiter, A., Aufenanger, S., Averbeck, I., Welling, S., & Wedjelek, M. (2013). *Medienintegration in Grundschulen: Untersuchung zur Förderung von Medienkompetenz und der Unterrichtlichen Mediennutzung in Grundschulen sowie ihrer Rahmenbedingungen in Nordrhein-Westfalen* (Landesanstalt für Medien Nordrhein-Westfalen, Ed.) (No. 73). Berlin: Vistas.

Breiter, A., & Welling, S. (2008). *Arbeitsräume im Internet für Schulen Bericht Nr 10.* Landesinstitut für Schulentwicklung (LS). Retrieved 2016-09-07, from http://www.ls-bw.de/projekte/sonstige_projekte/arimis/arimis10.pdf

Breuer, J. S., & Bente, G. (2010). Why so Serious? On the Relation of Serious Games and Learning. *Eludamos. Journal for Computer Game Culture*, *4*(1), 7–24. Retrieved 2015-11-10, from http://www.eludamos.org/index.php/eludamos/article/view/vol4no1-2

Broer, J. (2014). Gamification and the Trough of Disillusionment. In A. Butz, M. Koch, & J. Schlichter (Eds.), *Mensch & Computer 2014 - Workshopband* (pp. 389–395). Berlin: De Gruyter Oldenbourg.

Broer, J. (2015). Is This New? Family Resemblances in Gamification in Education. *Bulletin of the IEEE Technical Committee on Learning Technology*, *17*(4), 14–17. Retrieved from http://www.ieeetclt.org/issues/december2015/4_Broer.pdf

Broer, J., & Breiter, A. (2015). Potentials of Gamification in Learning Management Systems:

A Qualitative Evaluation. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & É. Lavoué (Eds.), *Design for Teaching and Learning in a Networked World* (pp. 389–394). Springer International Publishing.

Broer, J., & Poeppelbuss, J. (2013). Gamification – A New Phenomenon in Information Systems Research? In *Proceedings of the 24th Australasian Conference on Information Systems*. Melbourne. Retrieved from http://mo.bf.rmit.edu.au/acis2013/307.pdf

Browne, T., Jenkins, M., & Walker, R. (2006). A Longitudinal Perspective Regarding the use of VLEs by Higher Education Institutions in the United Kingdom. *Interactive Learning Environments*, *14*(2), 177–192. doi: 10.1080/10494820600852795

Brox, E., Luque, L. F., Evertsen, G. J., & Hernandez, J. E. G. (2011). Exergames for Elderly: Social Exergames to Persuade Seniors to Increase Physical Activity. In *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)* (pp. 546–549).

Bundesverband Interaktive Unterhaltungssoftware. (2016). *Der Durchschnittliche Gamer in Deutschland ist 35 Jahre Alt.* Retrieved 2016-09-28, from https://www.biu-online.de/2016/06/07/der-durchschnittliche-gamer-in-deutschland-ist-35-jahre-alt/

Bundeszentrale für Politische Bildung. (2015). *Kulturhoheit | bpb.* Retrieved 2016-09-07, from http://www.bpb.de/nachschlagen/lexika/recht-a-z/22499/kulturhoheit

Burkhard, M., Nutsi, A., & Koch, M. (2013). Einsatz von Spaß und Humor zur Förderung Sozialer Aktivität. In *Mensch & Computer Workshopband* (pp. 369–374).

Caillois, R. (1979). *Man, Play, and Games.* New York: Schocken Books.

Capterra. (2015). *Top LMS Software.* Retrieved 2015-03-10, from http://www.capterra.com/learning-management-system-software/#infographic

Carrol, J. M. (2002). *Human-Computer Interaction.* New York: ACM Press.

Celino, I. (2013). Location-Based Games for Citizen Computation. In P. Michelucci (Ed.), *Handbook of Human Computation* (pp. 297–316). Springer New York.

Chen, J. (2007). Flow in Games (and Everything Else). *Communications of the ACM*, *50*(4), 31. doi: 10.1145/1232743.1232769

Choi, J., Mogami, T., & Medalia, A. (2010). Intrinsic Motivation Inventory: An Adapted

Measure for Schizophrenia Research. *Schizophrenia Bulletin*, *36*(5), 966–976. doi: 10.1093/schbul/sbp030

Chou, Y.-k. (2015). *Octalysis: Complete Gamification Framework.* Retrieved 2015-03-31, from `http://www.yukaichou.com/gamification-examples/octalysis-complete-gamification-framework/`

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. doi: 10.1177/001316446002000104

Conati, C., & Gutica, M. (2016). Interaction with an Edu-Game: A Detailed Analysis of Student Emotions and Judges' Perceptions. *International Journal of Artificial Intelligence in Education.* doi: 10.1007/s40593-015-0081-9

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., … players, F. (2010). Predicting Protein Structures with a Multiplayer Online Game. *Nature*, *466*(7307), 756–760. doi: 10.1038/nature09304

Cornforth, D. J., & Adam, M. T. P. (2015). Player Profiling in Minecraft. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 135–155). Springer International Publishing.

Cotner, S. H., Fall, B. A., Wick, S. M., Walker, J. D., & Baepler, P. M. (2008). Rapid Feedback Assessment Methods: Can We Improve Engagement and Preparation for Exams in Large-enrollment Courses? *Journal of Science Education and Technology*, *17*(5), 437–443. doi: 10.1007/s10956-008-9112-8

Crawford, C. (1984). *The Art of Computer Game Design.* McGraw-Hill/Osborne Media.

Crawford, C. (2003). *The Art of Interactive Design.* San Francisco: No Starch Press.

Crawford, C. (2005). *Chris Crawford on Interactive Storytelling.* New Riders Pub.

Csíkszentmihályi, M. (2008). *Flow: The Psychology of Optimal Experience.* New York: Harper Perennial Modern Classics.

da Rocha Seixas, L., Gomes, A. S., & de Melo Filho, I. J. (2016). Effectiveness of Gamification in the Engagement of Students. *Computers in Human Behavior*, *58*, 48–63. doi: 10.1016/j.chb.2015.11.021

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and Intrinsic Motivation to Use Computers in the Workplace. *Journal of Applied Social Psychology*, 1111–1132.

de-Marcos, L., Domínguez, A., Saenz-de Navarrete, J., & Pagés, C. (2014). An Empirical Study Comparing Gamification and Social Networking on E-learning. *Computers & Education*, *75*, 82–91. doi: 10.1016/j.compedu.2014.01.012

de-Marcos, L., Garcia-Lopez, E., & Garcia-Cabot, A. (2016). On the Effectiveness of Game-like and Social Approaches in Learning: Comparing Educational Gaming, Gamification & Social Networking. *Computers & Education*, *95*, 99–113. doi: 10.1016/j.compedu.2015.12.008

de Smet, C., Bourgonjon, J., de Wever, B., Schellens, T., & Valcke, M. (2012). Researching Instructional Use and the Technology Acceptation of Learning Management Systems by Secondary School Teachers. *Computers & Education*, *58*(2), 688–696. doi: 10.1016/j.compedu.2011.09.013

De Charms, R. (2013). *Personal Causation: The Internal Affective Determinants of Behavior.* New York: Routledge.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research*, *71*(1), 1–27. doi: 10.3102/00346543071001001

Deci, E. L., & Ryan, R. M. (1985a). The General Causality Orientations Scale: Self-Determination in Personality. *Journal of Research in Personality*, *19*, 109–134.

Deci, E. L., & Ryan, R. M. (1985b). *Intrinsic Motivation and Self-Determination in Human Behavior.* Boston, MA: Springer US.

Deci, E. L., Ryan, R. M., & Koestner, R. (2001). The Pervasive Negative Effects of Rewards on Intrinsic Motivation: Response to Cameron (2001). *Review of Educational Research*, *71*(1), 43–51. doi: 10.3102/00346543071001043

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and Education: The Self-determination Perspective. *Educational Psychologist*, *26*(3-4), 325–346. doi: 10.1080/00461520.1991.9653137

Denny, P. (2013). The Effect of Virtual Achievements on Student Engagement. In *Proc. SIGCHI Conference on Human Factors in Computing Systems* (pp. 763–772). New York, NY, USA: ACM. doi: 10.1145/2470654.2470763

Deterding, S. (2011). Situated Motivational Affordances of Game Elements: A Conceptual Model. In *CHI 2011 Gamification Workshop.* Retrieved 2013-05-

02, from http://gamification-research.org/wp-content/uploads/2011/04/09-Deterding.pdf

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining "Gamification". In *MindTrek '11 Proceedings* (pp. 9–15). New York, NY, USA: ACM. doi: 10.1145/2181037.2181040

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in Education: A Systematic Mapping Study. *Journal of Educational Technology & Society*, *18*(3), 75–88. Retrieved from http://www.jstor.org/stable/jeductechsoci.18.3.75

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*, 380–392. doi: 10.1016/j.compedu.2012.12.020

Dougherty, J. (2015). *Gamification: Rethinking The Grading Process (Part 3 of 6).* Retrieved 2016-03-17, from http://www.teq.com/blog/2015/04/gamification-rethinking-grading-process-part-3-5/

Drieschner, E. (2009). *Bildungsstandards Praktisch: Perspektiven Kompetenzorientierten Lehrens und Lernens.* Wiesbaden: VS Verlag für Sozialwissenschaften. Retrieved from http://dx.doi.org/10.1007/978-3-531-91343-8

Ducheneaut, N., & Moore, R. J. (2004). The Social Side of Gaming: A Study of Interaction Patterns in a Massively Multiplayer Online Game. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (pp. 360–369). New York, NY, USA: ACM. doi: 10.1145/1031607.1031667

Duggan, K., & Shoup, K. (2013). *Business Gamification For Dummies.* Hoboken: For Dummies.

Duggan, M. (2015). *Gaming and Gamers.* Pew Research Center. Retrieved from http://www.pewinternet.org/2015/12/15/gaming-and-gamers/

Duncan, S. C. (2011). Minecraft, Beyond Construction and Survival. *Well Played*, *1*(1), 1–22. Retrieved 2016-06-15, from http://dl.acm.org/citation.cfm?id=2207096.2207097

Eck, R. V., & Dempsey, J. (2002). The Effect of Competition and Contextualized Advisement on the Transfer of Mathematics Skills in a Computer-based Instructional Simulation Game. *Educational Technology Research and Development*, *50*(3), 23–41. doi: 10.1007/BF02505023

Edmunds, J., Ntoumanis, N., & Duda, J. L. (2006). A Test of Self-Determination Theory in the Exercise Domain. *Journal of Applied Social Psychology*, *36*(9), 2240–2265. doi: 10.1111/j.0021-9029.2006.00102.x

Eickhoff, C., Harris, C. G., de Vries, A. P., & Srinivasan, P. (2012). Quality Through Flow and Immersion: Gamifying Crowdsourced Relevance Assessments. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 871–880). New York, NY, USA: ACM. doi: 10.1145/2348283.2348400

Falkner, N. J. G., & Falkner, K. E. (2014). "Whither, Badges?" or "Wither, Badges!": A Metastudy of Badges in Computer Science Education to Clarify Effects, Significance and Influence. In *Proceedings of the 14th Koli Calling International Conference on Computing Education Research* (pp. 127–135). New York, NY, USA: ACM. doi: 10.1145/2674683.2674698

Farzan, R., DiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W., & Brownholtz, E. A. (2008). Results from Deploying a Participation Incentive Mechanism within the Enterprise. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems* (pp. 563–572). Florence, Italy: ACM. doi: 10.1145/1357054.1357145

Filsecker, M., & Hickey, D. T. (2014). A Multilevel Analysis of the Effects of External Rewards on Elementary Students' Motivation, Engagement and Learning in an Educational Game. *Computers & Education*, *75*, 136–148. doi: 10.1016/j.compedu.2014.02.008

Finken, J., Marx, F., & Breiter, A. (2015). Computerbasiertes Testen zur Messung von Musikkompetenzen. In U. Lucke, A. Schwill, & R. Zender (Eds.), *DELFI 2016 - Die 14. E-Learning Fachtagung Informatik* (Vol. P-262, pp. 323–326). Bonn: Gesellschaft für Informatik.

Fogg, B. (2007). *BJ Fogg's Behavior Model.* Retrieved 2017-02-28, from `http://www.behaviormodel.org/`

Fogg, B. (2009). A Behavior Model for Persuasive Design. In *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 40:1–40:7). New York, NY, USA: ACM. doi: 10.1145/1541948.1541999

Ford, M. E. (1992). *Motivating Humans: Goals, Emotions, and Personal Agency Beliefs.* SAGE Publications.

Frasca, G. (2003). Ludologists Love Stories, Too: Notes from a Debate that Never Took Place. In *Digital Games Research Conference 2003.*

Frasca, G. (2007). Play the Message: Play, Game and Videogame Rhetoric. *Unpublished PhD dissertation. IT University of Copenhagen, Denmark.* Retrieved 2016-06-15, from http://www.powerfulrobot.com/Frasca_Play_the_Message_PhD.pdf

Fromme, J. (2003). Computer Games as a Part of Children's Culture. *Game Studies*, *3*(1). Retrieved 2016-09-28, from http://www.gamestudies.org/0301/fromme/

Gamification World S.L. (2016). *GWC Conference 2016: The Engagement & Gamification Conference.* Retrieved 2017-01-19, from http://www.gwc-conference.com/

García-Peñalvo, F. J., Conde, M. Á., Alier, M., & Casany, M. J. (2011). Opening Learning Management Systems to Personal Learning Environments. *Journal of Universal Computer Science*, *17*(9), 1222–1240. Retrieved 2016-09-07, from http://www.jucs.org/jucs_17_9/opening_learning_management_systems/jucs_17_09_1222_1240_garcia.pdf

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, *33*(4), 441–467. doi: 10.1177/1046878102238607

Gartner. (2013). *Hype Cycle for Emerging Technologies, 2013.* Retrieved from http://www.gartner.com/document/2571624

Gartner Inc. (2012). *Gartner Says by 2014, 80 Percent of Current Gamified Applications Will Fail to Meet Business Objectives Primarily Due to Poor Design.* Retrieved 2017-01-19, from http://www.gartner.com/newsroom/id/2251015

Gee, J. P. (2005). Learning by Design: Good Video Games as Learning Machines. *E-Learning and Digital Media*, *2*(1), 5–16. doi: 10.2304/elea.2005.2.1.5

Gee, J. P. (2008). *What Video Games Have to Teach Us About Learning and Literacy* (2nd revised, updated ed.). New York: St. Martin'S Press.

