TU WIEN Informatics

Social gamification elements on user engagement and user retention in gamified learning apps

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Software Engineering and Internet Computing

eingereicht von

Felix Olszewski, Bsc
Matrikelnummer 12022502

an der Fakultät für Informatik

der Technischen Universität Wien

Betreuung: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioglu
Mitwirkung: Univ.Ass. / PhD Kevin Marc Blasiak

Wien, 2. April 2025

Felix Olszewski	Hilda Tellioglu

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

TU WIEN Informatics

Social gamification elements on user engagement and user retention in learning apps

DIPLOMA THESIS

submitted in partial fulfillment of the requirements for the degree of

Diplom-Ingenieur

in

Software Engineering and Internet Computing

by

Felix Olszewski, Bsc
Registration Number 12022502

to the Faculty of Informatics

at the TU Wien

Advisor: Associate Prof. Dipl.-Ing.in Dr.in techn. Hilda Tellioglu
Assistance: Univ.Ass. / PhD Kevin Marc Blasiak

Vienna, April 2, 2025

Felix Olszewski	Hilda Tellioglu

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.at

Erklärung zur Verfassung der Arbeit

Felix Olszewski, Bsc

Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.
Ich erkläre weiters, dass ich mich generativer KI-Tools lediglich als Hilfsmittel bedient habe und in der vorliegenden Arbeit mein gestalterischer Einfluss überwiegt. Im Anhang „Übersicht verwendeter Hilfsmittel“ habe ich alle generativen KI-Tools gelistet, die verwendet wurden, und angegeben, wo und wie sie verwendet wurden. Für Textpassagen, die ohne substantielle Änderungen übernommen wurden, haben ich jeweils die von mir formulierten Eingaben (Prompts) und die verwendete IT- Anwendung mit ihrem Produktnamen und Versionsnummer/Datum angegeben.

Wien, 2. April 2025

Felix Olszewski

Kurzfassung

Da digitale Plattformen heutzutage zu einem zentralen Ort des Lernens geworden sind, ist es in unserem Interesse zu verstehen, wie wir die Motivation der Lernenden in einem solchen Kontext verbessern können. Ein Werkzeug, das als Lösung verwendet werden kann, ist das Werkzeug der Gamification. Mit unserer Forschung wollen wir herausfinden, wie soziale Gamification-Elemente wie ein Freunde-Newsfeed und ein "Lern-Streak" bei der Messung ihrer Effektivität zur Erhöhung der Benutzerbindung in einer gamifizierten Lern-App miteinander verglichen werden können. Zunächst haben wir eine qualitative Literaturrecherche durchgeführt, auf deren Grundlage wir eine gamifizierte mobile Lernanwendung entwickelt haben, gefolgt von einem quantitativen Benutzerexperiment mit dem Thema unserer entwickelten Lernanwendung, bei dem wir die Benutzerbindung mit Hilfe eines A/B-Testing-Ansatzes mit einer Kontrollgruppe A und zwei Versuchsgruppen B und C messen, denen jeweils zwei Gamification-Element-Sets bestehend aus einem Social Newsfeed und einem Lern-Streak Element zugewiesen wurden. Von den insgesamt 137 Nutzern, die an unserem Experiment teilgenommen haben, konnten wir 22 Nutzer während der Testphase an das Projekt binden. Unsere Ergebnisse zeigen, dass der Lern-Streak einen Unterschied von +19,41% zur Kontrollgruppe aufweist und der Newsfeed einen negativen Effekt von -35,69% hat. Aufgrund der geringen Teilnehmerzahl konnten unsere Ergebnisse jedoch keine statistische Signifikanz erreichen. Aus unserem Experiment und unserer Literaturübersicht schließen wir, dass einerseits Forscher mehr qualitative Experimente durchführen sollten, um Gamification auf einer tieferen Ebene zu verstehen, während andererseits Praktiker der Gamification sicherstellen sollten, dass Gamification mit einem ganzheitlichen Ansatz implementiert wird, um eine abgerundete spielähnliche Erfahrung zu gewährleisten.

Abstract

With digital platforms having become a central place where people nowadays learn, it is in our interest to understand how we can improve the learners motivation in such a context. One tool that can be used as a solution is the tool of gamification. With this research we aim to find out how social gamification elements like a friends news feed and a learning streak element compare when measuring their effectiveness of increasing user retention on a gamified learning app. First we performed a qualitative literature review based on which we developed a gamified mobile learning application, followed by a quantitative user experiment using the subject of our developed learning app where we measure user retention using an A/B testing approach with one control group A and two treatment variants B and C, getting assigned two gamification element sets consisting of a social news feed and a learning streak element respectively. From the total of 137 users that participated in our experiment we managed to retain 22 users during our testing trial. Our results indicate that the learning streak has a +19.41% retention difference from the control group and the news feed has a negative effect of -35.69%. However, the low number of participation did not allow our results to reach statistical significance. From our experiment and our literature review we conclude that on the one hand researchers should conduct more qualitative experiments to understand gamification on a deeper level while on the other hand gamification practitioners should make sure to implement gamification using a holistic approach, ensuring a well-rounded game-like experience.

Contents

Kurzfassung **vii**

Abstract **ix**

Contents **xi**

1 Introduction **1**

2 Background and Current State-of-the-Art **3**
2.1 Gamification 5
2.2 User Engagement and User Retention 21
2.3 A lack of quantitative research 22
2.4 A/B Testing 22
2.5 Hypotheses and Research Assumptions 24

3 Methods **27**
3.1 Pragmatism, Research Strategy 27
3.2 Research Methods 28
3.3 Implementation of GeoChamp Prior to Gamification 28
3.4 Literature Review 35
3.5 A/B Testing 39
3.6 The A/B Testing Framework implementation 41

4 Results **43**
4.1 Gamification Features Derived from the Literature Review 43
4.2 Analysis of the feature set of GeoChamp in regards to Gamification . 54
4.3 Results of the A/B Testing Experiment 56

5 Discussion **61**
5.1 The potential of gamification in learning 61
5.2 Practical Evaluation of Gamification 63

6 Conclusion and Future Work **65**

Overview of Generative AI Tools Used **67**

List of Figures **69**

List of Tables **71**

List of Algorithms **73**

Appendix **77**

Bibliography **81**

CHAPTER 1

Introduction

Learning is increasingly happening with the aid of digital tools that are ought to excite and enhance the learning process. During the COVID-19 Pandemic e-learning saw a spike because it facilitated educating students through digital means while meeting in person was not possible. This shift has highlighted the importance of effective digital learning tools in maintaining educational continuity. In this time period online learning platforms such as the language learning app Duolingo have reached an all time high with Duolingo reaching over 500 million total downloads [1] or the Programming Learning App Mimo having reached a total of 30 million users [2]. Currently, the gamified and "addictive" learning app Duolingo beats many of its other non-gamified counterparts such as Babbel or Busuu roughly ten-fold in terms of monthly downloads [3], which shows the big potential that gamification has in today's digital market. Gamification has also already been applied and studied in various other contexts such as fitness trackers [4], to increase mathematics competence [5] or even to enhance virtual tourism [6].

However, despite gamification being widely adopted there are numerous cases reported in studies where gamification has been applied incorrectly and then leading to negative outcomes [7][8]. Example negative outcomes include that badges might increase performance avoidance learning goals [7] or negative influence of badges, leaderboards, and experience points on students with sensing learning tendencies [8]. This indicates a need for more precise and effective design strategies. Furthermore, the specific element sets to be investigated in this paper – namely social elements like following friends and celebrations and learning streaks and their corresponding notifications – have not been investigated closely yet [9]. Moreover, most case-studies on gamification apps only rely on qualitative survey experiments and do not collect usage data directly to investigate the effects of specific gamification features on user engagement and user retention [10]. Finally, most studies that evaluate gamification have not quantitatively explored the effects of specific categories of gamification-features. This lack of empirical data on

user interaction highlights a significant gap in understanding the true impact of specific gamification elements.

Therefore, this study tries to address this gap in knowledge by using a mixed methods approach [11]. Our mixed methods approach consists of gathering qualitative findings from a literature review, programming a gamified mobile learning application based on findings from our literature review and evaluating the gamified e-learning app (GeoChamp [1,2]) that we developed to test the effect of gamification features on engagement and retention. The goal of this thesis is to ascertain which specific gamification features – either social elements or learning streaks – have the greatest effect on enhancing user engagement and retention which are important factors for successful e-learning since many studies see engagement as a measure of student motivation [12]. User Engagement determines how much people engage with learning in any given moment and user retention describes how long a user can be retained to return consistently to engage with the application over a longer period of time.

At the time of writing GeoChamp currently has over 1000 mobile app downloads, 200 MAU (monthly active users), 50 WAU (weekly active users) and 20 DAU (daily active users). The app already sets itself apart from other apps by including very few gamification features already, with more to be released like daily streaks which belong to the performance category from the taxonomy of Toda [13] and from the same taxonomy social gamification features such as receiving and celebrating friends accomplishments will also be included. Therefore we ask the research question:

> *How do two sets of gamification features, from the established categories "performance" and "social", affect and compare in increasing user retention in learning apps?*

We will address this question by conducting usage data collection based on gamification directly from within the app, where the effects on user retention of two sets of features, namely social elements and learning streaks are compared to each other using an A/B testing framework. This is possible since the GeoChamp application was developed as part of this thesis and hence it is possible for us to extend its functionality.

Our findings have implications on understanding how gamification features can influence successful learning experiences and how the impact of different variations of implementations differs. Moreover, our findings can help e-learning platforms to make more informed decisions when considering to add gamification into their learning experience with concrete guidance for the feature design process of what to pay attention to and what anti-patterns to avoid.