Gil, B., Cantador, I., & Marczewski, A. (2015). Validating Gamification Mechanics and Player Types in an E-learning Environment. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & É. Lavoué (Eds.), *Design for Teaching and Learning in a Networked World* (Vol. 9307, pp. 568–572). Cham: Springer International Publishing.

Gillison, F., Osborn, M., Standage, M., & Skevington, S. (2009). Exploring the Experience of Introjected Regulation for Exercise across Gender in Adolescence. *Psychology of Sport and Exercise*, *10*(3), 309–319. doi: 10.1016/j.psychsport.2008.10.004

Goehle, G., & Wagaman, J. (2015). The Impact of Gamification in Web Based Homework.

PRIMUS. doi: 10.1080/10511970.2015.1122690

Griffith, L. E., Cook, D. J., Guyatt, G. H., & Charles, C. A. (1999). Comparison of Open and Closed Questionnaire Formats in Obtaining Demographic Information From Canadian General Internists. *Journal of Clinical Epidemiology*, *52*(10), 997–1005. doi: 10.1016/S0895-4356(99)00106-7

Grolnick, W. S., Deci, E. L., & Ryan, R. M. (1997). Internalization within the Family: The Self-determination Theory Perspective. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 135–161). Hoboken, NJ, US: John Wiley & Sons Inc.

Grubbauer, M. (2011). Hauptrichtungen der Spieltheorie. In *Spielen als Pädagogische Maßnahme* (pp. 21–29). VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-92855-5_2

Guin, T. D.-L., Baker, R., Mechling, J., & Ruylea, E. (2012). Myths and Realities of Respondent Engagement in Online Surveys. *International Journal of Market Research*, *54*(5), 1–21. Retrieved from `http://www.marketstrategies.com/user_area/content_media/Online_Survey_Engagement.pdf`

Gwet, K. L. (2014). *Handbook of Inter-Rater Reliability, 4th Edition: The Definitive Guide to Measuring The Extent of Agreement Among Raters.* Gaithersburg, MD: Advanced Analytics, LLC.

Hakulinen, L., Auvinen, T., & Korhonen, A. (2013). Empirical Study on the Effect of Achievement Badges in TRAKLA2 Online Learning Environment. In *Learning and Teaching in Computing and Engineering (LaTiCE), 2013* (pp. 47–54). doi: 10.1109/LaTiCE.2013.34

Hamari, J. (2013). Transforming Homo Economicus into Homo Ludens: A Field Experiment on Gamification in a Utilitarian Peer-to-peer Trading Service. *Electronic Commerce Research and Applications*, *12*(4), 236–245. doi: 10.1016/j.elerap.2013.01.004

Hamari, J., & Koivisto, J. (2013). Social Motivations To Use Gamification: An Empirical Study Of Gamifying Exercise. *ECIS 2013 Completed Research.* Retrieved from `http://aisel.aisnet.org/ecis2013_cr/105`

Hamari, J., & Koivisto, J. (2015). Why Do People Use Gamification Services? *International Journal of Information Management*, *35*(4), 419–431. doi: 10.1016/j.ijinfomgt.2015.04.006

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of

Empirical Studies on Gamification. In *2014 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). doi: 10.1109/HICSS.2014.377

Hanus, M. D., & Fox, J. (2015). Assessing the Effects of Gamification in the Classroom: A Longitudinal Study on Intrinsic Motivation, Social Comparison, Satisfaction, Effort, and Academic Performance. *Computers & Education*, *80*, 152–161. doi: 10.1016/j.compedu.2014.08.019

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. doi: 10.2307/2346830

Hartmann, T. (2009). Let's compete! In T. Quandt, J. Wimmer, & J. Wolling (Eds.), *Die Computerspieler* (pp. 211–224). VS Verlag für Sozialwissenschaften. doi: 10.1007/978-3-531-91848-8_12

Hew, K. F., Huang, B., Chu, K. W. S., & Chiu, D. K. W. (2016). Engaging Asian Students through Game Mechanics: Findings from Two Experiment Studies. *Computers & Education*, *92–93*, 221–236. doi: 10.1016/j.compedu.2015.10.010

Höflich, J. R. (2003). Vermittlungskulturen im Wandel: Brief - E-Mail - SMS. In *Vermittlungskulturen im Wandel: Brief - E-Mail - SMS* (pp. 39–61). Frankfurt am Main: Peter Lang.

Hollingsed, T., & Novick, D. G. (2007). Usability Inspection Methods After 15 Years of Research and Practice. In *Proceedings of the 25th Annual ACM International Conference on Design of Communication* (pp. 249–255). New York, NY, USA: ACM. doi: 10.1145/1297144.1297200

Hotz, R. L. (2012). When Gaming Is Good for You. *Wall Street Journal*. Retrieved 2016-09-28, from `http://www.wsj.com/articles/SB10001424052970203458604577263273943183932`

Huang, B., & Hew, K. F. T. (2015). Do Points, Badges and Leaderboard Increase Learning and Activity: A Quasi-Experiment on the Effects of Gamification? In *Proceedings of the 23rd International Conference on Computers in Education.* Asia-Pacific Society for Computers in Education. Retrieved 2017-03-07, from `http://hub.hku.hk/handle/10722/223925`

Huizinga, J. (2004). *Homo ludens: vom Ursprung der Kultur im Spiel.* Hamburg: Rowohlt Taschenbuch Verlag.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A Formal Approach to Game Design and Game Research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (Vol. 4, pp. 1–5). Retrieved 2015-03-10, from `http://www.aaai.org/Papers/Workshops/2004/WS-04-04/WS04-04-001.pdf`

Huotari, K., & Hamari, J. (2012). Defining Gamification: A Service Marketing Perspective. In *Proceeding of the 16th International Academic MindTrek Conference* (pp. 17–22). New York, NY, USA: ACM. doi: 10.1145/2393132.2393137

Huotari, K., & Hamari, J. (2016). A Definition for Gamification: Anchoring Gamification in the Service Marketing Literature. *Electronic Markets*. doi: 10.1007/s12525-015-0212-z

Ihantola, P., Ahoniemi, T., Karavirta, V., & Seppälä, O. (2010). Review of Recent Systems for Automatic Assessment of Programming Assignments. In *Proceedings of the 10th Koli Calling International Conference on Computing Education Research* (pp. 86–93). New York, NY, USA: ACM. doi: 10.1145/1930464.1930480

International Standardization Organization. (2010). *Ergonomics of Human System Interaction - Part 210: Human-centred Design for Interactive Systems.*

Iosup, A., & Epema, D. (2014). An Experience Report on Using Gamification in Technical Higher Education. In *Proceedings of the 45th ACM Technical Symposium on Computer Science Education* (pp. 27–32). New York, NY, USA: ACM. doi: 10.1145/2538862.2538899

Jain, A. K. (2010). Data Clustering: 50 Years Beyond K-means. *Pattern Recognition Letters, 31*(8), 651–666. doi: 10.1016/j.patrec.2009.09.011

Juul, J. (2003). The Game, the Player, the World: Looking for a Heart of Gamesness. In *Level Up: Digital Games Research Conference Proceedings* (pp. 30–45). Utrecht: Utrecht University. Retrieved from `http://www.jesperjuul.net/text/gameplayerworld/`

Juul, J. (2005). *Half-real: Video Games between Real Rules and Fictional Worlds.* Cambridge, Mass.: MIT Press.

Kapp, K. M. (2012). *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education.* San Francisco: Pfeiffer.

Kappen, D. L., & Nacke, L. E. (2013). The Kaleidoscope of Effective Gamification: Deconstructing Gamification in Business Applications. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 119–122). New York, NY, USA: ACM. doi: 10.1145/2583008.2583029

Kappen, D. L., Nacke, L. E., Gerling, K. M., & Tsotsos, L. E. (2016). Design Strategies for Gamified Physical Activity Applications for Older Adults. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 1309–1318). doi: 10.1109/HICSS.2016.166

Kent, S. L. (2001). *The Ultimate History of Video Games: From Pong to Pokémon and Beyond: The Story Behind the Craze that Touched Our Lives and Changed the World.* Rocklin, CA: Prima Pub.

Kettle, K. L., & Häubl, G. (2010). Motivation by Anticipation Expecting Rapid Feedback Enhances Performance. *Psychological Science*, *21*(4), 545–547. doi: 10.1177/0956797610363541

Keusch, F., & Zhang, C. (2015). A Review of Issues in Gamified Surveys. *Social Science Computer Review*, 1–20. doi: 10.1177/0894439315608451

Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, *119*(2), 254–284. doi: 10.1037/0033-2909.119.2.254

Koestner, R., & Hope, N. (2014). A Self-Determination Theory Approach to Goals. In M. Gagné (Ed.), *The Oxford Handbook of Work-Engagement, Motivation, and Self-Determination Theory* (pp. 400–413). Oxford: Oxford University Press.

Konzack, L. (2007). Rhetorics of Computer and Video Game Research. In J. Williams & J. Smith (Eds.), *The Players' Realm: Studies on the Culture of Video Games and Gaming* (pp. 110–130).

Korn, O. (2012). Industrial Playgrounds: How Gamification Helps to Enrich Work for Elderly or Impaired Persons in Production. In *Proceedings of the 4th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (pp. 313–316). New York, NY, USA: ACM. doi: 10.1145/2305484.2305539

Koster, R. (2005). *A Theory of Fun for Game Design.* Scottsdale: Paraglyph Press.

Kringiel, D. (2009). *Computerspielanalyse Konkret: Methoden und Instrumente - Erprobt an Max Payne 2.* München: kopaed.

Kultusminister Konferenz. (n.d.). *Bildungsstandards der Kultusministerkonferenz.* Retrieved 2017-02-27, from `https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html`

Kultusminister Konferenz. (2016). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring.* Retrieved 2017-02-27, from https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf

Kultusministerkonferenz. (2012). *Bildungsstandards im Fach Deutsch für die Allgemeine Hochschulreife.* Retrieved 2017-02-27, from https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2012/2012_10_18-Bildungsstandards-Deutsch-Abi.pdf

Labovitz, S. (1967). Some Observations on Measurement and Statistics. *Social Forces*, *46*(2), 151–160. doi: 10.2307/2574595

Landers, R. N. (2014). Developing a Theory of Gamified Learning Linking Serious Games and Gamification of Learning. *Simulation & Gaming*, *45*(6), 752–768. doi: 10.1177/1046878114563660

Landesinstitut für Schule Bremen. (n.d.). *Landesweite Lernplattform: Bremen Arbeitet mit "itslearning".* Retrieved 2016-09-07, from http://www.lis.bremen.de/medien/itslearning-32095

Landesinstitut für Schulentwicklung Baden-Württemberg. (2016). *Kompetenzraster Deutsch Klasse 5-10 Bildungsplan 2016.* Retrieved from http://www.schule-bw.de/unterricht/individualisiertes_lernen/kompetenzraster2016/deutsch/d_kora_lfs_1_14.pdf

Laskaris, J. (2015). *LMS and Gamification.* Retrieved 2017-01-19, from https://www.talentlms.com/blog/lms-and-gamification/

Laudan, M. (2016). *Elektronischer Sport (e-Sport): Inhaltsanalyse von Computerspielen zur Identifikation von Anforderungen an erfolgreiche e-Sport Spiele.* unpublished bachelor thesis.

Law, E. L.-C., Roto, V., Hassenzahl, M., Vermeeren, A. P., & Kort, J. (2009). Understanding, Scoping and Defining User Experience: A Survey Approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 719–728). ACM. Retrieved 2017-04-03, from http://dl.acm.org/citation.cfm?id=1518813

Lazar, J., Feng, J. H., & Hochheiser, H. (2010). *Research Methods in Human-Computer Interaction* (2nd ed.). Wiley.

Lee, J. J., & Hammer, J. (2011). Gamification in Education: What, How, Why Bother? *Academic Exchange Quarterly*, *15*(2), 146–150. Retrieved 2012-12-06, from http://www.gamifyingeducation.org/files/Lee-Hammer-AEQ-2011.pdf

Lee, T. Y., Dugan, C., Geyer, W., Ratchford, T., Rasmussen, J. C., Shami, N. S., & Lupushor, S. (2013). Experiments on Motivational Feedback for Crowdsourced Workers. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* Retrieved 2015-10-29, from http://www.jamierasmussen.com/publications/ibmerswhotweet.pdf

Lehmann, J., Lalmas, M., Yom-Tov, E., & Dupret, G. (2012). Models of User Engagement. In J. Masthoff, B. Mobasher, M. C. Desmarais, & R. Nkambou (Eds.), *User Modeling, Adaptation, and Personalization* (pp. 164–175). Springer Berlin Heidelberg.