[1] GeoChamp Play Store: https://play.google.com/store/apps/details?id=com.felixolszewski.geochamp1&hl=de_AT
[2] GeoChamp App Store: https://apps.apple.com/at/app/geochamp/id6477741110

CHAPTER 2

Background and Current State-of-the-Art

Learning through digital means is becoming an integral part of many peoples lives. Today, a large number of learners depend on online platforms such as chatbots, mobile apps, and educational websites to gain knowledge. However, research indicates that students may struggle with motivation in online learning environments [14] and might not achieve learning outcomes equivalent to those in traditional classroom settings [15].

Learning is the process where we acquire new understanding, knowledge, behaviors, skills, values, attitudes and preferences [16]. In online settings, users acquire knowledge by engaging with content, receiving feedback, and applying what they've learned through hands-on activities. Many theories have gained popularity in the scientific field of learning, the most popular being constructivist learning theory, behaviorist learning theory and self-determination theory.

Constructivist learning theory implies that the learner constructs his knowledge through experience and reflection [17]. The theory of behaviorism often also gets applied to the context of learning where it implies that learning new behaviour is primarily driven by external stimuli and reinforced through rewards and punishments [18] which is connected to engagement with an application since engaging with a system is one type of behaviour and the same goes for user retention, which is the behaviour of returning to interacting with the system consistently. It has often been shown in studies that when sticking to Self-Determination Theory (SDT) when constructing an online learning environment, that this can have many benefits for the learners motivation and engagement [19]. In SDT, the focus lies on intrinsic motivation and the fulfillment of the three fundamental needs of autonomy, competence and relatedness which when fulfilled will lead to an increased motivation in humans. Gamification elements are able to address each of those needs in their unique way [20].

Constructivism exists in both learning theory which is the science of how we learn and in epistemology which is the science about the nature of knowledge. The main focus of constructivism is that learning is seen as an active process and not as a passive one. We actively learn new things be engaging with material and this way we construct or extend our present constructed knowledge tree by adding further knowledge into this tree. On the one hand, this learning theory implies that when teaching we should put a strong focus on the learner and his existing constructing knowledge tree when designing a learning environment, to make sure the learning content can be built upon or possibly has already been built. On the other hand this sheds a more subjective light on learning material as it would need to be constructed adjusted to each learners previous knowledge since according to constructivist learning theory there is no knowledge independent of the meaning attributed to experience (constructed) by the learner, or community of learners [17].

Behaviorist learning theory [21] is when the theory of behaviorism applies to learning. Behaviorism emphasizes observable and measurable aspects of human behavior, focusing on how environmental stimuli and consequences shape learning outcomes [22]. The theory posits that all learning occurs through stimulus-response associations, where behaviors are either reinforced through rewards or discouraged through punishments. In educational settings, behaviorist approaches rely on clear rules, structured environments, and systematic reinforcement schedules to shape desired learning behaviors [23]. The theory operates on four key functions: automatic responses, escape behaviors, tangible rewards, and social attention. For behaviorists, learning is only considered to have occurred when there is an observable change in behavior, rather than through internal mental processes [24].

Self-determination theory (SDT) represents a comprehensive framework for understanding human motivation and personality development in learning contexts [19]. At its core, SDT identifies three fundamental psychological needs that drive intrinsic motivation: autonomy (the need for control over one's actions), competence (the desire to master challenges), and relatedness (the need to feel connected to others) [25]. When these basic needs are satisfied in learning environments, students demonstrate greater engagement, persistence, and academic achievement. Unlike behaviorism's focus on external rewards, SDT emphasizes the importance of intrinsic motivation, suggesting that learners have natural tendencies to explore, grow, and develop when their psychological needs are met. The theory recognizes that motivation exists on a spectrum from highly extrinsic to highly intrinsic, acknowledging that most learning situations involve a combination of both types of motivation [26].

Thus, motivation of people who learn in online environments has been seen to be impacted by the differences to a traditional learning environment [14]. Research indicates that online learners often face disengagement due to feelings of isolation, delayed feedback, or monotonous content [27]. Moreover, to solve and address the latter issues, gamification has been proposed as a practical solution to increase the motivation of the learner [28].

2.1 Gamification

2.1.1 Definition and Evolution of Gamification

Gamification, as defined by Charles Hofacker, involves the integration of elements commonly found in game design into non-gaming goods and services. The objective is to elevate the value perceived by customers and motivate behaviors that create further value [29].

In line with this definition, it is important to distinguish between Gamification and Game-Based Learning where in Game-Based Learning a full-fledged game serves as a medium to teach a concept while in gamification the non-gaming good or service gets enhanced by game design elements [30].

The first appearance of gamification was in 2003 when developer and designer Nick Pelling created a company to bring game elements onto hardware-based user interfaces like vending machines to increase businesses revenue [31]. While the company had to shut down due to too little interest, another company named Bunch Ball launched in 2005 which published its first gamification product in 2007 which allowed businesses to use game elements to get closer to their business goals. The term gamification was not yet used at the time but shortly after it became a common term and soon scientific papers about gamification were also starting to get published at the beginning of the next decade [32].

While gamification has its origins in business domains, nowadays Gamification is also very commonly applied in learning environments since it can help to excite and motivate learners to pursue learning the target content due to an improved environment [33].

2.1.2 Classification of Gamification Strategies

A popular classification of gamification is the classification developed in 2019 by Toda et al.[34]. This classification which aims to classify gamification elements into categories is tailored for educational contexts and was derived from existing literature on gamification. It consists of the five dimensions of Performance, Ecological, Social, Personal and Fictional elements where each category has two to five subcategories. It is depicted in figure 2.1.

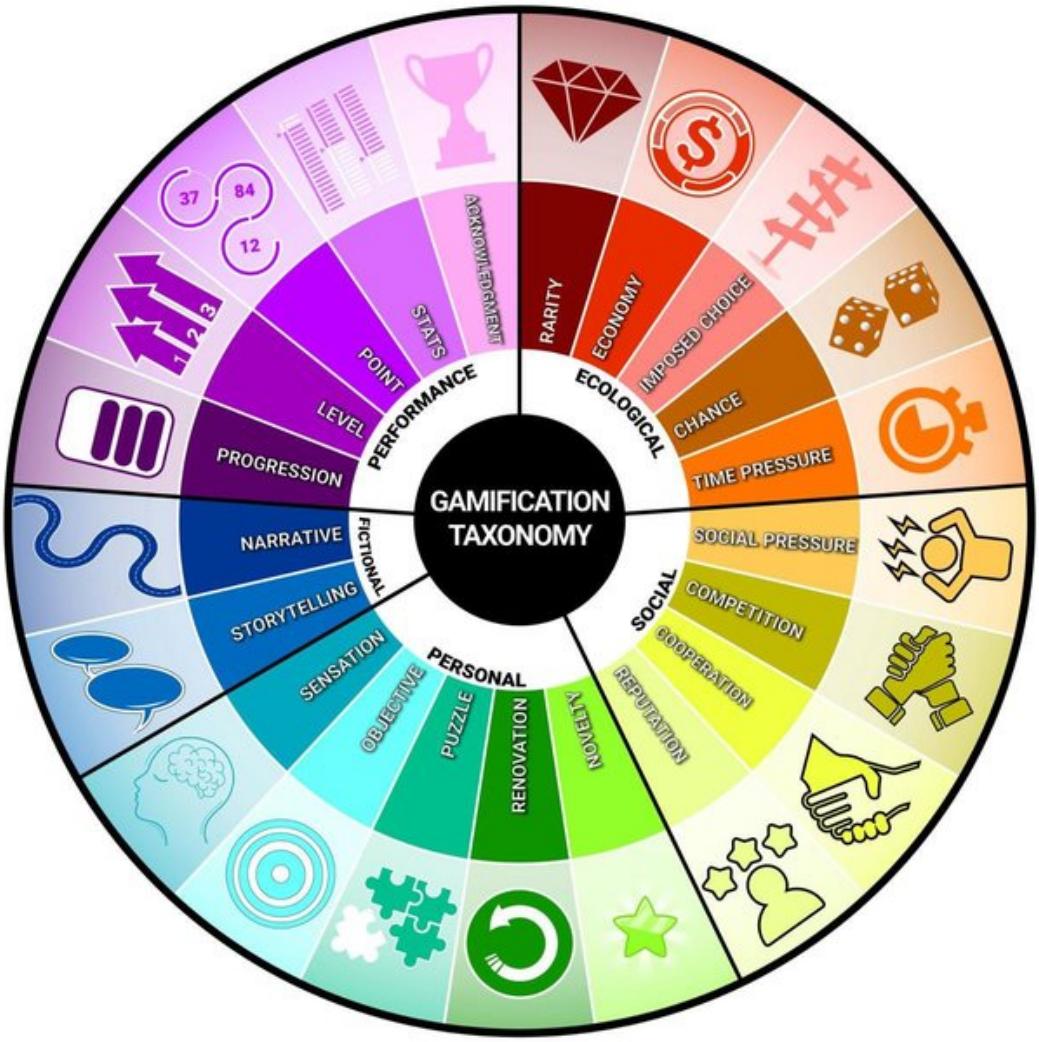

Figure 2.1: Gamification Taxonomy by Toda et al. [34]

Performance elements from the performance category of the classification of Toda et al, elements like Points, Progression, Levels, Stats, and Acknowledgement help students track their progress. Without these, learners may feel lost due to lack of feedback. Furthermore, Ecological elements (Chance, Imposed Choice, Economy, Rarity, Time Pressure) shape the gamification environment. Their absence makes the experience less engaging and interactive. Social elements enable learner interactions through Competition, Cooperation, Reputation, and Social Pressure. Missing these elements can lead to student isolation. Personal elements (Sensation, Objective, Puzzle, Novelty, Renovation) focus on individual learner engagement. Without these, students may lose motivation due to lack of personal meaning. Fiction combines user experience (Narrative) with environmental context (Storytelling). Its absence removes the contextual meaning and purpose behind

tasks in the immersive environment.