Leong, B., & Luo, Y. (2011). Application of Game Mechanics to Improve Student Engagement. In *International Conference on Teaching and Learning in Higher Education.* Retrieved 2015-03-04, from https://www0.comp.nus.edu/~bleong/teaching/tlhe2011-game.pdf

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the" Overjustification" Hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129. Retrieved 2015-02-17, from http://psycnet.apa.org/journals/psp/28/1/129/

Lepper, M. R., Keavney, M., & Drake, M. (1996). Intrinsic Motivation and Extrinsic Rewards: A Commentary on Cameron and Pierce's Meta-Analysis. *Review of Educational Research*, *66*(1), 5–32. doi: 10.3102/00346543066001005

Li, F. (1999). The Exercise Motivation Scale: Its Multifaceted Structure and Construct Validity. *Journal of Applied Sport Psychology*, *11*(1), 97–115. doi: 10.1080/10413209908402953

Li, W., Grossman, T., & Fitzmaurice, G. (2012). GamiCAD: A Gamified Tutorial System for First Time Autocad Users. In *Proceedings of the 25th annual ACM symposium on User interface software and technology* (pp. 103–112). New York, NY, USA: ACM. doi: 10.1145/2380116.2380131

Lindenberg, S. (2001). Intrinsic Motivation in a New Light. *Kyklos*, *54*(2-3), 317–342. doi: 10.1111/1467-6435.00156

Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a Fad? Gamification in

Health and Fitness Apps. *JMIR Serious Games*, *2*(2), e9. doi: 10.2196/games.3413

Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring Learning Outcomes in Higher Education Motivation Matters. *Educational Researcher*, *41*(9), 352–362. doi: 10.3102/0013189X12459679

Livingston, I. J., Gutwin, C., Mandryk, R. L., & Birk, M. (2014). How Players Value Their Characters in World of Warcraft. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1333–1343). New York, NY, USA: ACM. doi: 10.1145/2531602.2531661

Lobel, A., Granic, I., & Engels, R. C. (2013). Stressful Gaming, Interoceptive Awareness, and Emotion Regulation Tendencies: A Novel Approach. *Cyberpsychology, Behavior, and Social Networking*, *17*(4), 222–227. doi: 10.1089/cyber.2013.0296

Locke, E. A. (1996). Motivation Through Conscious Goal Setting. *Applied and Preventive Psychology*, *5*(2), 117–124. doi: 10.1016/S0962-1849(96)80005-9

Locke, E. A. (2003). Good Definitions: The Epistemological Foundation of Scientific Progress. In J. Greenberg (Ed.), *Organizational Behavior: The State of the Science* (pp. 395–425). Mahwah, NJ: Lawrence Erlbaum.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015). Serious Games Analytics: Theoretical Framework. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 3–29). Springer International Publishing.

Lonn, S., & Teasley, S. D. (2009). Saving Time or Innovating Practice: Investigating Perceptions and Uses of Learning Management Systems. *Computers & Education*, *53*(3), 686–694. doi: 10.1016/j.compedu.2009.04.008

Lübke, S.-I. (2013). *Schule ohne Noten: Lernberichte in der Praxis der Laborschule.* Springer-Verlag.

Mabed, M., & Köhler, T. (2012). The Impact of Learning Management System Usage on Cognitive and Affective Performance. In *Communities in New Media: Virtual Enterprises, Research Communities & Social Media Networks: 15. Workshop GeNeMe '12 Gemeinschaften in Neuen Medien* (pp. 273–284). Dresden: TUDpress. Retrieved 2017-02-28, from `http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-101320`

Machado, M., & Tao, E. (2007). Blackboard vs. Moodle: Comparing User Experience of

Learning Management Systems. In *2007 37th Annual Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports* (pp. S4J–7–S4J–12). doi: 10.1109/FIE.2007.4417910

Malone, T. W. (1982). Heuristics for Designing Enjoyable User Interfaces: Lessons from Computer Games. In *Proceedings of the 1982 Conference on Human Factors in Computing Systems* (pp. 63–68). New York, NY, USA: ACM. doi: 10.1145/800049.801756

Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. doi: 10.1214/aoms/1177730491

Marczewski, A. (2012). *Gamification: A Simple Introduction.* Andrzej Marczewski.

Marczewski, A. (2015). *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design.* CreateSpace Independent Publishing Platform.

Markland, D., & Hardy, L. (1997). On the Factorial and Construct Validity of the Intrinsic Motivation Inventory: Conceptual and Operational Concerns. *Research Quarterly for Exercise and Sport*, *68*(1), 20–32. doi: 10.1080/02701367.1997.10608863

Mayer, R. E. (2014). *Computer Games for Learning: An Evidence-Based Approach.* Cambridge, Massachusetts: The Mit Press.

McAuley, A., Stewart, B., Siemens, G., & Cornier, D. (2010). *The MOOC model for digital practice.* University of Prince Edward Island. Retrieved 2017-05-30, from http://www.elearnspace.org/Articles/MOOC_Final.pdf

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric Properties of the Intrinsic Motivation Inventory in a Competitive Sport Setting: A Confirmatory Factor Analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48–58. doi: 10.1080/02701367.1989.10607413

McGill, T. J., & Klobas, J. E. (2009). A Task–technology Fit View of Learning Management System Impact. *Computers & Education*, *52*(2), 496–508. doi: 10.1016/j.compedu.2008.10.002

McGonigal, J. (2011). *Reality is Broken.* New York: Penguin Press.

McHugh, M. L. (2012). Interrater Reliability: The Kappa Statistic. *Biochemia Medica*, *22*(3), 276–282. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

McNamara, T. (2004). *World of Warcraft Review.* Retrieved 2016-07-26, from http://www.ign.com/articles/2004/12/10/world-of-warcraft-review

Medved, J. (2015). *6 Gamified LMSs Compared - Capterra Blog.* Retrieved 2016-09-08, from http://blog.capterra.com/6-gamified-lmss-compared/

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Do Points, Levels and Leaderboards Harm Intrinsic Motivation?: An Empirical Analysis of Common Gamification Elements. In *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 66–73). New York, NY, USA: ACM. doi: 10.1145/2583008.2583017

Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2015). Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior*. doi: 10.1016/j.chb.2015.08.048

Meske, C., Brockmann, T., Wilms, K., & Stieglitz, S. (2015). Gamify Employee Collaboration-A Critical Review of Gamification Elements in Social Software. In *Australasian Conference on Information Systems 2015.* Adelaide. Retrieved 2016-01-04, from https://acis2015.unisa.edu.au/wp-content/uploads/2015/11/ACIS_2015_paper_106.pdf

Michael, D. R., & Chen, S. L. (2006). *Serious Games: Games That Educate, Train, and Inform.* Boston, Mass.: Thomson Course Technology.

Möll, D. G. (2014). Von Fischen und Haien. Zur Mediatisierung des Glücksspiels am Beispiel Online-Poker. In T. Grenz & G. Möll (Eds.), *Unter Mediatisierungsdruck* (pp. 145–168). Springer Fachmedien Wiesbaden.

Montgomery, D. C. (2009). *Design and Analysis of Experiments* (7th ed.). Hoboken: Wiley.

Morschheuser, B., Henzi, C., & Alt, R. (2015). Increasing Intranet Usage through Gamification – Insights from an Experiment in the Banking Industry. In *2015 48th Hawaii International Conference on System Sciences (HICSS)* (pp. 635–642). doi: 10.1109/HICSS.2015.83

Mozilla. (2016). *About | Open Badges.* Retrieved 2016-03-17, from http://openbadges.org/about/

Myers, D. (2009). In Search of a Minimalist Game. In *Proceedings of the 2009 DIGRA International Conference: Breaking New Ground: Innovation in Games, Play, Practice and Theory.* Retrieved 2016-06-15, from http://www.digra.org/digital-library/publications/in-search-of-a-minimalist-game/

Nacke, L. (2009). From Playability to a Hierarchical Game Usability Model. In *Proceedings of the 2009 Conference on Future Play on @ GDC Canada* (pp. 11–12). New York, NY, USA: ACM. doi: 10.1145/1639601.1639609

Nicholson, S. (2012). A User-Centered Theoretical Framework for Meaningful Gamification. In *Games Learning Society 2012: Madison.* Retrieved 2013-05-02, from `http://scottnicholson.com/pubs/meaningfulframework.pdf`

Nielsen. (2015). *Distribution of Gamers in the United States as of April 2015, by Age Group.* Retrieved 2016-09-28, from `https://www.statista.com/statistics/494857/distribution-of-gamers-by-age-usa/`

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2000). Why Are You Learning a Second Language? Motivational Orientations and Self-Determination Theory. *Language Learning*, *50*(1), 57–85. doi: 10.1111/0023-8333.00111

O'Brien, H. L., & Toms, E. G. (2008). What is User Engagement? A Conceptual Framework for Defining User Engagement with Technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. doi: 10.1002/asi.20801

O'Donovan, S., Gain, J., & Marais, P. (2013). A Case Study in the Gamification of a University-level Games Development Course. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference* (pp. 242–251). New York, NY, USA: ACM. doi: 10.1145/2513456.2513469

Oh, Y., & Yang, S. (2010). Defining exergames and exergaming. In *Proceedings of Meaningful Play* (pp. 1–17). Retrieved from `http://meaningfulplay.msu.edu/proceedings2010/mp2010_paper_63.pdf`

Phillips, C., Johnson, D., Wyeth, P., Hides, L., & Klarkowski, M. (2015). Redefining Videogame Reward Types. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction* (pp. 83–91). New York, NY, USA: ACM. doi: 10.1145/2838739.2838782

Piaget, J. (1975). *Nachahmung, Spiel und Traum* (No. 5). Stuttgart: Klett.

Pink, D. H. (2011). *Drive. The surprising truth about what motivates us.* Edinburgh: Canongate.

Prensky, M. (2001). *Digital Game-based Learning.* McGraw-Hill.

Prensky, M. (2003). Digital Game-based Learning. *Computers in Entertainment*, *1*(1), 21–21.

doi: 10.1145/950566.950596

Reeve, J. (2002). Self-determination theory applied to educational settings. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 183–203). Rochester, NY, US: University of Rochester Press.

Reeves, B., & Read, J. L. (2009). *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete.* Boston: Harvard Business Press.

Riddle, K. (2010). Remembering Past Media Use: Toward the Development of a Lifetime Television Exposure Scale. *Communication Methods and Measures*, *4*(3), 241–255. doi: 10.1080/19312458.2010.505500

Robertson, M. (2010). *Can't play, won't play.* Retrieved 2014-04-17, from `http://hideandseek.net/2010/10/06/cant-play-wont-play/`

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, *25*(1), 54–67. doi: 10.1006/ceps.1999.1020

Ryan, R. M., Koestner, R., & Deci, E. L. (1991). Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and Emotion*, *15*(3), 185–205. doi: 10.1007/BF00995170

Saatchi & Saatchi S. (2011). *Engagement Unleashed: Gamification for Business, Brands, and Loyalty.* Retrieved 2014-05-30, from `http://de.slideshare.net/Saatchi_S/gamification-study`

Salen, K., & Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals.* Cambridge, Mass.: MIT Press.

Sarabi, M. E. (2014). *Von Game Elements in die Gamification.* unpublished bachelor thesis.

Schering, S. (2014). Ein Design Space für internale Gamification-Anwendungen. In A. Butz, M. Koch, & J. Schlichter (Eds.), *Mensch & Computer 2014 - Workshopband* (pp. 361–368). Berlin: De Gruyter Oldenbourg.

Schmidt, R. (2014). *Press Play! Der erste Gamification-Workshop war ein großer Erfolg | Gamification-Workshop.* Retrieved 2017-06-01, from `http://gamification-workshop.org/press-play-der-erste-gamification-workshop-war-ein-grosser-erfolg/`

Schönberger, K. (2003). "... dass jemand mal vorbeischreibt." E-Mail im Alltag - zur Kulturanalyse eines neuen Mediendispositivs. In *Vermittlungskulturen im Wandel: Brief - E-Mail - SMS* (pp. 111–146). Frankfurt am Main: Peter Lang.

Schubert, P., Hager, J., & Paulsen, L. (2014). Auswirkungen von Gamification in Enterprise Collaboration Systems. In A. Butz, M. Koch, & J. Schlichter (Eds.), *Mensch & Computer 2014 - Tagungsband* (pp. 003–014). Berlin: De Gruyter Oldenbourg.

Schulmeister, R. (2005). Zur Didaktik des Einsatzes von Lernplattformen. In M. Franzen & Fachhochschule Solothurn Nordwestschweiz (Eds.), *Lernplattformen: Optimierung der Ausbildung oder didaktischer Rückschritt? Web-based Training 2005* (1. ed., pp. 11–19). Dübendorf: Empa-Akademie.

Scott, A. (2014). How Playcentric Research Methods are Contributing to New Understanding and Opportunities for Design. In P. Rodgers & J. Yee (Eds.), *The Routledge Companion to Design Research* (pp. 400–414).

Seifert, R., & Jöckel, S. (2009). Die Welt der Kriegskunst. In T. Quandt, J. Wimmer, & J. Wolling (Eds.), *Die Computerspieler* (pp. 297–311). VS Verlag für Sozialwissenschaften.

Seligman, M. (2011). *Flourish: A Visionary New Understanding of Happiness and Well-being.* New York: ATRIA.

Siang, A. C., & Rao, R. K. (2003). Theories of Learning: A Computer Game Perspective. In *Fifth International Symposium on Multimedia Software Engineering, 2003. Proceedings* (pp. 239–245). doi: 10.1109/MMSE.2003.1254447

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. doi: 10.1016/j.chb.2012.06.007

Skinner, B. F. (1938). *The Behavior of Organisms: An Experimental Analysis.* Oxford, England: Appleton-Century.

Skinner, B. F. (2011). *About Behaviorism.* Knopf Doubleday Publishing Group.

Snow, E. L., Allen, L. K., & McNamara, D. S. (2015). The Dynamical Analysis of Log Data Within Educational Games. In C. S. Loh, Y. Sheng, & D. Ifenthaler (Eds.), *Serious Games Analytics* (pp. 81–100). Springer International Publishing.

Stantchev, V., Colomo-Palacios, R., Soto-Acosta, P., & Misra, S. (2014). Learning Management

Systems and Cloud File Hosting Services: A Study on Students' Acceptance. *Computers in Human Behavior*, *31*, 612–619. doi: 10.1016/j.chb.2013.07.002

Statista. (2016a). *Market Share of Leading MMO Games 2014.* Retrieved 2016-03-17, from http://www.statista.com/statistics/343083/market-share-leading-mmo-games/

Statista. (2016b). *Schüler an Allgemeinbildenden Schulen in Deutschland nach Bundesländern 2016/2017.* Retrieved 2017-04-20, from https://de.statista.com/statistik/daten/studie/1321/umfrage/anzahl-der-schueler-an-allgemeinbildenden-schulen/

Statista. (2016c). *Schüler an Beruflichen Schulen in Deutschland nach Bundesländern 2016/2017.* Retrieved 2017-04-20, from https://de.statista.com/statistik/daten/studie/1330/umfrage/schueler-an-beruflichen-schulen-in-deutschland-nach-bundeslaendern/

Statistisches Landesamt Baden-Württemberg. (2016). *Allgemeinbildende Schulen nach Schularten.* Retrieved 2017-02-27, from http://www.statistik.baden-wuerttemberg.de/BildungKultur/SchulenAllgem/abschulen.jsp

Stenros, J., Holopainen, J., Waern, A., Montola, M., & Ollila, E. (2011). Narrative Friction in Alternate Reality Games: Design Insights from Conspiracy for Good. In *Proceedings of DiGRA 2011 Conference: Think Design Play, Utrecht, The Netherlands.* DiGRA.

Stevens, S. (1946). On the Theory of Scales of Measurement. *Science*, *103*(2684), 677–680. Retrieved 2017-06-14, from http://psychology.okstate.edu/faculty/jgrice/psyc3214/Stevens_FourScales_1946.pdf

Straub, E. T. (2009). Understanding Technology Adoption: Theory and Future Directions for Informal Learning. *Review of Educational Research*, *79*(2), 625–649. doi: 10.3102/0034654308325896

Suits, B. (2005). *The Grasshopper.* Peterborough: Broadview Press.

Sweetser, P., & Wyeth, P. (2005). GameFlow: A Model for Evaluating Player Enjoyment in Games. *Comput. Entertain.*, *3*(3). doi: 10.1145/1077246.1077253

Tang, S.-H., & Hall, V. C. (1995). The Overjustification Effect: A Meta-analysis. *Applied Cognitive Psychology*, *9*(5), 365–404. doi: 10.1002/acp.2350090502

Tavinor, G. (2008). Definition of Videogames. *Contemporary Aesthetics*, *6*. Retrieved from http://hdl.handle.net/2027/spo.7523862.0006.016

ter Vrugte, J., de Jong, T., Vandercruysse, S., Wouters, P., van Oostendorp, H., & Elen, J. (2015). How Competition and Heterogeneous Collaboration Interact in Prevocational Game-based Mathematics Education. *Computers & Education*, *89*, 42–52. doi: 10.1016/j.compedu.2015.08.010

Thom, J., Millen, D., & DiMicco, J. (2012). Removing Gamification from an Enterprise SNS. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 1067–1070). New York, NY, USA: ACM. doi: 10.1145/2145204.2145362

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 229–243). New York, NY, USA: ACM. doi: 10.1145/2967934.2968082

Toups, Z. O., Crenshaw, N. K., Wehbe, R. R., Tondello, G. F., & Nacke, L. E. (2016). "The Collecting Itself Feels Good": Towards Collection Interfaces for Digital Game Objects. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (pp. 276–290). New York, NY, USA: ACM. doi: 10.1145/2967934.2968088

Tsigilis, N., & Theodosiou, A. (2003). Temporal Stability of the Intrinsic Motivation Inventory. *Perceptual and Motor Skills*, *97*(1), 271–280. doi: 10.2466/pms.2003.97.1.271

Vallerand, R. J. (2000). Deci and Ryan's Self-Determination Theory: A View from the Hierarchical Model of Intrinsic and Extrinsic Motivation. *Psychological Inquiry*, *11*(4), 312–318. Retrieved 2016-03-07, from http://www.jstor.org/stable/1449629

von Ahn, L. (2006). Games with a Purpose. *Computer*, *39*(6), 92–94. doi: 10.1109/MC.2006.196

Vorderer, P., Hartmann, T., & Klimmt, C. (2003). Explaining the enjoyment of playing video games: The role of competition. In *Proceedings of the second international conference on Entertainment computing* (pp. 1–9). Carnegie Mellon University. Retrieved 2016-03-02, from http://dl.acm.org/citation.cfm?id=958735

Waern, A. (2012). Framing Games. In *Proceedings of Nordic DiGRA* (Vol. 2012). Retrieved 2016-06-15, from http://www.digra.org/dl/db/12168.20295.pdf

Weaver, D., Spratt, C., & Nair, C. S. (2008). Academic and Student Use of a Learning Management System: Implications for Quality. *Australasian Journal of Educational technology, 24*(1), 30–41. Retrieved 2016-08-31, from http://leading2learn.ca/viu/OLTD504/papers/might%20use/Web%20CT%20study%20at%20Monash.pdf

Weber, K., Martin, M. M., & Cayanus, J. L. (2005). Student Interest: A Two-study Re-examination of the Concept. *Communication Quarterly*, *53*(1), 71–86. doi: 10.1080/01463370500055996

Weibel, A., Wiemann, M., & Osterloh, M. (2014). A Behavioral Economics Perspective on the Overjustification Effect: Crowding-In and Crowding-Out of Intrinsic Motivation. In M. Gagné (Ed.), *The Oxford Handbook of Work-Engagement, Motivation, and Self-Determination Theory* (pp. 72–84). Oxford: Oxford University Press.

Welling, S., Breiter, A., & Brüggemann, M. (2008). *Arbeitsräume im Internet für Schulen Bericht Nr. 9.* Landesinstitut für Schulentwicklung (LS). Retrieved 2016-09-07, from http://www.ls-bw.de/projekte/sonstige_projekte/arimis/arimis10.pdf

Werbach, K., & Hunter, D. (2015). *The Gamification Toolkit: Dynamics, Mechanics, and Components for the Win.* Wharton Digital Press.

Whitmer, J., Nuñez, N., Harfield, T., & Forteza, D. (2016). *Patterns in Blackboard Learn tool use: Five Course Design Archetypes.* Retrieved from http://bbbb.blackboard.com/coursedesignresearch

Williams, G. C., McGregor, H. A., Zeldman, A., Freedman, Z. R., & Deci, E. L. (2004). Testing a Self-Determination Theory Process Model for Promoting Glycemic Control Through Diabetes Self-Management. *Health Psychology*, *23*(1), 58–66. doi: 10.1037/0278-6133.23.1.58

Wise, S. L., & DeMars, C. E. (2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment*, *10*(1), 1–17. doi: 10.1207/s15326977ea1001_1

Wittgenstein, L. (2007). *Philosophical Investigations: The German Text, with a Revised English Translation* (3. ed.; G. E. M. Anscombe, Trans.). Oxford: Blackwell.

Wonnacott, R. J., & Wonnacott, T. H. (1985). *Introductory Statistics* (4. ed. ed.). New York u.a.: Wiley.

Wu, P.-H., Hwang, G.-J., Milrad, M., Ke, H.-R., & Huang, Y.-M. (2012). An Innovative Concept Map Approach for Improving Students' Learning Performance with an Instant Feedback Mechanism. *British Journal of Educational Technology*, *43*(2), 217–232. doi: 10.1111/j.1467

-8535.2010.01167.x

Yee, N. (2001). *The Norrathian Scrolls: A Study of EverQuest.* Retrieved 2016-03-17, from `http://www.nickyee.com/eqt/skinner.html`

Yee, N. (2006). The Labor of Fun How Video Games Blur the Boundaries of Work and Play. *Games and Culture*, *1*(1), 68–71. doi: 10.1177/1555412005281819

Yee, N. (2007). Motivations for Play in Online Games. *CyberPsychology & Behavior*, 772–775.

Zhang, P. (2008). Technical Opinion: Motivational Affordances: Reasons for ICT Design and Use. *Commun. ACM*, *51*(11), 145–147. doi: 10.1145/1400214.1400244

Zichermann, G. (2014). *Gamification is Scientifically Validated - Now What?* Retrieved 2016-03-16, from `http://www.gamification.co/2014/08/06/gamification-scientifically-validated-now/`

Zichermann, G., & Cunningham, C. (2011). *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps.* Sebastopol: O'Reilly Media.

Zichermann, G., & Linder, J. (2013). *The Gamification Revolution: How Leaders Leverage Game Mechanics to Crush the Competition.* New York: McGraw Hill Professional.

Zyda, M. (2005). From Visual Simulation to Virtual Reality to Games. *Computer*, *38*(9), 25–32. doi: 10.1109/MC.2005.297

Appendices

A Starting Terms for the Expert Survey

Table A.1: Terms used in the initial expert survey and example references to literature describing them. Many of these terms can be found in more than one publication.

nr	term	example reference
41	accomplishment	(Seligman, 2011)
1	achievable task(s)	(Csíkszentmihályi, 2008)
11	aesthetics	(Kapp, 2012)
21	autonomy	(Deci, Koestner, & Ryan, 2001)
31	avatars / virtual worlds	(Blohm & Leimeister, 2013)
51	badges / achievements	(Blohm & Leimeister, 2013)
2	being the hero	(Zichermann & Cunningham, 2011)
12	challenge	(Malone, 1982)
22	challenges / quests / missions / tasks	(Blohm & Leimeister, 2013)
32	clear goals	(Csíkszentmihályi, 2008)
42	collecting	(Blohm & Leimeister, 2013)
52	competence	(Deci, Koestner, & Ryan, 2001)
3	competition	(Blohm & Leimeister, 2013)
13	concentration	(Csíkszentmihályi, 2008)
23	conflict	(Kapp, 2012)
33	control	(Lepper et al., 1996)
43	cooperation	(Blohm & Leimeister, 2013)
53	curiosity	(Malone, 1982)
4	curve of interest	(Kapp, 2012)
14	customization	(Zichermann & Cunningham, 2011)
24	documentation of behavior	(Blohm & Leimeister, 2013)
34	effortless involvement	(Csíkszentmihályi, 2008)
44	fame, getting attention	(Zichermann & Cunningham, 2011)
54	fantasy	(Malone, 1982)
5	feedback	(Csíkszentmihályi, 2008)

nr	term	example reference
15	fixed rules	(Juul, 2005)
25	gaining status	(Zichermann & Cunningham, 2011)
35	gifting	(Zichermann & Cunningham, 2011)
45	goals	(Kapp, 2012)
55	group tasks	(Blohm & Leimeister, 2013)
6	intangible rewards	(Deci, Koestner, & Ryan, 2001)
16	leaderboards	(Blohm & Leimeister, 2013)
26	leading others	(Zichermann & Cunningham, 2011)
36	learning / mastery	(Pink, 2011)
46	levels	(Blohm & Leimeister, 2013)
56	meaning	(Seligman, 2011)
7	negotiable consequences	(Juul, 2005)
17	nurturing, growing	(Zichermann & Cunningham, 2011)
27	onboarding	(Zichermann & Cunningham, 2011)
37	organizing and creating order	(Blohm & Leimeister, 2013)
47	pattern recognition	(Zichermann & Cunningham, 2011)
57	player attachment to outcome	(Juul, 2005)
8	player effort	(Juul, 2005)
18	points	(Blohm & Leimeister, 2013)
28	positive emotions	(Seligman, 2011)
38	progress	(Huotari & Hamari, 2012)
48	purpose	(Pink, 2011)
58	relatedness	(Deci, Koestner, & Ryan, 2001)
9	relationships	(Seligman, 2011)
19	replay or do over	(Kapp, 2012)
29	reward structures	(Kapp, 2012)
39	social engagement loops	(Zichermann & Cunningham, 2011)
49	social incentives	(Hamari & Koivisto, 2013)
59	storytelling	(Kapp, 2012)
10	surprise and unexpected delight	(Zichermann & Cunningham, 2011)
20	tangible rewards	(Deci, Koestner, & Ryan, 2001)
30	time pressure	(Blohm & Leimeister, 2013)
40	valorization of outcome	(Juul, 2005)
50	variable outcome	(Juul, 2005)
60	virtual trade	(Blohm & Leimeister, 2013)

1

B Terms with High Expert Ratings

Table B.1: Terms that at least 66% of experts considered to be relevant in either category. Percentages rounded to the nearest integer.

	spec. implementation		gamification in general	
Term	mentions	%	mentions	%
achievable task(s)	13	100%	13	100%
accomplishment	12	92%	13	100%
goals	12	92%	13	100%
challenge	11	85%	13	100%
points	12	92%	12	92%
player effort	11	85%	12	92%
progress	11	85%	12	92%
badges/achievements	11	85%	12	92%
competition	10	77%	12	92%
clear goals	10	77%	12	92%
challenges/quests/missions/tasks	9	69%	12	92%
reward structures	12	92%	11	85%
positive emotions	11	85%	11	85%
feedback	10	77%	11	85%
fixed rules	10	77%	11	85%
documentation of behavior	10	77%	11	85%
learning/mastery	9	69%	11	85%
collecting	8	62%	11	85%
leaderboards	7	54%	11	85%
levels	6	46%	11	85%
intangible rewards	5	38%	11	85%
social incentives	5	38%	11	85%
purpose	9	69%	10	77%
concentration	7	54%	10	77%
cooperation	6	46%	10	77%
curiosity	6	46%	10	77%
variable outcome	5	38%	10	77%
storytelling	4	31%	10	77%

	spec. implementation		gamification in general	
Term	mentions	%	mentions	%
group tasks	2	15%	10	77%
fame, getting attention	8	62%	9	69%
autonomy	7	54%	9	69%
relatedness	6	46%	9	69%
surprise and unexpected delight	5	38%	9	69%
nurturing, growing	5	38%	9	69%
tangible rewards	5	38%	9	69%
time pressure	5	38%	9	69%
social engagement loops	5	38%	9	69%
player attachment to outcome	10	77%	8	62%
competence	9	69%	8	62%
Term	mentions	%	mentions	%
achievable task(s)	13	100%	13	100%
accomplishment	12	92%	13	100%
goals	12	92%	13	100%
challenge	11	85%	13	100%
points	12	92%	12	92%
player effort	11	85%	12	92%
progress	11	85%	12	92%
badges/achievements	11	85%	12	92%
competition	10	77%	12	92%
clear goals	10	77%	12	92%
challenges/quests/missions/tasks	9	69%	12	92%
reward structures	12	92%	11	85%
positive emotions	11	85%	11	85%
feedback	10	77%	11	85%
fixed rules	10	77%	11	85%
documentation of behavior	10	77%	11	85%
learning/mastery	9	69%	11	85%
collecting	8	62%	11	85%
leaderboards	7	54%	11	85%
levels	6	46%	11	85%
intangible rewards	5	38%	11	85%
social incentives	5	38%	11	85%
purpose	9	69%	10	77%
concentration	7	54%	10	77%
cooperation	6	46%	10	77%
curiosity	6	46%	10	77%
variable outcome	5	38%	10	77%
storytelling	4	31%	10	77%

Term	spec. implementation		gamification in general	
	mentions	%	mentions	%
group tasks	2	15%	10	77%
fame, getting attention	8	62%	9	69%
autonomy	7	54%	9	69%
relatedness	6	46%	9	69%
surprise and unexpected delight	5	38%	9	69%
nurturing, growing	5	38%	9	69%
tangible rewards	5	38%	9	69%
time pressure	5	38%	9	69%
social engagement loops	5	38%	9	69%
player attachment to outcome	10	77%	8	62%
competence	9	69%	8	62%

C Gamification in LMS- Result Tables

Table C.1: All responses from the four raters (R1–4) for each analyzed LMSs. A yes was recoded as 1, a maybe as 0.5 and a no as 0, so that the variance for each item for the respective system could be calculated.