The Performance dimension encompasses the crucial elements that provide feedback to users within a gamified environment. This dimension consists of Point, Progression, Level, Stats, and Acknowledgement components. These elements work together to create a comprehensive feedback system that guides and motivates users through their journey. Points serve as the fundamental scoring mechanism, while Progression helps users visualize their advancement through progress bars or maps. Levels add a hierarchical structure that unlocks new content or challenges as users advance. Stats provide detailed information through dashboards or displays about user performance and achievements. Acknowledgements, such as badges, medals, and trophies, celebrate specific accomplishments and milestones, reinforcing positive behaviors and achievements.

The Ecological dimension relates to the environmental properties and mechanics that create dynamic interactions within the system. This dimension includes Chance, Imposed Choice, Economy, Rarity, and Time Pressure elements. These components work together to create an engaging and interactive environment that challenges users in various ways. Chance introduces unpredictability through random events or rewards, while Imposed Choice forces users to make meaningful decisions that impact their progress. The Economy system enables transactions and exchanges within the environment, creating value and scarcity. Rarity adds exclusivity through limited resources or special items, making certain achievements more desirable. Time Pressure introduces urgency through deadlines or countdowns, although it must be implemented carefully to avoid causing stress or disengagement.

The Personal dimension addresses individual user experiences and motivations, incorporating Sensation, Objective, Puzzle, Novelty, and Renovation elements. This dimension focuses on creating meaningful and engaging experiences that resonate with individual users. Sensation elements engage users through visual and auditory stimulation, while Objectives provide clear goals and purpose. Puzzles challenge users cognitively and test their understanding of content. Novelty keeps the experience fresh through updates and new content, preventing stagnation and maintaining user interest over time. Renovation provides second chances and opportunities for improvement, allowing users to learn from mistakes and continue progressing despite initial failures.

The Fictional dimension weaves together the narrative elements that create context and meaning within the gamified environment, consisting of Narrative and Storytelling components. This dimension bridges the gap between user actions and the broader context of their experience. The Narrative element shapes how events unfold based on user choices and actions, creating a personalized journey through the system. Storytelling provides the framework for how this narrative is communicated to users, whether through text, audio, or other sensory means. Together, these elements create an immersive experience that gives meaning to user actions and helps maintain engagement by providing context and purpose to various tasks and achievements within the system.

The Social dimension focuses on the interpersonal aspects and user interactions within

the gamified environment, comprising Competition, Cooperation, Reputation, and Social Pressure elements. This dimension creates a community-driven experience where users can engage with one another in meaningful ways. Competition drives engagement through leaderboards and challenges, while Cooperation encourages teamwork and collaborative problem-solving. Reputation systems create social hierarchies and status indicators that reflect user contributions and achievements within the community. Social Pressure, while potentially motivating, needs to be carefully balanced to maintain a positive and supportive environment rather than creating unnecessary stress or anxiety among users. The Social dimension is the main focus of this thesis as this thesis tries to analyze the impact of social features. For this reason and due to its diversity of features and potential outcomes we view this dimensions as the most important dimension from the taxonomy of Toda et al.

The main gamification elements that this paper discusses are elements being able to follow friends, celebrating their achievements and also learning streaks and their corresponding notifications. The follow-feature and achievement celebrations fall into the category of social elements, specifically achievement celebrations could be attributed to the reputation category. Learning streaks are a reflection of a users performance of consistenly learning and are hence attributable to the performance category of "Level" but since other users can view each others learning streak, the streak also falls into the social category of "Reputation". Streak notification reminders also have a more nuanced categorization where on the one hand they and the streak together can be seen as an ecological element of "Time Pressure" but the textual messages contained within notifications in GeoChamp also contain the fictional dimension with "Storytelling" since they convey the narrative that the mascot Geo does not want the user to loose their streak or else that will make Geo sad or angry.

There also exist other less fine-grained gamification classifications such as the one proposed by Xi and Hamari which splits gamification elements into three kinds namely immersion-related features, achievement-related features, and social interaction-related features [35]. It can be seen as a generalization of the classification proposed by Toda et al. since each category encompasses one or more dimensions of the Toda classification.

Xu et al. classified gamification elements in a different way, into two categories where elements such as points, badges, feedback and challenges fall into the extrinsic category and leaderboards, levels, avatars, and privacy control are seen as intrinsic gamification elements [36].

The distinction is made based on self-determination theory which draws upon the three human fundamental psychological needs, namely competence, autonomy and relatedness which are all responsible for intrinsic motivation.

Research suggests that while rewards like points badges and feedback can boost short-term activity, it is intrinsic motivation that fosters long-term engagement and enjoyment. Hamari finds that extrinsic motivation, like rewards, does not lead to an increase in playing activity[35]. Researchers recommend that both intrinsic and extrinsic motivations

should be integrated to develop a meaningful gamification experience [37].

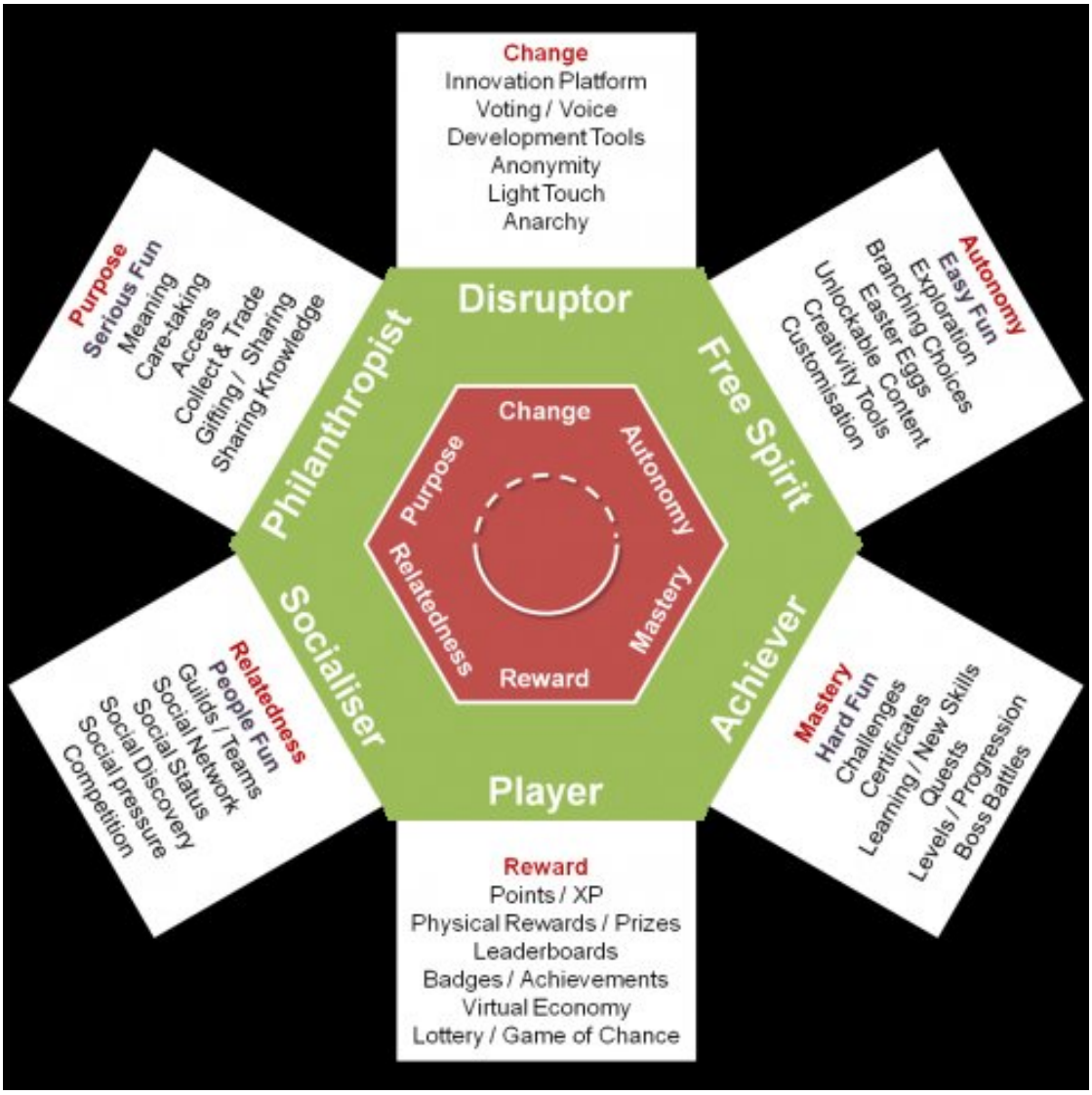

Figure 2.2: HEXAD Player Type Framework [38]

Another common classification in the domain of gamification is the HEXAD scale which can be seen in Figure 2.2 does not categorize game elements but instead splits users into six different groups based on their personality type and their core motivations[38]. These groups include the Philantropist, Socialiser, Free Spirit, Achiever, Player and Disruptor. Each user type is driven by his values. In the middle of the diagram a circle splits those six values in half, differentiating between intrinsic and extrinsic motivations. Purpose, Change and Autonomy fall into the intrinsic category while Relatedness, Reward and Mastery are seen as extrinsic motivations.

A more recent development of gamification classification has been the approach proposed by Schöbel et al [39]. Their work integrates and refines existing frameworks to offer a comprehensive set of 15 "construction elements," each categorized along 10 dimensions that distinguish foundational game logic (such as interdependency or development) from optional design choices (like rewards or competition). By systematically mapping these elements and dimensions, their taxonomy enables both the analysis of existing gamification designs and even the creation of new, context-sensitive gamification concepts [39]. The taxonomy applied in a concrete use case can be seen in Figure 2.3 with the use case of the Nike+ app.