System	Item	R1	R2	R3	R4	Variance
Blackboard	E1 - accomplishment	1	1	NA	1	0
Canvas	E1 - accomplishment	1	1	NA	1	0
Edmodo	E1 - accomplishment	1	1	NA	1	0
Moodle	E1 - accomplishment	1	1	NA	1	0
Schoology	E1 - accomplishment	1	1	NA	1	0
Blackboard	E2 - autonomy	1	1	1	1	0
Canvas	E2 - autonomy	1	1	1	1	0
Edmodo	E2 - autonomy	1	1	1	1	0
Moodle	E2 - autonomy	1	1	1	1	0
Schoology	E2 - autonomy	1	0.5	1	1	0.046875
Blackboard	E3 - challenge	1	1	0	0.5	0.171875
Canvas	E3 - challenge	1	0.5	0.5	0.5	0.046875
Edmodo	E3 - challenge	1	1	0.5	0.5	0.0625
Moodle	E3 - challenge	1	0.5	1	0.5	0.0625
Schoology	E3 - challenge	1	0.5	0	0.5	0.125
Blackboard	E4 - competence	0.5	0	0.5	0.5	0.046875
Canvas	E4 - competence	0.5	0.5	1	0.5	0.046875
Edmodo	E4 - competence	0.5	0.5	1	0.5	0.046875
Moodle	E4 - competence	0.5	0.5	1	0.5	0.046875
Schoology	E4 - competence	0.5	0.5	0	0.5	0.046875
Blackboard	E5 - curiosity	0	0	0	0	0
Canvas	E5 - curiosity	0	0	1	0	0.1875
Edmodo	E5 - curiosity	0	0	1	0	0.1875
Moodle	E5 - curiosity	0	0	0	0	0
Schoology	E5 - curiosity	0	0	0	0	0
Blackboard	E6 - learning/mastery	1	0	0	1	0.25
Canvas	E6 - learning/mastery	1	1	0	1	0.1875

System	Item	R1	R2	R3	R4	Variance
Edmodo	E6 - learning/mastery	1	0	0.5	1	0.171875
Moodle	E6 - learning/mastery	1	0	0	1	0.25
Schoology	E6 - learning/mastery	1	0	0.5	1	0.171875
Blackboard	E7 - player attachment to outcome	0.5	0	0.5	0.5	0.046875
Canvas	E7 - player attachment to outcome	0.5	0	0.5	0.5	0.046875
Edmodo	E7 - player attachment to outcome	0.5	0	NA	0	0.055556
Moodle	E7 - player attachment to outcome	0.5	0	0	0	0.046875
Schoology	E7 - player attachment to outcome	0.5	0	0.5	0	0.0625
Blackboard	E8 - positive emotions	0.5	0	0	1	0.171875
Canvas	E8 - positive emotions	0.5	0	0	1	0.171875
Edmodo	E8 - positive emotions	0.5	0	0	1	0.171875
Moodle	E8 - positive emotions	0.5	0	0	1	0.171875
Schoology	E8 - positive emotions	0.5	0	0.5	1	0.125
Blackboard	E9 - purpose	0.5	0	0	0	0.046875
Canvas	E9 - purpose	0.5	0	0.5	0	0.0625
Edmodo	E9 - purpose	0.5	0	0.5	0	0.0625
Moodle	E9 - purpose	0.5	0	0.5	0	0.0625
Schoology	E9 - purpose	0.5	0	0.5	0	0.0625
Blackboard	E10 - relatedness	1	0.5	0	1	0.171875
Canvas	E10 - relatedness	1	1	1	1	0
Edmodo	E10 - relatedness	1	1	1	0	0.1875
Moodle	E10 - relatedness	1	1	0.5	0.5	0.0625
Schoology	E10 - relatedness	1	1	1	0	0.1875
Blackboard	E11 - concentration	0	0	0.5	0	0.046875
Canvas	E11 - concentration	0	0	1	0	0.1875
Edmodo	E11 - concentration	0	0	0.5	0	0.046875
Moodle	E11 - concentration	0	0	0.5	0	0.046875
Schoology	E11 - concentration	0.5	0	1	0	0.171875
Blackboard	G1 - achievable task(s)	0	0.5	0.5	0.5	0.046875
Canvas	G1 - achievable task(s)	0	0	0.5	0.5	0.0625
Edmodo	G1 - achievable task(s)	0	0	0.5	0.5	0.0625
Moodle	G1 - achievable task(s)	0	0	0.5	0.5	0.0625
Schoology	G1 - achievable task(s)	0	0	0	0.5	0.046875
Blackboard	G2 - challenges/quests/missions/tasks	0	1	0	1	0.25
Canvas	G2 - challenges/quests/missions/tasks	0	0	1	0.5	0.171875
Edmodo	G2 - challenges/quests/missions/tasks	0	0	1	0.5	0.171875
Moodle	G2 - challenges/quests/missions/tasks	0	0	1	0.5	0.171875
Schoology	G2 - challenges/quests/missions/tasks	0	0	0	0.5	0.046875
Blackboard	G3 - clear goals	0.5	0.5	1	1	0.0625
Canvas	G3 - clear goals	0.5	0	0	1	0.171875

System	Item	R1	R2	R3	R4	Variance
Edmodo	G3 - clear goals	0.5	0	1	0.5	0.125
Moodle	G3 - clear goals	0.5	0	1	1	0.171875
Schoology	G3 - clear goals	0.5	0	0.5	1	0.125
Blackboard	M1 - collecting	1	0	0	1	0.25
Canvas	M1 - collecting	1	0.5	1	1	0.046875
Edmodo	M1 - collecting	1	1	1	1	0
Moodle	M1 - collecting	1	1	1	1	0
Schoology	M1 - collecting	1	1	0	1	0.1875
Blackboard	M10 - variable outcome	0.5	0.5	0.5	0.5	0
Canvas	M10 - variable outcome	0.5	0.5	0.5	0.5	0
Edmodo	M10 - variable outcome	0.5	1	0.5	0.5	0.046875
Moodle	M10 - variable outcome	0.5	1	0.5	0.5	0.046875
Schoology	M10 - variable outcome	0.5	0.5	0.5	0.5	0
Blackboard	M2 - documentation of behavior	0.5	0	1	1	0.171875
Canvas	M2 - documentation of behavior	0.5	0.5	0	1	0.125
Edmodo	M2 - documentation of behavior	0.5	0	0.5	0.5	0.046875
Moodle	M2 - documentation of behavior	0.5	1	1	1	0.046875
Schoology	M2 - documentation of behavior	0.5	1	0.5	1	0.0625
Blackboard	M3 - feedback	0.5	0.5	0	0.5	0.046875
Canvas	M3 - feedback	0.5	0.5	0.5	0.5	0
Edmodo	M3 - feedback	0.5	0	0	1	0.171875
Moodle	M3 - feedback	0.5	0	0.5	0.5	0.046875
Schoology	M3 - feedback	0.5	0	0.5	0.5	0.046875
Blackboard	M4 - fixed rules	0.5	0	0	1	0.171875
Canvas	M4 - fixed rules	0.5	0	0.5	0	0.0625
Edmodo	M4 - fixed rules	0.5	0	0	0	0.046875
Moodle	M4 - fixed rules	0.5	0	0.5	0	0.0625
Schoology	M4 - fixed rules	0.5	0	0	1	0.171875
Blackboard	M5 - nurturing, growing	1	0.5	0.5	0	0.125
Canvas	M5 - nurturing, growing	1	0.5	0	0	0.171875
Edmodo	M5 - nurturing, growing	1	0	1	0	0.25
Moodle	M5 - nurturing, growing	1	0	0.5	0	0.171875
Schoology	M5 - nurturing, growing	1	0	0	0	0.1875
Blackboard	M6 - player effort	0.5	1	1	1	0.046875
Canvas	M6 - player effort	0.5	1	1	1	0.046875
Edmodo	M6 - player effort	0.5	1	1	0	0.171875
Moodle	M6 - player effort	0.5	1	1	1	0.046875
Schoology	M6 - player effort	0.5	0.5	0.5	1	0.046875
Blackboard	M7 - storytelling	0	0	0	0	0
Canvas	M7 - storytelling	0	0	0	0	0

System	Item	R1	R2	R3	R4	Variance
Edmodo	M7 - storytelling	0	0	0	0	0
Moodle	M7 - storytelling	0	0	0	0	0
Schoology	M7 - storytelling	0	0	0	0	0
Blackboard	M8 - surprise and unexpected delight	0	0	0.5	0	0.046875
Canvas	M8 - surprise and unexpected delight	0	0	0	0	0
Edmodo	M8 - surprise and unexpected delight	0	0	0.5	0	0.046875
Moodle	M8 - surprise and unexpected delight	0	0	0.5	0	0.046875
Schoology	M8 - surprise and unexpected delight	0	0	0	0	0
Blackboard	M9 - time pressure	0.5	1	1	0	0.171875
Canvas	M9 - time pressure	0.5	1	1	0	0.171875
Edmodo	M9 - time pressure	0.5	0	1	0	0.171875
Moodle	M9 - time pressure	0.5	1	0	0	0.171875
Schoology	M9 - time pressure	0.5	1	0.5	1	0.0625
Blackboard	R1 - badges achievements	1	0	0	1	0.25
Canvas	R1 - badges achievements	1	1	0	1	0.1875
Edmodo	R1 - badges achievements	1	1	1	1	0
Moodle	R1 - badges achievements	1	1	1	1	0
Schoology	R1 - badges achievements	1	1	1	1	0
Blackboard	R2 - intangible rewards	1	0	0.5	1	0.171875
Canvas	R2 - intangible rewards	1	1	0.5	1	0.046875
Edmodo	R2 - intangible rewards	1	1	0.5	1	0.046875
Moodle	R2 - intangible rewards	1	1	0.5	1	0.046875
Schoology	R2 - intangible rewards	1	1	0	1	0.1875
Blackboard	R3 - levels	0	0	0	0	0
Canvas	R3 - levels	0	0	0	0	0
Edmodo	R3 - levels	0	0	0	0	0
Moodle	R3 - levels	0	0	0.5	0	0.046875
Schoology	R3 - levels	0	0	0.5	0	0.046875
Blackboard	R4 - points	0	0.5	0	0	0.046875
Canvas	R4 - points	0	1	1	0	0.25
Edmodo	R4 - points	0	1	0	0	0.1875
Moodle	R4 - points	0	1	1	0	0.25
Schoology	R4 - points	0	1	0	0	0.1875
Blackboard	R5 - progress	1	0.5	0.5	1	0.0625
Canvas	R5 - progress	1	1	0	1	0.1875
Edmodo	R5 - progress	1	1	1	0.5	0.046875
Moodle	R5 - progress	1	1	0.5	1	0.046875
Schoology	R5 - progress	1	0	0.5	1	0.171875
Blackboard	R6 - rewards	1	0	0	1	0.25
Canvas	R6 - rewards	1	1	1	1	0

System	Item	R1	R2	R3	R4	Variance
Edmodo	R6 - rewards	1	1	1	1	0
Moodle	R6 - rewards	1	1	1	1	0
Schoology	R6 - rewards	1	1	1	1	0
Blackboard	R7 - tangible rewards	0	0	0	0	0
Canvas	R7 - tangible rewards	0	0	0.5	0	0.046875
Edmodo	R7 - tangible rewards	0	0	1	0	0.1875
Moodle	R7 - tangible rewards	0	0	1	0	0.1875
Schoology	R7 - tangible rewards	0	0	0	0	0
Blackboard	S1 - competition	0	0	0.5	0	0.046875
Canvas	S1 - competition	0	1	0.5	0	0.171875
Edmodo	S1 - competition	0	0.5	0.5	0	0.0625
Moodle	S1 - competition	0	0	1	1	0.25
Schoology	S1 - competition	0	0.5	0.5	0	0.0625
Blackboard	S2 - cooperation	0.5	0	0.5	1	0.125
Canvas	S2 - cooperation	0.5	1	1	1	0.046875
Edmodo	S2 - cooperation	0.5	1	1	0	0.171875
Moodle	S2 - cooperation	0.5	1	1	0.5	0.0625
Schoology	S2 - cooperation	0.5	1	1	0	0.171875
Blackboard	S3 - fame, getting attention	0	0	0	0	0
Canvas	S3 - fame, getting attention	0	0	0	0	0
Edmodo	S3 - fame, getting attention	1	0	1	0	0.25
Moodle	S3 - fame, getting attention	0	0	0.5	0.5	0.0625
Schoology	S3 - fame, getting attention	0	0	0	0	0
Blackboard	S4 - group tasks	1	0	0.5	0.5	0.125
Canvas	S4 - group tasks	0	0	0.5	0.5	0.0625
Edmodo	S4 - group tasks	0.5	0	0.5	0	0.0625
Moodle	S4 - group tasks	1	0	0.5	0	0.171875
Schoology	S4 - group tasks	0.5	0	0.5	0	0.0625
Blackboard	S5 - leaderboards	0	0	0	0	0
Canvas	S5 - leaderboards	0	0	0.5	0	0.046875
Edmodo	S5 - leaderboards	0	0	0.5	0	0.046875
Moodle	S5 - leaderboards	0	0	0	1	0.1875
Schoology	S5 - leaderboards	0	0	0	0	0
Blackboard	S6 - social engagement loops	1	0	1	1	0.1875
Canvas	S6 - social engagement loops	1	0	0	1	0.25
Edmodo	S6 - social engagement loops	1	0	1	0	0.25
Moodle	S6 - social engagement loops	1	1	0.5	0	0.171875
Schoology	S6 - social engagement loops	1	1	0.5	0	0.171875
Blackboard	S7 - social incentives	0.5	0	0	0	0.046875
Canvas	S7 - social incentives	0.5	0	0	0	0.046875

System	Item	R1	R2	R3	R4	Variance
Edmodo	S7 - social incentives	0.5	0	0.5	0	0.0625
Moodle	S7 - social incentives	0.5	1	0.5	0	0.125
Schoology	S7 - social incentives	0.5	1	0.5	0	0.125

Table C.2: Results of the evaluation of gamification in LMSs for all categories.