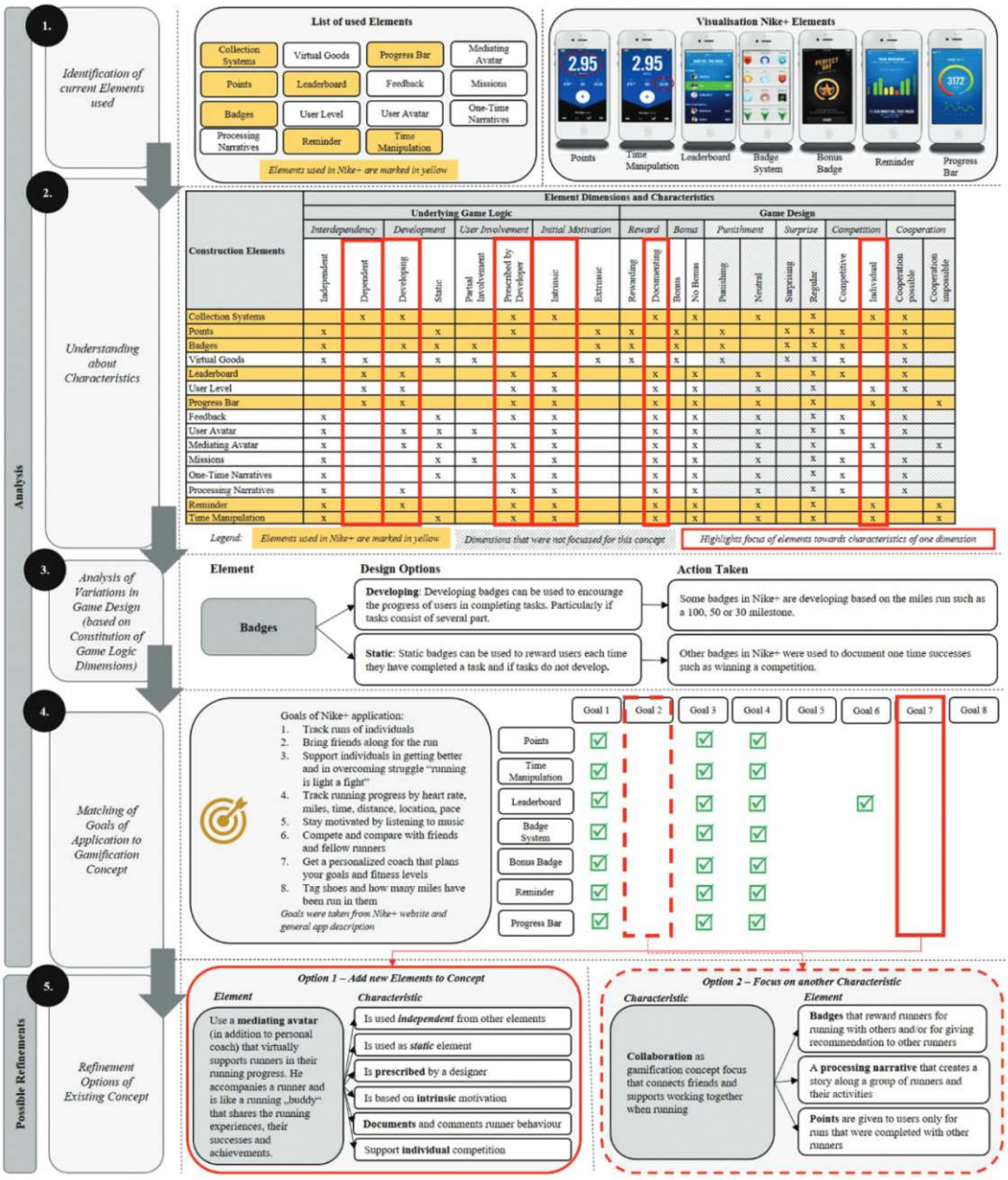

Figure 2.3: Gamification Framework developed by Schöbel et al

2.1.3 Game Based Learning and Video Games

To understand how exactly gamification manifests itself and how it distinguishes from other applied techniques, we need to look closer at Game-Based Learning and Video Games. Game-Based Learning is a technique that is defined separately from gamification where the main difference between gamification and Game-Based Learning is that in Game-Based Learning an existing application does not just get enhanced by adding gamification elements to make the existing application still fulfill its primary purpose but instead in Game-Based Learning there is no existing software application but instead a full-fledged game gets taken as the main software entity and this game differs from other games in that it not also entertains but also contains content that teaches the user a specific concept. The concept of Video Games itself is even more different from gamification and can be categorized closer to Game-Based Learning, the only difference to Game-Based Learning being that the purpose of creating a Video Game was not to teach a concept but to entertain and hence a Video Game does not contain any artificially added learning content.

Some famous examples of Game-Based Learning include the computer game "Minecraft: Education Edition", the mobile app for kids called "Toca Boca World", and we even see the learning application LingoKids being an example of Game-Based Learning. With Minecraft being the best-selling video game at the time of writing and its educational edition just being a special version of Minecraft with added Educational Challenges [40] this makes us see this as a perfect example to explain Game-Based learning. A more special case is the highly trending mobile app "Toca Boca World" which is also primarily a game but it did not start out or emerge from a pure video game but instead even since its very beginnings had the aim to be a teaching application combining gaming and learning [41]. Those two examples show us that applications of Game-Based Learning definitely can but do not have to start out or be based on a non-learning video game.

Studies like Al-Azawi [42] have discussed the concrete differences between Game-Based Learning and educational gamification. The study of Al-Azawi even differentiates Game-Based Learning even further, distinguishing between Game-Based Learning (GBL), educational games and gamified education, where the distinction between GBL and education games is made in a way that educational games are one specific kind of GBL, but using the classification of the authors, not just the educational game "Minecraft: Educational Edition" can be seen as GBL but even the "vanilla" version of Minecraft includes Game-Based Learning implicitly since the user learns new concepts through playing [42].

2.1.4 The application of gamification outside of the educational context

While today gamification finds the most application in education, this is certainly not the only domain that gamification is being applied in today [43]. In fact gamification did not even have its origins in education but instead in business to increase purchases with Bunch

Ball providing a software as a service to integrate gamification into existing applications [44]. Domains besides education where gamification is also applied are in digital business, at the workplace, in tourism [37], e-commerce [45], health, exercise, content sharing, sustainable consumption, work/crowd-sourcing, data gathering, industrial work processes [46] and even Software Engineering [47] [48].

Gamification in digital business might be applied in a similar manner as the Bunch Ball framework which aims to increase the revenue of a business by providing a gamified engaging experience to its users. This makes users more likely to engage deeply with the platform and in turn increases the probability of users spending money. One example of digital business is e-commerce where in a gamified e-commerce platform users could for example earn a certain number of points, level or achievements after purchasing a certain number of items [49].

In industrial work processes gamification might be applied in a different manner than in e-commerce, due to the different circumstances at hand. In industrial work processes, often manual steps have to be performed many times and applied gamification has the potential to motivate the worker to perform those tasks faster, or with higher quality or for a longer period of time. Those circumstances are very different from e-commerce as are its users and the goals of its users [46].

While neither e-commerce nor industrial work processes involve frequent and strong interaction between many human actors within those processes, at the workplace this is often the case. This is why at the workplace, gamification might aim to use this network of interaction by including social gamification elements into the gamification flow to foster cooperation or even friendly competition between coworkers [50].

Another domain where gamification is applied which also involves social networks is in online forums with the example of the Software Engineering Stack-Exchange StackOverflow where studies claim that gamification has led to immense success [48]. StackOverflow is a prime example of successful application of gamification as its context proved great potential, where the problem of e-commerce where the user already has a different metric that he optimizes for when engaging, that is the value-money equation, was not yet present in StackOverflow. In StackOverflow users interact in two ways mainly, on the one side users can be "producers" by creating and answering questions and on the other hand users can be "consumers" by reading questions when a programmer faces a problem with a certain technology [48].

For consumers of StackOverflow, the main way that they interact with the gamified system is by reading the number of points of questions and answers visible on the left hand side on Figure 2.4 and by reading the reputation of the user that is answering, which can be seen in the bottom right corner on Figure 2.4. But it is in fact way more the producers on StackOverflow that are influenced by gamification, since producers value their own reputation points and medals greatly as a way of reflecting their knowledge and contribution status within the programming community and producers are motivated to take their time to create well-written questions if those questions were not posed

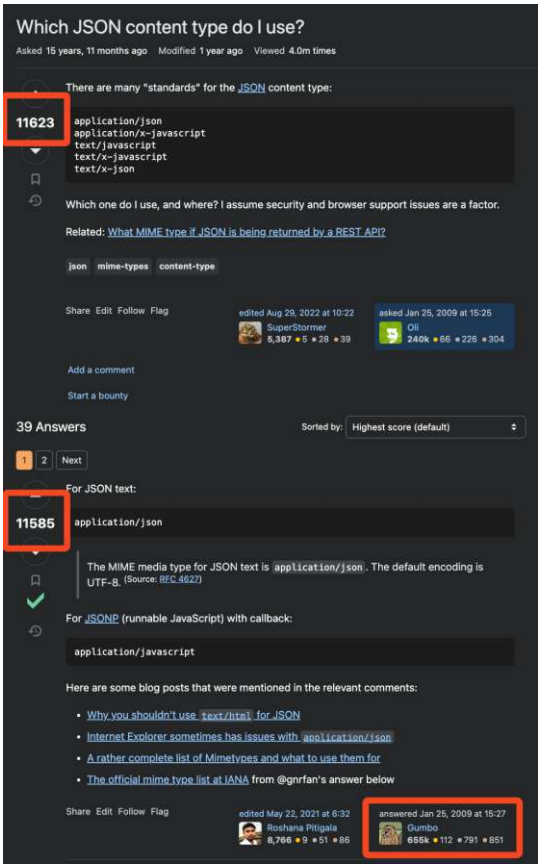

Figure 2.4: A StackOverflow post

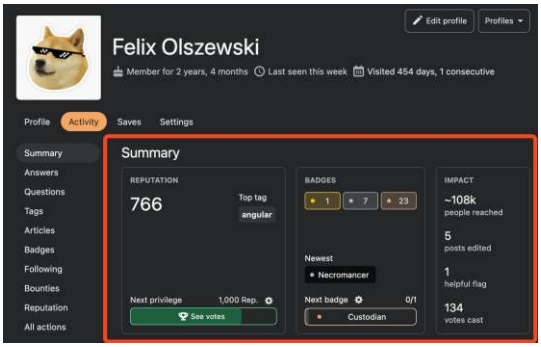

Figure 2.5: The StackOverflow Profile Page

yet or to be the first person to answer unanswered difficult questions, since this is the most effective way to gain points and medals on StackOverflow [48]. Points, medals and progress bars for a producer profile can be seen in Figure 2.5 in the section highlighted in red. The key difference between gamification in StackOverflow and gamification in e-commerce is that the element of "points" that StackOverflow added created a new value equation and reputation within an existing system in which prior to the addition, producers did not interact with the forum motivated by a visible and measurable value equation.

An application of gamification similar to its application in the forum StackOverflow is the application of gamification into work/crowd-sourcing since the answers on StackOverflow themselves could be seen as work that was sourced from a crowd. Another famous example of gamification in crowdsourcing is present in the app Foldit which gamified the manual task of folding proteins such that the online community received gamified rewards when contributing to the folding database [51].

Further examples of gamification are the application of gamification to data gathering where users are motivated by a game-like user interface to gather more data, which is

very similar to a crowdsourcing approach of applying gamification, since data is also gathered by users who are not deeply involved with the system on a daily [48].

Sustainability and particularly sustainable consumption of consumers is also an area where gamification is applied more frequently today. Being and acting sustainably, especially when consuming is often a task where it is hard to get any concrete numbers on how much impact a users actions have. This can be solved by measuring it numerically and reflecting numbers back to the user, which in turn provides an opportunity to gamify the display of those numbers [52]. Famous examples of gamification in sustainability include the app "Cool Choices" which is a game where players compete in teams to reduce energy usage over a multi-week period [52].

When applying gamification to health the goal is to motivate the user to improve their health by giving them rewards upon completing health-fostering actions. One way to stay healthy is by exercising, which is why exercise apps also frequently use elements of gamification, with the famous exercise app Strava or Fitbit having been the subject of multiple gamification studies [4] [53]. Exercise apps also provide the ability to view the accounts of other users which is why social gamification elements are often implemented in exercise applications [4].

While gamification has reached lots of application and thus scientific discussion, the results of whether and how gamification can be effective are still not crystal clear as there have been multiple concerns raised lately [54]. A literature review of empirical gamification studies concluded that most studies show that gamification provides positive effects but those effects depend on the context in which gamification is applied [45].

The review from Hamari et al [45] also stated that the result of gamification depends greatly on the context in which the gamification is being implemented as well as on the users using it. In contexts such as education and learning the review from Hamari concluded an overall positive effect. However, the results non-educational contexts are more nuanced.

In a utilitarian context such as in e-commerce sites gamification might not be successful due to the nature of the context being rationally-driven where the user would optimize for economic exchanges [45].

2.1.5 Evaluating the effectiveness of gamification outside of the educational context

As Hamari et al stated, the result of gamification depends greatly on the context in which the gamification is being implemented as well as on the users using it [45]. We try to look at each of the contexts in which gamification is applied in today's world one by one to find out how scientific literature evaluates the effectiveness of gamification in the specific context. Our observed contexts are in e-commerce, at the workplace, in tourism, in exercise, sustainable consumption, industrial work processes and in crowdsourcing.

According to Garcia et al gamification in e-commerce has a positive effect on engaging users [55]. The study studied the behavior of 253 Spanish Amazon users and evaluated the PBL triad as a system of reputation points that forms a consistent construct. Nevertheless, according to Garcia et al the research of gamification in e-commerce is still sparse and further research is required [55].

At the workplace gamification has been investigated under strong scrutiny with the most recent research highlighting negative effects of gamification on frontline employees (FLEs) in retailing and telemarketing [56]. The negative impacts manifest in worsened employee engagement and well-being, moderated by the willingness of such employees to participate in such gamified work.

In 2023 a meta review of the application of gamification in tourism has concluded positive effects of gamification in tourism where gamification driving engagement and experience and thereby driving behavioral change [57].

Many health and exercise apps today implement gamification to boost their products outcome. A recent meta-review of the effectiveness of gamification for physical activity applications has concluded that gamification has the definitive potential of having a positive effect for creating desired behavior [58], its effectiveness being especially present in the short term.

When looking at gamifications effectiveness in sustainable consumption, a recent study found that gamification actually shows the opposite of the desired impact after being applied into an existing applications context [59]. Possible explanations were that moral licensing took place which made users not feel bad about performing environmentally unfriendly acts due to them having established green points inside the app or that the external rewards undermined intrinsic motivation for environmental protection.

In 2022 researchers aimed to find out whether gamification could have a positive effect in the industrial work process using an experimental approach. The researchers found that turning work tasks into game-like experiences produces measurable efficiency gains, creates an overall positive emotional atmosphere, and enhances workers' motivational levels [46].

Finally, the case of StackOverflow is a case where researchers were able to measure a positive effect of gamification to motivate users to contribute to crowdsourcing frequently and with high quality contributions [60]. According to the authors there are multiple other different motivations at play that need to be considered such as the extrinsic motivation of a StackOverflow Profile serving as a possible career prospect or intrinsic motivations such as self-improvement or helping others [60].

2.1.6 Benefits of Gamification in Learning

The use of gamification promises to bring many benefits including increasing user engagement and user retention and the user being more motivated to perform the intended actions in the applied context [61]. However in educative contexts also other benefits

such as improved collaboration skills, problem-solving and communication skills have been shown. It is also said to increase the learners engagement in online forums, projects and other learning activities [62].

2.1.7 Common Negative Effects and Ethical Considerations of Gamification

While literature on gamification design, categorization and implementation presents itself as a clear topic, the concrete benefits of gamification have been questioned increasingly in recent years [54]. Despite research on the negative effect of gamification still being sparse [63], critics mention many biases and effects that led to gamification gaining significance in scientific literature and in the industry [64] but there are also mentions of pitfalls in gamification and ethical concerns raised about it [54].

According to the mapping study done by Almeida et al, commonly negative investigated effects of gamification are lack of effect, worsened performance, motivational issues, lack of understanding, and irrelevance [63]. Gaming the system and cheating were also reported by Almeida as ethically negative effects of gamification [63].

Dah et al concluded in 2024 that there are four factors why gamification could be failing, namely shallow gamification, the overjustification effect, an overly strong focus on the Points, Badges, Leaderboards triad and an overreliance on narrow scientific models and theories to explain or design gamified experiences [54].

Shallow Gamification is the superficial application of game elements on a system without transforming the core underlying experience or without deeply applying game-design principles [54]. The overjustification effect refers to the effect of too much extrinsic motivation reducing intrinsic motivation, discovered already more than 30 years ago by Lepper et al. and Ryan [65] [66]. An overly strong focus on points, badges and leaderboards is also said to reduce the motivation in learners, closely tied to the shallow gamification principle. Finally, focusing only on models like SDT and Flow Theory to evaluate and conceptualize gamification is another reason why gamification is said to still be failing [54].

2.1.8 Review of Psychological Impacts and User Behavior Change

Landers sees Gamification as not a "new" instructional technique per se but instead as solely a new combination of existing motivational techniques [67]. Further, the techniques are based on a multitude of theories including gamified instructional design, classic conditioning theories of learning, expectancy-based theories, goal-setting theory, and SDT (Self Determination Theory) [67].

Classical conditioning theories explain how learning occurs through association between stimuli and responses [68]. In gamified learning, this manifests when inherently neutral elements (like points or badges) become associated with positive learning outcomes, eventually triggering desired learning behaviors independently. The process involves

key principles like acquisition, where associations are established and strengthened, and extinction, where responses decrease when reinforcement stops [68]. This behavioral foundation helps explain how gamification elements can shape learner behavior over time.

For instance, goal-setting theory suggests the effectiveness of setting specific, difficult goals and the generality of its effects across people, tasks, countries, time spans, experimental designs and goal sources [69].

In addition, Self Determination Theory (SDT) draws upon the three fundamental psychological needs of humans namely autonomy, competence and relatedness [25]. It states that those three innate needs drive human behavior and choices [25]. SDT also distinguishes between intrinsic and extrinsic motivation, where we are intrinsically motivated when we are performing an action that is inherently interesting and satisfying and extrinsically, if that action provides an external reward or punishment [25].

2.1.9 Successes and Criticisms of Gamified Learning Platforms

Today's most famous example of gamification and its success potential is the language learning app Duolingo [34]. Duolingo made engaging in learning a second language deeply attractive for lots of people worldwide, having reached a total of 800 million downloads, 75 million monthly active users and 20 million daily active users [70].