Category	Moodle	Edmodo	Blackboard	Canvas	Schoology
E1 - accomplishment	Y	Y	Y	Y	Y
E10 - relatedness	M	N	Y	Y	N
E11 - concentration	N	N	N	N	N
E2 - autonomy	Y	Y	Y	Y	Y
E3 - challenge	M	M	M	M	M
E4 - competence	M	M	M	M	M
E5 - curiosity	N	N	N	N	N
E6 - learning/mastery	Y	Y	Y	Y	Y
E7 - player attachment to outcome	N	N	M	M	N
E8 - positive emotions	Y	Y	Y	Y	Y
E9 - purpose	N	N	N	N	N
G1 - achievable task(s)	M	M	M	M	M
G2 - challenges/quests/missions/tasks	M	M	Y	M	M
G3 - clear goals	Y	M	Y	Y	Y
M1 - collecting	Y	Y	Y	Y	Y
M10 - variable outcome	M	M	M	M	M
M2 - documentation of behavior	Y	M	Y	Y	Y
M3 - feedback	M	Y	M	M	M
M4 - fixed rules	N	N	Y	N	Y
M5 - nurturing, growing	N	N	N	N	N
M6 - player effort	Y	N	Y	Y	Y
M7 - storytelling	N	N	N	N	N
M8 - surprise and unexpected delight	N	N	N	N	N
M9 - time pressure	N	N	N	N	Y
R1 - badges achievements	Y	Y	Y	Y	Y
R2 - intangible rewards	Y	Y	Y	Y	Y
R3 - levels	N	N	N	N	N
R4 - points	N	N	N	N	N
R5 - progress	Y	M	Y	Y	Y
R6 - rewards	Y	Y	Y	Y	Y
R7 - tangible rewards	N	N	N	N	N
S1 - competition	Y	N	N	N	N
S2 - cooperation	M	N	Y	Y	N
S3 - fame, getting attention	M	N	N	N	N
S4 - group tasks	N	N	M	M	N
S5 - leaderboards	Y	N	N	N	N
S6 - social engagement loops	N	N	Y	Y	N
S7 - social incentives	N	N	N	N	N

Table C.3: Coded qualitative responses by all raters for each analyzed system. Numbers in brackets indicate the multiple mentions.

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
E1	badges (3), grades (2), attendance, quiz	badges (2), grades (2), points from tests, feedback, wiki	badges (2), points (3), mastery system	badges (2), grades, progress overview, points in quizzes, successful completion of tasks/ courses	badges (2), grades (2), quiz system
E2	free choice of task and time (3), menus, navigation	free choice of task and time (3), menus, navigation	free choice of task and time (2), menus, navigation, mastery system	free choice of task and time (3), free navigation	free choice of task and time (3), free navigation
E3	depends on instructor input (2), quizzes	depends on instructor input (2), quizzes	depends on instructor input (2)	depends on instructor input (3), quizzes	quizzes and tasks (2), depends on instructor input
E4	depends on instructor input (2), assignments, quizzes (1,m)	difficult navigation, assignments, quizzes, wiki, depends on instructor input	depends on instructor input (2)	depends on instructor input (2)	depends on instructor input (2), submission of assignments and quizzes
E5	-	clear navigation	design, menus, descriptions, symbols	clearly arranged	design encourages exploration

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
E6	mastery system, retaking of quests, assignments, quizzes	assignments, quizzes, communication, retaking of assignments/ quizzes	mastery system, retaking of assignments, clearly visible points	retaking of quizzes and assignments, assignments, quizzes, exchange of communication	assignments, quizzes, exchange with teacher/ students, formative assessment
E7	social network component, badges	social network component, badges, personalized responses	audio and video feedback, personalisation possible	social network component, badges	social network component, badges
E8	badges (2), grades, quizzes, assignments, events	successful completion (2), badges (2), feedback, wiki	badges, grades	completing assignments (2), badges (2), wiki, progress	completion of assignments (2), badges (2)
E9	fulfilling assignments (1,m)	fulfilling assignments, wiki	-	completing assignments, wiki	fulfilling assignments
E10	social network features (3)	social features (3), wiki	communication tools (2), group tasks	social network functions (4), wiki, blogs	social network component, news on front page, social contact limited on purpose
E11	depends on instructor input, current tasks marked red	-	announcements	-	-
G1	depends on instructor input	depends on instructor input (2), not visible	depends on instructor input	depends on instructor input	depends on instructor input

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
G2	depends on instructor input	encouraged, supported (2)	depends on instructor input, tasks/ quizzes	depends on instructor input	assignments, depends on instructor input
G3	instructor can set goals (2,y,m), overloaded front page	depends on instructor input (2), tasks clearly shown, tasks aligned to course goals, options clearly visible	depends on instructor input, many choices in menu	notifications (2), teacher can set goals	intuitive menu, depends on instructor input (2)
M1	badges (3), grades, attendance	badges (2), grades	badges, points, tool roll call	badges (3), grades	badges (4), grades
M2	course progress, activity digest for parents, postings, quizzes, grades, attendance, unclear, analytics	system history, progress overview (2)	activity stream, tool roll call	progress/ activity log	progress area, postings, quizzes, grades, parents can monitor behavior
M3	depends on instructor input (2)	completed assignments, points, grades, depends on instructor input	depends on instructor input, mastery system shows goal alignment, points	depends on instructor input (2)	based on instructor input (2), snapshot system
M4	depends on instructor input (2)	depends on instructor input, fixed criteria for badges	-	instructors can set rules for assignments/ quizzes	teachers can set rules for assignments and quizzes

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
M5	badges, attendance, grades	completed assignments, points, badges, wiki, course completion	points in tasks and quizzes	badges, wiki, course completion	badges, creating one's own tasks
M6	course/task completion (2)	course/task completion (3), creation of courses/ tasks	successful completion of tasks and quizzes (2)	completing assignments/ course/ quizzes (3)	fulfilling assignments (2), depends on difficulty, success is defined externally
M7	-	-	-	-	-
M8	-	-	-	-	-
M9	in quizzes (3)	due dates (3)	deadlines (2), time limit in quiz	time limit in quizzes, deadlines	deadlines (2), time limits
M10	grades (3), pass/fail	grades/ numerical outcomes (3)	numerical outcomes (2)	teacher grades and comments assignments, in quizzes, pass/fail and grades	numerical outcomes (grades/ points) (3)
R1	badges (4)	badges (2)	badges (2)	badges (4)	badges (3), points
R2	badges (3), grades, points	badges (2), awards, grades	badges (2)	badges (3), grades, feedback in quizzes	badges (4), grades
R3	list of absences, grades	-	-	-	-

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
R4	in quizzes and tasks	points for tests	points in tasks and quizzes	grades	points for answering questions
R5	course and mastery progress, grades	grades, course progress, number of completed tasks	mastery system, points	grades (2), badges, feedback	badges, grades, progress area, snapshots
R6	badges (3), grades, points	badges (2), awards, grades	badges, points/ grades, tool roll call	badges (3), grades, feedback in quizzes	badges (4), grades
R7	-	-	-	-	fulfillment of tasks, participation in surveys/quizzes
S1	badges (1,m)	-	tool roll call allows for observation	leaderboard, badges, points	badges
S2	communication features (3)	tools for collaboration (3)	collaboration systems (3)	communication systems (4)	groups/ communication possibilities
S3	-	-	-	through leaderboard plug-in	alerts, student of the month badge
S4	group assignments	depends on instructor input (2), wiki	depends on instructor input	group assignments, wiki	group assignments (2)
S5	-	-	grade list, results of courses/tasks	as a plug-in	-

Term	Schoology	Blackboard	Canvas	Moodle	Edmodo
S6	social network functions (2), badges, attendance	notifications (2), social network functions (2), calendar events, wiki, badges	notifications	badges (2), social network functions (2)	group messages, social network functions, badges
S7	social network functions (2)	social network, discussion boards, wiki	-	social network functions (2), wiki	social network component

D The Gamification Inventory

D.1 Version 1 of the Gamification Inventory

The following pages contain the Gamification Inventory in its first version. This is the version that was used to evaluate the five learning management systems (LMSs).

Experiential factors		
accomplishment		Does the system provide the user with a sense of accomplishment?
Y/N/M		
autonomy		Does the system give the user a feeling of autonomy? Can the user decide what to to next?
Y/N/M		
challenge		Does the system include tasks designed to be challenging for the user to complete?
Y/N/M		
competence		Does the system give the user a feeling of competence?
Y/N/M		
curiosity		Does the system include measures to pique the user's curiosity?
Y/N/M		
learning/ mastery		Does the system promote learning / mastery?
Y/N/M		
player attachment to outcome		Does the system include affordances to increase the (emotional) attachment of the user to the outcome of his or her actions?
Y/N/M		
positive emotions		Does the system trigger positive emotions in the user?
Y/N/M		
purpose		Does the system provide the user with a feeling of purpose?
Y/N/M		
relatedness		Does the system provide the user with a feeling of relatedness to other users? Is there social contact?
Y/N/M		
concentration		Does the system require users to concentrate on the task at hand?
Y/N/M		

System mechanics		
collecting		Does the system provide opportunities for the user to collect things?
Y/N/M		
documentation of behaviour		Does the system visibly document the user's behavior?
Y/N/M		
feedback		Does the system provide meaningful feedback to the user?
Y/N/M		
fixed rules		Does the system implement a set of fixed rules?
Y/N/M		
nurturing, growing		Does the system provide the user with the capability to nurture or grow something?
Y/N/M		
player effort		Does the system require user effort for a successful outcome?
Y/N/M		
storytelling		Does the system make use of storytelling?
Y/N/M		
surprise and unexpected delight		Does the system make use of the element of surprise?
Y/N/M		
time pressure		Does the system limit the amount of time a user has to complete a task?
Y/N/M		
variable outcome		Does the system provide different outcomes?
Y/N/M		

Rewards		
rewards		Does the system provide rewards of any kind to the user?
Y/N/M		
tangible rewards		Does the system provide tangible rewards to the user?
Y/N/M		
intangible rewards		Does the system provide users with intangible rewards?
Y/N/M		
progress		Does the system provide a measure of progress to the user?
Y/N/M		
points		Does the system reward users with points of any kind?
Y/N/M		
levels		Does the system include levels or ranks that users can reach?
Y/N/M		
badges/ achievements		Is the user rewarded with badges or achievements?
Y/N/M		

Goals		
clear goals		Does the system provide the user with clear goals / ideas about what to do next?
Y/N/M		
challenges/quests/missions/tasks		Does the system divide work to be done into a set of challengest/quests/missions/tasks?
Y/N/M		
achievable task(s)		Does the system divide these tasks in such a way that they seem achievable to the user?
Y/N/M		

Social elements		
competition		Does the system foster competition between users?
Y/N/M		
leaderboards		Does the system include a form of leaderboard?
Y/N/M		
cooperation		Does the system provide the opportunity for cooperation between users?
Y/N/M		
group tasks		Does the system include tasks that are meant to be solved in a group?
Y/N/M		
fame, getting attention		Does the system provide means for the user to gain fame / attention?
Y/N/M		
social engagement loops		Does the system make use of social features to keep the users engaged or to re-engage them?
Y/N/M		
social incentives		Does the system make use of social incentives?
Y/N/M		

D.2 Version 2 of the Gamification Inventory

The following pages contain the second, revised version of the Gamification Inventory.

E1 accomplishment	**How is a sense of accomplishment…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E2 autonomy	**How is a feeling of autonomy on the side of the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E3 challenge	**How is appropriate challenge of the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E4 competence	**In what ways is a feeling of competence on the side of the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E5 curiosity	**How is a feeling of curiosity on the side of the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E6 learning/ mastery	**How are learning and mastery…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E7 player attachment to outcome	**How is an increase in the (emotional) attachment to the outcome of the user's actions…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E8 positive emotions	**How is the creation of positive emotions…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E9 purpose	**How is a feeling of purpose on the side of the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E10 relatedness	**How is a feeling of relatedness to other users…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
E11 concentration	**How is a requirement for concentration…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:

M1 collecting	**What can users collect in the system, that is…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M2 documentation of behaviour	**How is visible documentation of user behaviour …**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M3 feedback	**How is feedback about the user's actions…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M4 fixed rules	**What types of rules are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M5 nurturing, growing	**How is nurturing or growing…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M6 player effort	**What actions that require effort from the user are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M7 storytelling	**What kind of narrative is…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M8 surprise and unexpected delight	**In what way are surprise or unexpected delight…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M9 time pressure	**How are spontaneous decisions or reactions or a sense of urgency…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
M10 variable outcome	**How are varied outcomes...**	
… built into the system?	… possible with proper user input?	I'm not sure, because:

R1 rewards	**What kinds of rewards are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R2 tangible rewards	**What kinds of tangible rewards are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R3 intangible rewards	**What kinds of intangible rewards are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R4 progress	**In what way is the communication of progress to the user…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R5 points	**In what way are points…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R6 levels	**In what way are levels…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
R7 badges/ achievements	**What kinds of badges/achievements are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:

G1 clear goals	**How are clear goals…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
G2 challenges/quests/missions/tasks	**What kinds of challenges/quests/missions/tasks are...**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
G3 achievable task(s)	**How is the appearance that tasks are achievable…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:

S1 competition	**How is comparison with others…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S2 leaderboards	**How are leaderboards…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S3 cooperation	**How are cooperation and incentives for cooperation…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S4 group tasks	**What kind of group tasks are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S5 fame/ getting attention	**What ways of gaining fame or attention are…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S6 social engagement loops	**How are social features meant to keep users engaged or to reangage them…**	
… built into the system?	… possible with proper user input?	I'm not sure, because:
S7 social incentives	**What other social features are...**	
… built into the system?	… possible with proper user input?	I'm not sure, because:

D.3 Evaluation Guideline

The following is a guideline for use with the Gamification Inventory, explaining the included items in more detail and giving references to additional reading for each one. Please refer to these explanations whenever you are unsure about the meaning of a term. It is good practice to read the guidelines at least once before the first application of the inventory.

Experiential

The experiential category contains terms dealing with the user's experience, or, more specifically, with affordances for specific experiences. Since the instrument is intended to be used for the evaluation of a system, we cannot use the actual user experience in the evaluation but only the affordances therefor.

E1 - accomplishment: Accomplishment/achievement is part of Seligman's well-being theory (Seligman, 2011) as well as one of the components of player motivation in the studies of Yee (2007). Badges are even often referred to as "achievements" in gaming circles. A feeling of accomplishment can be caused by such rewards, but also by praise or a feeling of uniqueness. Please describe in this category, which features of the system you are evaluating can cause such a feeling of accomplishment and how that is done.

E2 - autonomy: Autonomy is one of the core drivers of intrinsic motivation in self-determination theory (Ryan & Deci, 2000) and has been adopted by various later authors as a key component of motivation and gamification. In a gamified system, autonomy usually refers to the ability of the user to choose what to do next and when to do it. Please describe here, how feelings of autonomy are created, e.g. what choices are left to the user.

E3 - challenge: Challenge can be found at the core of many definitions for games as well as Czikszentmihalyi's flow (Csíkszentmihályi, 2008). It generally refers to the difficulty of a task compared to the user's skill. Note, that the terms "challenge" and "challenges", while related, refer to different concepts. Please describe here both features that create difficulty for the user as well as those designed to adapt to the user's abilities.

E4 - competence: Perceived competence is seen as an integral part to intrinsic motivation in cognitive evaluation theory (Ryan & Deci, 2000). In a way, this can bee seen as a mix of accomplishment and challenge. In order to feel competent, a user must be able to succeed at the tasks given by the system while at the same time feeling challenged by them so as to not make them seem trivial. Please describe here, what properties of the system can lead to users feeling competent (while not necessarily actually being competent.)

E5 - curiosity: Curiosity is a property adopted from what is possibly the earliest work on digital gamification available: Malone (1982) on learning from computer games to inform user interface design. Curiosity usually involves a process of exploration or discovery. Please describe here features that enable such exploratory behavior and the way in which users are guided towards it.

E6 - learning/mastery: Koster (2005) sees learning at the core of game design and as the main cause for fun, while Pink (2011) stresses the importance of mastery for motivation. This item is not so much about the actual learning content of the system, but about elements that promote the mastery of certain skills. This could, for example, include features that slowly reveal the complexity of a task or repetitions intended to improve mastery.

E7 - player attachment to outcome: According to Juul (2003), in order for a game to be a game, it most not only have a variable and quantifiable outcome, the player must also care about that outcome (be attached to it). Please describe here those features of the system that have the potential to create an attachment of the user to the outcome of his or her actions. Essentially, why do users care about achieving a specific outcome?

E8 - positive emotions: Another term adopted from (Seligman, 2011), positive emotions may be seen as one goal of gamification. (For example to mediate attitude in Landers' model of gamified learning.) Please describe here the kind of positive emotions created and the means used to do so. One example could be praise given to the user, another could be the joy of success.

E9 - purpose: Pink (2011) discusses a similar concept to Seligman's "meaning", the reason behind taking specific action. Often, this refers to something more than personal gain - a larger community goal to reach, for example. Please describe here, how the user is given such a feeling of purpose for his or her actions that goes beyond the actions themselves.

E10 - relatedness: Relatedness, adopted from self-determination theory (Ryan & Deci, 2000), is an odd item as there is an entire category of items dedicated to social elements. We have included it in the experiential section of the instrument instead, as it is about the experience of the user. Please describe here the features that make users feel related to others. Seeing the progress and difficulties of others working on the same task as oneself might created such a feeling, for example.

E11 - concentration: Concentration is another requirement for the flow state (Csíkszentmihályi, 2008). Games are often good at making the user concentrate, gamification can attempt to use their elements to achieve a similar experience. Please describe here both the features that require a user's concentration and those that help the user concentrate on the task at hand.

System Mechanics

The second category, system mechanics, deals purely with the functionality of the system, rather than the experience of the user.

M1 - collecting: Collecting is a game-design element (more specifically: a game dynamic) mentioned by Blohm & Leimeister (2013). Many games contain such collection mechanics, be it the collection of items in the game or the collection of badges. Some games, such as the variety of Pokémon games, even include it as a core motivator.

M2 - documentation of behavior: Documentation of behavior is a game mechanic also mentioned by Blohm & Leimeister (2013). According to them, "The continuous documentation of one's own behavior visualizes progress, facilitates the derivation of achievable personal goals and offers immediate feedback so that users perceive feelings of high individual performance" (Blohm & Leimeister, 2013, p.4). Look for features that list or aggregate a user's past actions or milestone achievements.

M3 - feedback: Feedback is a value-laden response to the user's actions, usually giving the user information about whether an action was positive or negative and, optimally, why. Feedback can come in various form and at various speeds. See Gee (2005) for a discussion of feedback in games for learning and (Kettle & Häubl, 2010) for the impact of temporal proximity of feedback.

M4 - fixed rules: Fixed rules are a direct adoption from game definitions, such as Suits (2005) or Juul (2003) and at the core of the distinction between ludus and paidia. Unfortunately, rules are difficult to define, and potentially difficult to detect. Look for rules that are explicit, spelled out, rather than rules that describe the operation of a software system.

M5 - nurturing, growing: Zichermann & Cunningham (2011) mention nurturing or growing as a game mechanic that could be used for gamification. Examples include the Tamagotchi line of toys or games such as The Sims or Sim City. Interestingly, neither is usually considered to be an actual game. (Juul (2003) calls Sim City a borderline case.)

M6 - player effort: Player effort is another of Juul's criteria for games (Juul, 2003) and directly connected to unnecessary obstacles as described by Suits (2005). Look for elements that require users to think, concentrate, or put a significant amount of time into progress.

M7 - storytelling: The relevance of storytelling for games has been questioned in the past, specially in the debate between narratology and ludology (c.f. Frasca, 2003). While one might argue that storytelling is not an essential element of all games, it is a prominent element in many games and essential to some of them. Stories can be told explicitly, but characters and other narrative elements can also be a sign of storytelling.

M8 - surprise and unexpected delight: Another game element mentioned by Zichermann & Cunningham (2011), surprise and unexpected delight can take many forms in gamification. One example is a badge that the user did not know existed - which can, among other things, encourage exploration.

M9 - time pressure: Time pressure is another element that can be found in many games but does not seem to be prominent in gamification yet. Blohm & Leimeister (2013) mention it as a game mechanic for use in gamification, however, as do Deterding et al. (2011) in reference to Reeves & Read (2009). Time pressure can have a variety of effects when used in a game, such as increasing concentration or challenge. Look for short-term time pressure through countdowns, moving targets, etc., rather than long-term time-pressure, such as a deadline.

M10 - variable outcome: Yet another characteristic of games according to Juul (2003), there must be different outcomes of a game. A variable outcome can also be found in smaller segments, such as quests, if there is no actual defined end to using a system. Look for features such as grades or ratings or winning/losing.

Rewards

The rewards category highlights some mechanics that are very common in gamification, those dealing with extrinsic rewards – such as points or badges.

R1 - rewards: This is a rather generic catch-all item for the use of rewards for gamification. It was originally called "reward structures" in the questionnaire and adopted from Kapp (2012). Look for anything that users are rewarded for and with. R2 and R3 are subcategories of this item.

R2 - tangible rewards: Look for any rewards that actually affect the user outside the examined system, such as cash rewards or a free beverage. Note that a tangible reward does not necessarily need to be physical - a money transfer to your bank account would also be considered tangible, for example.

R3 - intangible rewards: All rewards that are not tangible. This could be fame and status, but also rewards within the examined system, such as points or badges.

R4 - progress: Progress can be shown to the user in a variety of ways. Progress bars are a common method, but levels and milestone badges are also often used. Huotari & Hamari (2016) provide a variety of examples.

R5 - points: Along with the following two items, points are ubiquitous in gamification. Points are often a numeric measure of progress (as in experience points) but can also be used as a currency that can be spent. Sometimes points can also be lost.

R6 - levels: Levels are even mentioned by Deterding et al. (2011), making them and badges the only means common to all definitions of gamification. That said, the term actually has two meanings in games. For one, a level can be an area of a game - be it a physical area of the game world or a new setup for the game to be played in. These types of levels usually have progressing difficulty. The term is also used as an indicator of the progress of characters in video games - the avatar of a player in a role-playing game typically gains levels as the player progresses through the game, for example, indicating strength. When we talk about levels in gamification, the latter meaning is what we are referring to. That is why the question also includes the term “rank”. Nevertheless, one needs to be careful not to confuse the two.

R7 - badges/achievements: Badges are the poster child of gamification. The concept has been transferred from video games, where players are awarded badges for a variety of achievements within the game, but originates in military and similar traditions. The underlying concept is always that there is a specific achievement that is being honored with a badge (as opposed to points, which can be gained from a variety of actions and do not show their origin.)

Goals

The goals category includes those items that stress the importance of goals and structured tasks in games. It is the smallest of the categories and it could be argued that the items therein could be included in other categories (such as mechanics) instead. Goals are integral to games, however, which is why we gave them a more prominent position in the instrument. The original survey actually included both “goals” and “clear goals” as terms. We combined these for simplicity’s sake, as the two terms were very similar.

G1 - clear goals: Many theories of motivation stress the importance of clear goals. Among others, clear goals are an element for the flow state as described by Csíkszentmihályi (2008). Look for features that let you know, what to do next within the examined system.

G2 - challenges/quests/missions/tasks: This item combines a variety of terms with similar meanings. The general idea is the subdivision of a final goal into smaller elements that the user can tackle. These are often called quests or missions in games. Blohm & Leimeister (2013) mention them in their list of game design elements. Look for a list of tasks, or tasks that are subdivided into smaller ones.

G3 - achievable task(s): It is important, that a task seems achievable to the user. Csíkszentmihályi (2008) stresses this as an important condition for flow and it also speaks to the feeling of competence suggested by self-determination theory (Ryan & Deci, 2000). Please note, that a task does not necessarily actually have to be achievable, it must only seem to be so when looking for this item. Look for how users are made to feel that the tasks they are given are achievable.

Social

Finally, the social category encompasses social functionalities that one might adopt from games.

S1 - competition: Competition involves the comparison of success between users. This can be achieved in a variety of ways, e.g. by showing a users success in comparison to others, but also by making users compete for rare resources.

S2 - leaderboards: Leaderboards are special form of competition and the main gamification element used for creating competition in common approaches to gamification. Leaderboards always show a ranking of users, but can come in a variety of forms, such as the no-disincentive-leaderboard which always shows direct competitors of the user instead of simply showing the top performers in the system. Leaderboards are usually based on points, but can use other measures of progress as well.

S3 - cooperation: Cooperation is relatively rare in gamified systems we have studied, but it can have powerful motivating effects (see Ryan & Deci, 2000). Look for features that enable users to work together with others and those that promote such cooperation.

S4 - group tasks: Group tasks are a gamification element that can be used to foster cooperation, as mentioned by Blohm & Leimeister (2013) among others. Unlike in S3, look specifically for tasks that are meant to be completed in a group (or even have to be).

S5 - fame, getting attention: Getting fame or attention is mentioned by Zichermann & Cunningham (2011) as another possible game element to use in gamification. Various gamification elements can be used to implement this - a leaderboard can draw attention to the most successful users or they might get special avatars, for example.

S6 - social engagement loops: Yet another element from Zichermann & Cunningham (2011), the social engagement loop is built on the idea that the interaction between users makes users continue using the system. This could involve notifications about the actions of other users as well as calls for help with a group activity. Also look for messages sent to the user outside the system (such as an e-mail, reminding the user of a pending invitation).

S7 - social incentives: This is a catch-all item for users being incentivized to use or continue to use as system through social means. This may be the usage of the system by peers or the possibility to connect to others socially. Hamari & Koivisto (2013) include a variety of examples of such incentives.

E Gamification Plug-ins for Moodle and DAKORA

E.1 Requirements

Table E.1: High level requirements for a system for use in an experiment to measure the effectiveness of gamification elements. The origin column contains a brief reason for each requirement, as they are derived from our previous work.

#	Requirement	Origin
1	The system must be able to automatically reward users with badges.	We identified badges as ubiquitous both in existing studies of gamification and in existing gamified systems. We have further discovered that features need to be independent of instructor input to avoid uncontrollable variables.
1.1	Badge rewards should be infrequent but prominent.	Related work shows that badges are used as infrequent but impactful rewards.
2	The system must be able to automatically reward users with points.	We identified points as ubiquitous both in existing studies of gamification and in existing gamified systems. We have further discovered that features need to be independent of instructor input to avoid uncontrollable variables.
2.1	Point rewards should be frequent but unintrusive.	Related work shows that points are best rewarded frequently. A high frequency requires low intrusiveness so as to avoid interrupting the use of the system.
3	All rewards should reflect progress.	Progress is seen as a core attribute of gamified systems both in literature and by our experts.