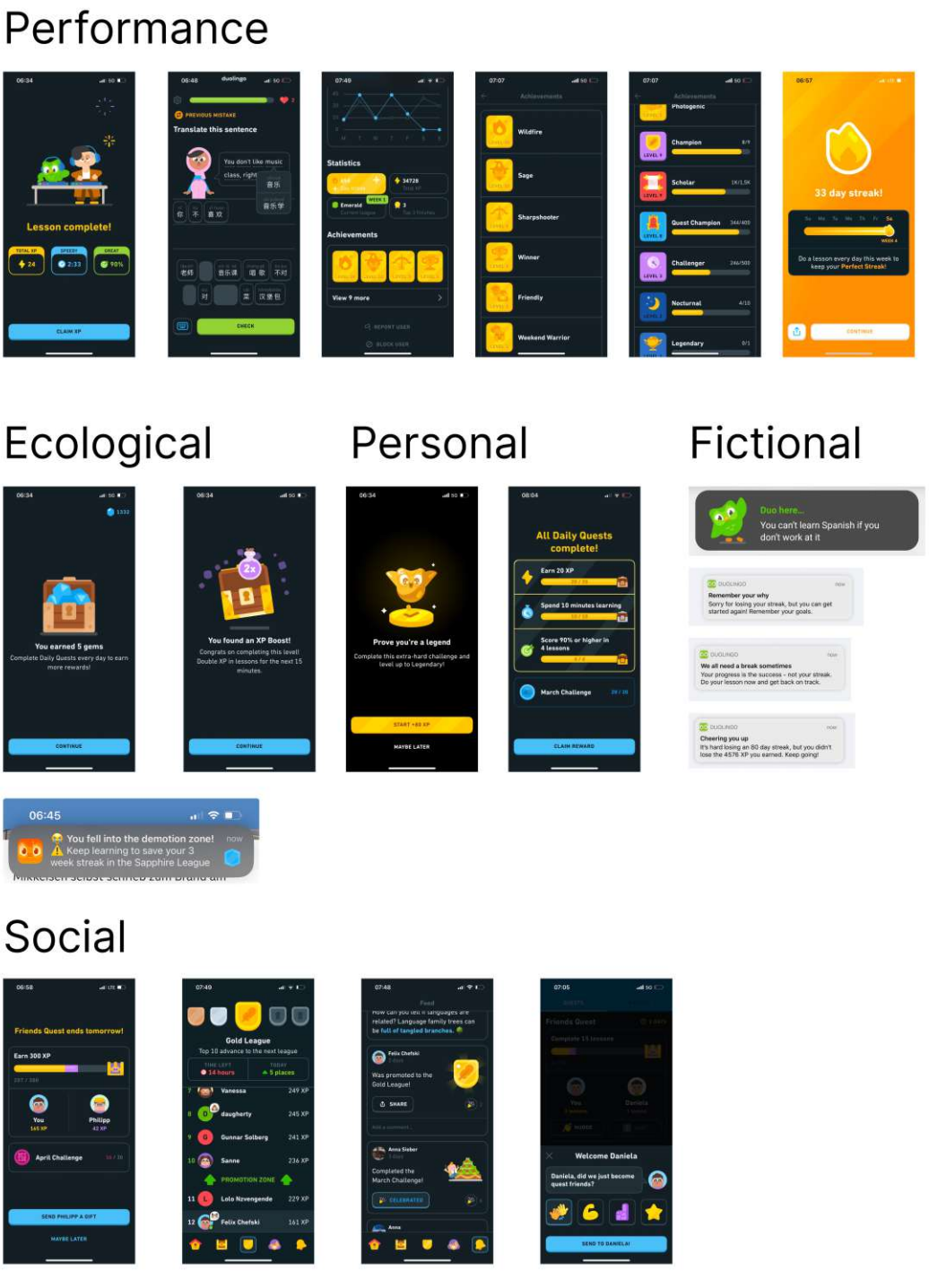

Figure 2.6: Gamification Elements in Duolingo (2024)

As can be seen in Figure 2.6 the gamification approach of Duolingo is very involved. Duolingo also features a leaderboard and a league system as one of its gamification elements, all of which make us classify Duolingo as using a "deep" gamification set as opposed to a "shallow" gamification set, by the definition of Lieberoth [71] where deep gamification means an app really uses full game design principles to design the app, not just visual framing elements[71]. This deep gamification approach could attribute some of the big successes that Duolingo had, related to gamification.

However, recently learning apps have also been criticized a lot in public discussion [72] and also in scientific discussion [73] for misusing gamification to attract user engagement while deteriorating its actual learning services. In 2022 Hadi et al coined this phenomenon with the term "gamification misuse", which happens when the gamification in a learning environment becomes so strong that it starts negatively impacting the learning performance [73]. The Google Search Query "duolingo is bad" and similar queries also only reached significant numbers in the last two years, coinciding with the publication of the mentioned negative paper, which shows a correlation, possibly the causality of a trend that gamification is being viewed more critically by the public additionally to just the scientific community [74].

Prior to the arise of the studied concept of "gamification misuse" studies focused solely on measuring whether gamification actually created behavioral engagement, which separate from measuring the effect of gamification on learning performance. Gamification misuse falls into three categories namely user learning, well-being, and ethics [73].

Gamification misuse is actually described by Hadi et al as an action that the user itself takes and not an action of how the application itself misuses gamification [73]. This misuse can manifest itself in three main ways namely in Reduced Learning, Poor Well-Being and Threatening Ethics.

The misuse of gamification in learning apps can significantly impair users' learning outcomes. One common consequence is a loss of confidence in learning, as users who overly rely on gamified features such as streaks or leaderboards may feel inadequate when they fail to meet the app's expectations, undermining their belief in their learning capabilities. Additionally, reduced interest in learning occurs when users prioritize collecting rewards over acquiring language skills, leading to frustration and decreased intrinsic motivation. This disengagement can ultimately result in users withdrawing from learning or dropping out entirely. The pressure to maintain gamified achievements, such as streaks, can lead learners to abandon the platform when those goals are not met, depriving them of opportunities for meaningful language acquisition [73].

Gamification misuse also poses ethical concerns, particularly in creating an unfair learning environment. For instance, users who exploit gamification elements like leaderboards through unethical practices, such as cheating, can discourage honest learners by distorting the competition. Furthermore, such misuse risks encouraging learning misconduct, especially among younger or impressionable users. When cheating or manipulating the system becomes normalized within the learning community, it sets a dangerous precedent,

potentially fostering a culture where rewards take precedence over genuine learning. This undermines the educational integrity of the platform and jeopardizes its role in promoting honest, effective learning practices [73].

The negative impact of gamification misuse extends beyond learning to users' overall well-being. Psychological well-being suffers as users experience disappointment, apprehension, or even self-recrimination when their obsessive pursuit of gamified rewards fails to yield meaningful progress or satisfaction. This can lead to stress, anxiety, and a diminished sense of accomplishment. Additionally, physical well-being is at risk due to prolonged, compulsive use of the app. Users have reported experiencing physical strain, disrupted sleep patterns, and neglected daily responsibilities due to their fixation on maintaining streaks or climbing leaderboards. These effects highlight the potential for gamification to detract from users' mental and physical health, rather than supporting a balanced and enjoyable learning experience [73].

2.2 User Engagement and User Retention

Increasing User Engagement and User Retention is often one of the key objectives when developing a digital product. Lalmas et al. defined user engagement as follows:

> User engagement refers to the quality of the user experience that emphasizes the positive aspects of interacting with an online application and, in particular, the desire to use that application longer and repeatedly. [75]

User retention is often seen as a part of user engagement but it places a distinct focus on retaining the engagement of users over a period of time, with the user coming back multiple times to reengage with the product.

2.2.1 Measurements

User Engagement, including user retention, can be measured in multiple different ways with the three popular ways of measuring being self-reporting, physiological approaches and web analytics [75]. Self reporting means that users reported themselves on their engagement in questionnaires, surveys and interviews. In physiological approaches the reports are not verbally produced by the user but are instead measured from physiological aspects of the user such as facial expression analysis or speech analysis. Lastly, web analytics are a straightforward approach to measure user engagement since they include many different numeric values such as clickthrough rates, number of page views and time spent on a site for example.

There are also a number of measurement dimensions for user engagement including scale, setting, temporality, objectivity, and the unit of measurement. All of those measurements can be used to measure user engagement and they each highlight a different aspect of user engagement.

One of the most common ways of measuring user engagement in scientific literature is the User Engagement Scale (UES) questionnaire which falls into the self-reported measuring category [76]. It consists of 31 items measuring six dimensions of engagement namely aesthetic appeal, focused attention, novelty, perceived usability, felt involvement, and endurability. A recent development has been the User Engagement Scale Short Form (SF) which is a shorter version of the UES questionnaire consisting of only 12 items [77].

2.2.2 Pitfalls when measuring user engagement

Commonly reported pitfalls when measuring user engagement are a bias toward heavily engaged users, focusing too much on acquisition rates, measuring the wrong metrics and neglecting other dimensions while only focusing on web-analytical data [75].

It is important to note that when relying on event-level metrics, heavily engaged users will be over-represented and this might skew performance metrics and it is wise to not see acquisition rates alone as an indicator of engagement with a product since a high churn rate might indicate that users are actually not sticking with the use of the application [78]. When also only measuring purely quantitatively with web-analytics then some underlying dynamics are not being revealed and it might not accurately reflect the users engagement [79].

2.3 A lack of quantitative research

Currently, there is very little quantitative evidence on the success of gamification. Learning however, is a process that happens over a longer period of time to be successful [80]. This long process can be best understood by employing a quantitative comparison of gamification features being included or excluded from an application, to properly measure their effects over the longer time period.

However, the current state of scientific knowledge on gamification focuses heavily on collecting only qualitative data to determine whether gamification was successful, a common form of collection being self-report questionnaires or interviews [81], which evaluate the effect of gamification only superficially during a moment in time.

As a remedy we should evaluate the effect of gamification on the learning process more quantitatively over a longer period of time. One common way to quantitatively measure the effect of certain features on a software application is the widely used A/B testing method. According to the research from Luo and Zhanni [81] the A/B testing method was not at all used a lot yet in the field of gamification research [81].

2.4 A/B Testing

A/B testing is defined by Quin et al. as follows [82]:

> A/B testing, also referred to as online controlled experimentation or continuous experimentation, is a form of hypothesis testing where two variants of a piece of software are compared in the field from an end user's point of view.

Hence A/B testing delves deeper than standard statistical hypothesis testing by typically applying to software and most of the time focusing on user metrics for comparison. The A & B in the naming stands for comparing variant A to variant B of a software [83].

In an A/B test the users are split into at least two groups where one group is the treatment group which receives a new changed version of the software product and the second group is the control group which receives the current version of the product. Then, for a certain duration the users engage with the software and analytics are being collected after which the metrics are compared to determine the effect that the new software version had on its users in comparison to the old product [82].

A/B tests are popularly used by software product companies with very large user numbers and very high revenue such as Google, Facebook, Amazon, Microsoft. They perform thousands of small A/B tests every year, where even small positive changes can account for enormous revenue increases. But it is also small companies like startups that use A/B tests to their advantage just that the experiments they perform are more drastic and they hence expect bigger effects of experiments [83].

2.4.1 A/B test design process

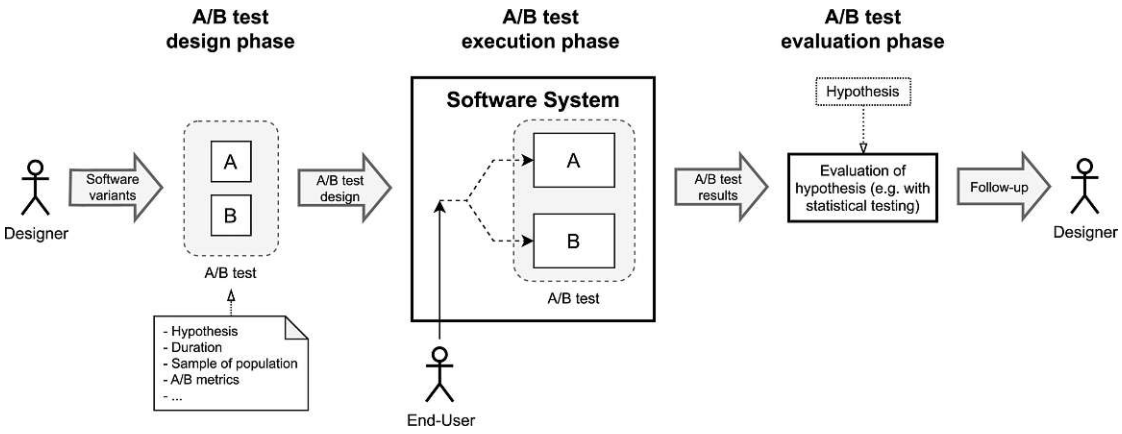

Figure 2.7: General A/B testing process from Quin et al. [82]

As can be seen in figure 2.7, the A/B testing process starts with the design of the A/B test, where the designer chooses two or more different software variants A & B and also comes up with other details of the test such as the duration, the specific sample from the population and metrics. The Hypothesis which is to be proven is also formulated in this step.

Engagement metrics are the most commonly used type of metrics in these tests, with other significant metrics including click metrics, monetary metrics, and performance metrics. Hypothesis testing for equality stands out as the most frequently utilized statistical method in A/B testing [82].

Then in the next phase the test is introduced to the existing software system and all end users are assigned to a variant and interact with the software until the test ends. After this phase, the A/B test results are handed over to the evaluation phase where statistical methods are used to check whether the hypothesis holds. The results from the evaluation are then forwarded to the designer who initially conceived the A/B test [82].

2.4.2 Statistical theory behind A/B testing

The underlying concept of almost all A/B tests is a two-sample t-test [84]. Such two-sample t-tests compare the difference of two distributions means relative to the variance [85]. The goal is to find out whether the difference in means between the two groups is greater than what could be expected by random chance. The smaller the p-value in such tests, the stronger of an evidence that the treatment is different from the control [83].

Two types of errors can occur when performing a statistical test: type I errors and type II errors. Type I errors, also known as false-positives, conclude that there is a significant difference between Treatment and Control when there is in fact no difference. Type II errors are false-negatives which conclude that the change had no effect when in fact it did. Type I errors are controlled at 5% by accepting statistical significance if and only if the p-value is smaller than 0.05 [86].

2.4.3 The subject of A/B testing

A/B testing primarily focuses on four key areas: algorithms, visual elements, workflows and processes, and back-end features. These tests are widely used across various domains, including web applications, search engines, e-commerce platforms, interaction software, and financial services. Notably, algorithms are a consistent focus of A/B testing in all these fields [83].

Moreover, back-end features such as server performance are especially important in the realm of search engines, making them frequent targets for A/B testing in this domain. This highlights the diverse application and importance of A/B testing in enhancing various aspects of technology and service delivery across different sectors [87].

2.5 Hypotheses and Research Assumptions

Finally, based on the above literature, we derive the following two hypotheses that we aim to investigate with our A/B testing approach:

- *Users in the "Social Gamification Elements" group (Variant B) will show higher user retention compared to the control group (Variant A).*
- *Users in the "Learning Streak and Notifications" group (Variant C) will show higher user retention compared to the control group (Variant A).*

CHAPTER 3

Methods

To propose practical strategies for integrating gamification into learning apps, we must cultivate a comprehensive understanding of the theoretical foundations of gamification. Thus, we adopt a pragmatic epistemological stance and employ a developmental mixed-method approach to synthesize insights from our research [88].

In particular, we use both a systematic literature review and a controlled hypothesis testing experiment, which are qualitative and quantitative methods because they provide complementary perspectives: the literature review synthesizes existing knowledge to identify trends, gaps, and theoretical foundations, while the controlled experiment allows us to empirically test specific hypotheses and establish causal relationships. We then present an integrative view of our findings by developing a consensus between the qualitative and quantitative research findings [89]. This allows us to develop inferences that go beyond each study's findings which helps us to gain a comprehensive understanding of gamification in learning apps.

3.1 Pragmatism, Research Strategy

To enhance our ability to understand gamification and provide future recommendations, we adopt a pragmatic epistemology by examining and testing the outcomes of strategies aimed at promoting a gamified learning experience. We prioritize the practical implications of our findings, particularly in identifying actionable suggestions for implementing gamification, as the most crucial aspect of our research.

We adopt a sequential mixed methods approach [90], guided by general principles for conducting mixed methods research [88], to develop a comprehensive understanding of gamification's role in learning applications, an area where existing research remains fragmented and inconclusive [64]. Grounded in our research questions, motivation, and context, this approach enables us to draw meta-inferences with practical significance,

leading to the development of gamification feature guidelines that go beyond surface-level elements such as Points, Badges, and Leaderboards.

3.2 Research Methods

First, we conducted a qualitative systematic literature review to guide the development of our hypotheses [88]. Also stemming from our literature review we developed gamification into the learning app GeoChamp. Based on the insights derived from the qualitative literature analysis, we then performed quantitative experiments using the GeoChamp application to test our formulated hypotheses. Figure 2.7 provides a summary of the methodology used in this study.

3.3 Implementation of GeoChamp Prior to Gamification

The development process of the GeoChamp application can be logically split up into two parts: version 1, which is the version prior to adding many gamification features to GeoChamp and version 2, which is the the version which was created following the literature review on gamification. In the following we would like to describe into detail the technical aspects of GeoChamp and what feature set version 1 of the GeoChamp app entailed.

3.3.1 The underlying technologies of the GeoChamp application

GeoChamp was developed using WebView technologies also known as Hybrid-App technologies which are different from true native mobile applications. While it can be considered a native mobile app, what makes GeoChamp differ is that the Native App renders a Local Website. The only part of the app that uses true native code is the widget implementation on both iOS and Android which uses the programming languages and approaches of Swift-UI and Java-XML respectively.

Furthermore, GeoChamp was developed with an offline first approach where most of the business logic runs on the client which makes it very suitable to be used in bad network conditions as well, which might happen in the classroom or if learning while commuting. The framework we chose for the WebView was Capacitor since it supersedes its error-prone predecessor Cordova. The Capacitor framework also provided the interface between our WebView and the native SQLite database by offering a library, where the native SQLite databases is where all of the entity data from GeoChamp is stored. Finally, we also hosted the GeoChamp Client as a website on Cloudflare, available under the domain of https://www.geochamp.app.

For the user interface components we used the Ionic Framework which works seamlessly with Capacitor and for our frontend framework we chose Angular since it was at the time the most supported and hence most used framework along with Ionic Capacitor, even

when at the end of our development Ionic's and Capacitor's support for React.js gained more traction than Angular.

While our client package was hosted to be downloaded on the App Store and Play Store respectively, we decided to host the backend of GeoChamp on the Backend as a Service (BAAS) platform firebase since it coupled seamlessly with our use of Firebase for further purposes, such as user authentication, analytics, data storage and A/B testing. Between our client and the firebase backend we implemented a thin server layer with a NestJS backend which faciliated implementation aspects such as data validation and data transformation, which otherwise would have to be written in firebase-firestores error-prone security rules language. This layer was also hosted on firebase using the newly launched Cloud Functions hosting service.

All of our used technologies can be viewed in the architectural diagram in figure 3.1 which shows the client separate from the server and the server split into the NestJS API and the firebase BAAS.

3.3.2 The domain model of GeoChamp

GeoChamp is an application which allows users to learn about and quiz themselves on geography. For example users can learn the districts of Vienna by getting asked where the first district is and then having to click the correct district on a map.

The domain entities of GeoChamp are modelled with the help of the composite structural design pattern. There is currently a single subject in GeoChamp, namely geography. The geography subject has multiple topics like for example the Districts of Vienna being one topic. Furthermore, every topic has multiple locations that it contains, where one location of the Districts of Vienna is "Innere Stadt" which is the first of the 23 districts. This creates a composition hierarchy in the form of a mathematical tree, of which a concrete instance is visible in Figure 3.2.