#	Requirement	Origin
4	The system must be configurable to use either badges, points, both, or neither.	A 2x2 experimental design requires that each experimental treatment can be toggled on or off.
5	The system must be able to record all user interactions that involve gamified features, even if the features are disabled.	Proper data analysis requires tracking of user interaction with the system. For comparison between different groups, it is important to track behavior relevant to our research question even if the gamified element they are relevant to is not enabled for the group. (e.g. actions that lead to points need to be tracked for all groups, even the ones that do not use points in order to see whether the use of points causes an increase in point-related activities.)
6	The system must display rewards to the user as soon as possible.	Related work shows that feedback in general and feedback in gamification in particular should be as directly connected to the action that caused it as possible.
7	The system must not promote competition	Related work shows that competition can impact a subset of users negatively. Also, experts were divided on its use.
	Field access specific requirements	
8	The system must integrate into DAKORA and the competency grid for Moodle.	Field access requires integration into the system used in Baden-Württemberg.
8.1	The system must include a plug-in to Moodle that handles interaction with the competency grid module, written in PHP.	Moodle uses PHP plug-ins for all extensions.
8.2	The system must include a plug-in to DAKORA, handling all display of user information, written in JavaScript	Students interact with the existing system using DAKORA, therefore our gamification elements need to be visible there. DAKORA uses JavaScript plug-ins for all extensions.
8.3	The two plug-ins must be able to communicate.	Data from the competency grid form the basis of the gamification elements in DAKORA.

#	Requirement	Origin
8.4	The system must be able to track interactions relevant to the gamified features, both in DAKORA and in Moodle.	In order to provide rewards for actions and in order to record interaction data, the system has to be informed of such actions.
9	All records of user behavior must be stored pseudonomously.	Privacy requirements for the use of the application in schools state, that we must not be able to connect the collected data to actual students.
10	The system should be unintrusive, adding no negative burden to the use of the existing system.	Requirement for field access.
11	The system should be extendable.	The system as proposed includes only very basic gamification elements, but once a large user base and experimental setup are established, additional experiments with more complicated gamification elements should be performed.

E.2 Source Code

The source code for our plug-ins is available online. The Moodle plug-in can be found at https://github.com/scrusi/dakora-gamification-plugin. While DAKORA also supports plug-ins, they are, at the current stage of development, not clearly separable from the rest of the applications. Therefore we can not provide a stand-alone version of our additions to DAKORA for download. The complete DAKORA app including our modifications can be found at https://github.com/ariepl/dakora in the experimental branch. Our modifications can be seen in detail at https://github.com/ariepl/dakora/commit/682b819b8df6c13767c46e3d4440b8a9b31684cc. All major functionality of our plug-in is handled in app.extern.gtn.dakora/www_debug/js/plugin/plugin.Gamification.js while the added page for progress is defined in app.extern.gtn.dakora/www_debug/js/page/page.gamification.js.

E.3 Editing Gamification Element Definitions

The gamification elements included in the plug-in (badges and points with levels) are defined in the source code, but can be easily modified to adapt the system to different experimental contexts. Each badge is defined in the badger class and consists of a name, a description, and an image. Additional badges can be adding by extending the array of badges in that file and adding an image (180x180px) to the appropriate folder. Array elements are numbered and it is important to keep the numbering consistent between upgrades as the numbers are used to map the badges defined in the badger class to the badges imported to Moodle upon installation and update. Listing E.1 shows an excerpt from these badge definitions.

Listing E.1: Excerpt from the badger class, defining badges to be used in the plug-in.

```
// Imports badges and fills the badgemap table
class badger {
    /**
    * array of badges to import
    **/
    private static $badges = array(
        5 => array (
            'name' => 'Ich kann das! - Bronze',
            'description' => 'Du hast eine deiner Kompetenzen
               ↪ eingeschätzt.',
            'image' => 'tutorial_completion_1.png' // in local/
               ↪ gamification/files/badgeimages
        ),
        6 => array (
            'name' => 'Ich kann das! - Silber',
            'description' => 'Du hast 10 deiner Kompetenzen
               ↪ eingeschätzt.',
            'image' => 'tutorial_completion_2.png' // in local/
               ↪ gamification/files/badgeimages
        ),
```

The awarding of points and badges is handled through a system of watchers and listeners in the event_handler class, as described above. Watchers and listeners are functions defined in that class. Both can include arbitrary code, but watchers are passed only the id of the triggering user while listeners are passed an object that defines the triggering event. In order to add a new watcher or listener, one needs to include an appropriate function and register it in either the watchers or the listeners array. These arrays define the conditions under which the function is called. Listeners look for a specific type of event to be triggered (usually a user action in DAKORA) while watchers check for counters to reach a certain value. Listing E.2 and Listing E.3 show excerpts from listener and watcher definitions, respectively.

Listing E.2: Excerpt from the event_handler class, showing example definitions of listeners.

```
class event_handler {
    /**
    * array of functions that listen to gamification events.
    * Actions without a listener will not be logged either.
    * key is the name of the event, value is an array of
      ↪ function names.
    **/
    private static $listeners = array(
        'grade_set_as_student' => array ('count_rate_self'),
        'example_added_to_calendar' => array('count_add_example'
          ↪ ),
        'grade_set_as_teacher' => array('
          ↪ check_correct_self_rating','check_all_siblings'),
        'example_submitted' => array('count_submit_example')
    );
    /**
    * Functions for use in triggers.
    **/
    private static function count_rate_self($event) {
```

```
        global $USER, $DB;
        $sql = 'SELECT count(*) as cnt FROM
           ↪ mdl_local_gamification_log WHERE  user_id = ? AND
           ↪ action = ? AND descriptor_id = ? AND value >=?';
        if ($DB->count_records_sql($sql,array($USER->id,'
           ↪ grade_set_as_student',$event->descriptor_id,$event
           ↪ ->value)) > 1) return false;
        $responses = new responses();
        $responses->add( self::update_counter(COUNT_RATE_SELF,
           ↪ $USER->id,'inc'));
        $responses->add(self::issue_points(POINTS_RATE_SELF,
           ↪ $USER->id));
        return $responses;

    }
```

Listing E.3: Excerpt from the event_handler class, showing examples for defining watchers.

```
/**
* functions that watch for certain counter levels to be reached
* outer key is the id of the counter to watch
* inner key is the number to look for
* values in the array are functions to call
**/
private static $watchers = array(
    COUNT_RATE_SELF => array(
        1 => array('badge_rate_self_bronze'),
        10 => array('badge_rate_self_silver'),
        30 => array('badge_rate_self_gold'),
        100 => array('badge_rate_self_diamond')
    ),
    COUNT_CORRECT_RATING => array(
```

```
        1 => array('badge_correct_rating_bronze'),
        10 => array('badge_correct_rating_silver'),
        30 => array('badge_correct_rating_gold')
    ),
);

/**
* Functions for use in watchers.
**/
private static function badge_rate_self_bronze($user_id) {
    return self::issue_badge(BADGE_RATE_SELF_BRONZE, $user_id);
}
private static function badge_rate_self_silver($user_id) {
    return self::issue_badge(BADGE_RATE_SELF_SILVER, $user_id);
}
```

F German Manual for our Plug-Ins

The following pages contain a brief German manual for our plug-ins, as provided to our partners at the Konzeptgruppe Moodle and the ministry of education and culture in Baden-Württemberg.

F.1 Einleitung

Dieses Plugin erweitert Dakora um einige Spielelemente, insbesondere Belohnungen für die Verwendung der App und das erreichen von Zielen. Das Plugin ist so gestaltet, dass es sich ohne weitere Erklärung in Dakora einfügt, Nutzer sollten für die Verwendung keine Erklärung benötigen.

F.2 Installation

Das Gamification Plugin besteht aus 2 Elementen - einem Plugin für Moodle und einem für Dakora.

F.2.1 Moodle Plugin installieren

- Entpacken Sie die local_gamification.zip Datei in das Unterverzeichnis "local" ihrer Moodle Installation (z.B. C:\xampp\htdocs\moodle\local oder /var/www/htdocs/moodle/local).
- Öffnen sie die Administrationsoberfläche Ihrer Moodle Installation.
- Moodle sollte Ihnen eine Aufforderung zur Installation des neuen Plugins anbieten.

- Klicken sie "Datenbank aktualisieren".

F.2.2 Dakora Plugin installieren

- In der finalen Version wird das Plugin bereits in Dakora enthalten sein.
- Zu Testzwecken können Sie die dakora_gamified.zip Datei in das document root ihres Webservers entpacken (z.B. C:\xammp\htdocs oder /var/www/htdocs)
- Eine modifizierte Version von Dakora ist jetzt auf ihrem Webserver under /dakora_gamified zu erreichen (z.B. http://localhost/dakora_gamified)

F.2.3 Moodle Plugin konfigurieren

- Die Konfigurationsmöglichkeiten für das Moodle Plugin sind minimal.
- Über die Moodle Konfigurationsdatei config.php können Sie das Plugin in einen von vier Modi versetzen in dem Sie die Variable $CFG->gamificationmode setzen.

```
$CFG->gamificationmode = 3;    // Settings for the
    ↪ gamification plugin.
                        // 0 = No visible changes
                        // 1 = Points only
                        // 2 = Badges only
                        // 3 = Points and Badges
                        // Note: This is based on a bitmask,
                        // allowing for hassle-free addition
                        // of binary options
```

- 0: Dakora verhält sich als ob das Plugin nicht installiert wäre
- 1: Nur die Punktfunktionen sind aktiviert (siehe unten)
- 2: Nur die Abzeichenfunktionen sind aktiviert (siehe unten)
- 3: Sowohl Punkte als auch Abzeichen sind aktiviert

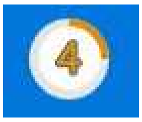
F.3 Verwendung

F.3.1 In Dakora

Das Plugin fügt, je nach gewähltem Modus, einige visuelle Elemente zu Dakora hinzu.

Punkte

- Nutzer und Nutzerinnen werden für bestimmte Aktionen mit Punkten belohnt. Aktuell sind das: Einschätzung einer eigenen Kompetenz, Bestätigung einer Kompetenz durch eine Lehrkraft, hinzufügen einer Aufgabe zum Kalender, erfüllen aller Teilkompetenzen, Abgabe einer Lösung
- Bei Punktgewinn wir kurz eine Meldung (z.B. "+50 Punkte") in der Kopfzeile eingeblendet.
- Bei ausreichend Punkten erhalten die Nutzer ein neues Level. Ein Levelaufstieg wird in einer Nachricht angezeigt.
- Das aktuelle Level der Nutzer wird in der Kopfzeile in einem Kreis angezeigt. Die Füllung des Kreises zeigt den Fortschritt zum nächsten Level.



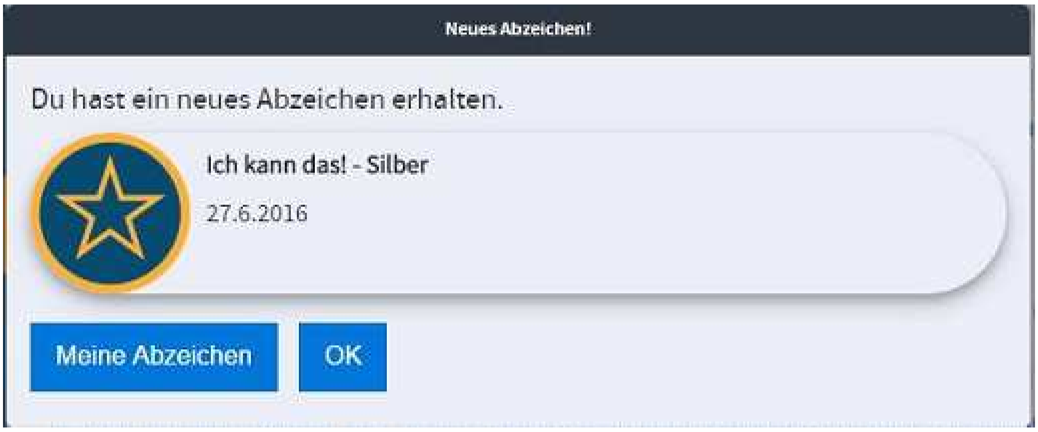
Levelanzeige. Ein Viertel der nötigen Punkte für Level 5 ist bereits erreicht.

- Jeder Aufstieg benötigt mehr Punkte als der vorhergehende. Nutzer und Nutzerinnen sollten also am Anfang schnell aufsteigen, dann immer langsamer.

Abzeichen

- Nutzer und Nutzerinnen werden für bestimmte erreichte Meilensteine mit Abzeichen belohnt. So bekommen sie z.B. ein Abzeichen für die erste (Bronze), zehnte(Silber), dreißigste (Gold), und hundertste (Diamant) abgegebene Selbsteinschätzung. Des weiteren gibt es aktuell Abzeichen für bestätigte Kompetenzen, hinzufügen von Aufgaben zum Kalender, und erreichen aller Teilkompetenzen in einem Lernfortschritt.

- Erhaltene Abzeichen werden in einer Nachricht angezeigt. Werden sie in Abwesenheit erhalten (z.B. weil eine Lehrkraft eine Kompetenz bestätigt hat.), werden sie beim nächsten Start der App angezeigt.



Benachrichtigung bei neuem Abzeichen.

- Von der Anzeige neuer Abzeichen können Nutzer und Nutzerinnen eine Übersichtsseite bereits erhaltener Abzeichen aufrufen. Diese Seite wird ebenfalls durch einen Klick auf die Levelanzeige erreicht.

- Im Modus 2 (nur Abzeichen) gibt es aktuell keinen direkten Link auf die Übersichtsseite. Dieser wird noch nachgerüstet.

F.3.2 In Moodle

Das moodle Plugin dient in erster Linie der Kommunikation mit Dakora und der Speicherung der Daten. Alle erreichbaren Abzeichen werden als normale Moodle Abzeichen gespeichert und sind auch in Moodle sichtbar und verfügbar. Die Abzeichen können in Moodle auch verändert werden, für die Testphase ist das allerdings nicht vorgesehen.

Fortschrittsseite.