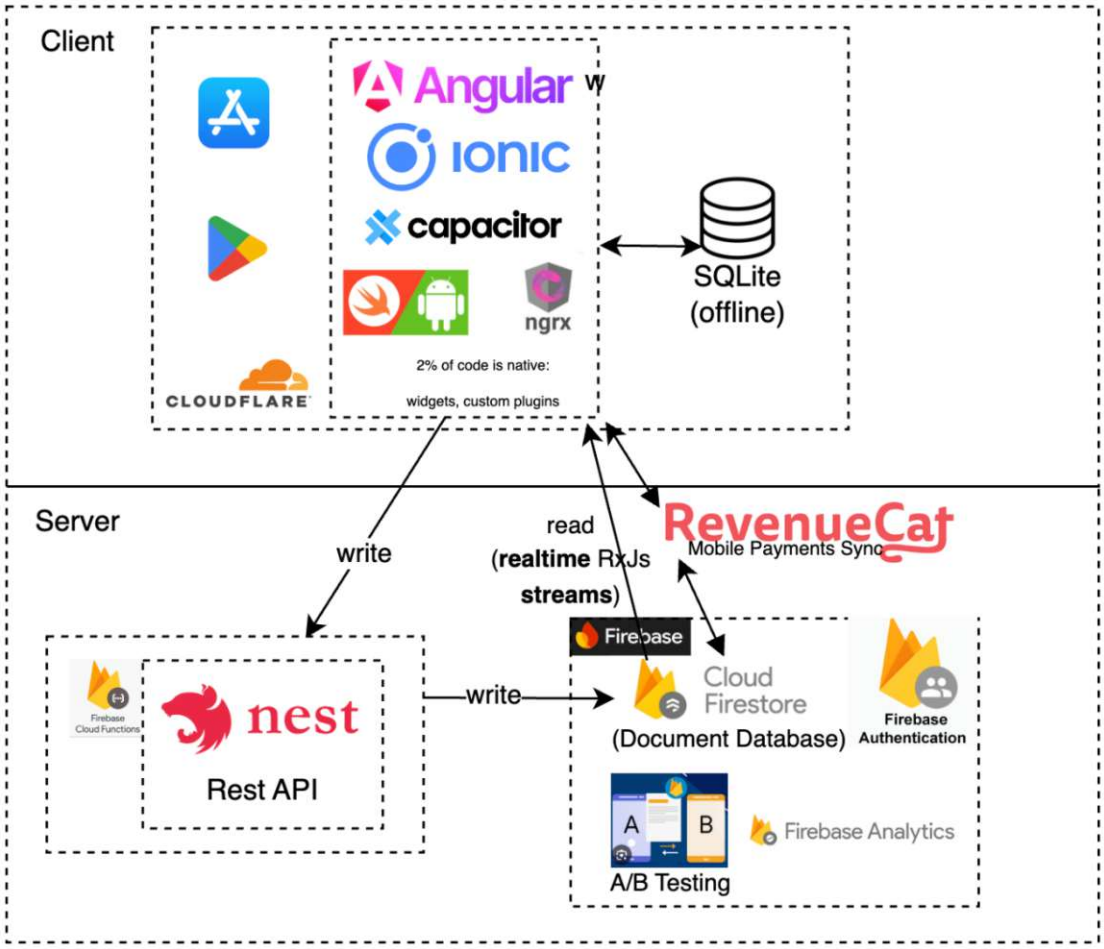

Figure 3.1: An architectural overview of the technologies used in the GeoChamp application

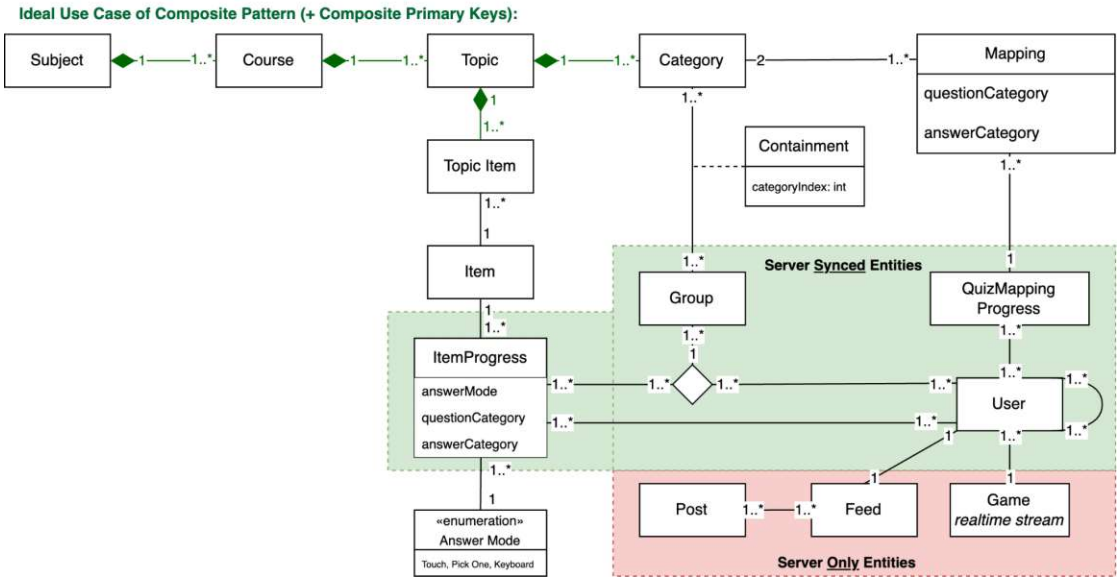

Figure 3.2: UML diagram showing the entity relations within GeoChamp

The key trait of the composite design pattern is that the relations between parents and children are always of type "one to many" respectively and not "many to many", as one child entity can only have one parent entity, not multiple. For this reason, the composite pattern also frequently uses composite primary keys, which define the containment of a child inside its parent [91]. In GeoChamp, composite primary keys are critical for defining the belonging of locations to topics and in version 2 of GeoChamp, the belonging of topics to courses. So in version 2 of GeoChamp, "Innere Stadt" is contained in the topic "Districts" inside the course "Vienna" inside the subject geography. Hence the containment association of "Innere Stadt" has four primary keys: first, the primary key of its own name "Innere Stadt" and further the primary key of the name of its parent topic "Districts", its parent course "Vienna" and its parent subject "Geography".

Topics in GeoChamp also always have one or more categories. Categories represent different things that can be guessed about locations within a topic. Some examples of categories are map, flag, capital, population, picture but also the name of a location itself is a category of topics since the name of a location can be guessed. The special thing about the category-topic association is that it does not use the composite pattern and it is simply modelled as a many to many relation instead. This decouples the association of a new category with a topic without having to modify its locations, since the association of locations with a topic should not have to change if a new category is added. This relation can be seen in Figure 3.3.

Table 3.1 shows all Courses, Topics and Categories that exist within GeoChamp. Courses are grouped into groups for their display in the user interface and Categories are grouped to facilitate the logic of displaying certain UI Components based on which category the user selected, where if the user is learning with the map category, the map component is displayed. Figure 3.4 shows the Course Vienna, Topic Districts and Category Map having been selected. In Table 3.1 this exact selection is also highlighted in red.

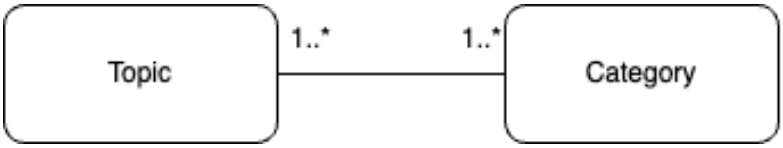

Figure 3.3: UML diagram showing the category-topic relationship.

Figure 3.4: The screen that appears if the user chooses to learn about the Course Vienna, the Topic Districts and the Category Map.

Table 3.1: Overview of Courses, Topics, and Categories

Courses	Topics	Categories
The Whole World:	Countries	**Non-Attribute Categories:**
World Countries	Regions	Flag
World Cities	Districts	Picture
World Nature	Territories	Map
World Sights	Dependent Territories	
World History	Provinces	**Textual Categories:**
	Provinces & Territories	Name
Continent Courses:	States	Capital
Asia	Federal States	Country
Africa	Lands	Country Code
Europe	Cantons	Continent Code
North & Middle America	Subdistricts	
South America	British Isles	**Numeric Categories:**
Australia & Oceania	Islands	Population
	Boroughs	Area
Country Courses:	Manhattan	Height
Ukraine	Neighborhoods	Length
United States	Oblasts	District Number
United Kingdom	Voivodeships	
Canada	Rivers	
Australia	Volcanoes	
Austria	Mountains	
Germany	Mountain Ranges	
Switzerland	Deserts	
Poland	Waterfalls	
France	Tectonic Plates	
Italy	Lakes	
Spain	Oceans	
India	Non-Capital Cities	
Russia	Cities	
China	Capital Cities	
	Sights	
City Courses:	Skyscrapers	
New York City	Religious Sites	
Vienna	Bridges	
	Museums	
	Nice Buildings	
	Parks & Squares	
	Tallest Skyscrapers	
	Theatres	
	Famous Buildings	
	Churches	
	Parks	
	Castles & Palaces	
	Most Famous Sights	
	Universities	
	States and Union Territories	

3.3.3 The implementation of GeoChamp Version 1

Upon installing and opening version 1 of GeoChamp onto a users device, the user is faced with the home screen where he is greeted by an older version of the mascot Geo the globe as can be seen in Figure 3.5. To proceed, the user needs to press the blue play button located at the lower center of the screen.

In the next screen the user is prompted to select the topic that he wants to study. By default the category countries is selected but the user can select a different category where other options are Flags, Cities, Nature, Capitals, Regions, Attributes and Pictures. In the top right corner of the same screen the user can find the game mode selector, where the user can switch between the Quiz and Learn Mode of playing. This screen is displayed in Figure 3.6.

Figure 3.5: The Home Screen

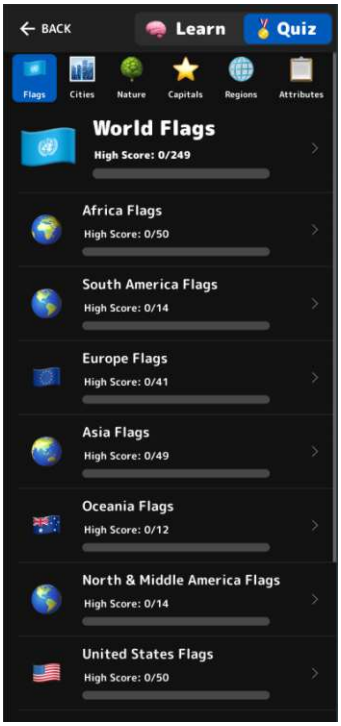

Figure 3.6: The Topic Selection

Figure 3.7: World Flags

When engaging in the Quiz Mode the user is quizzed on his current state of knowledge and after finishing a quiz his score is saved, if it was his highest score so far for the specific selection. The high score for each topic is also displayed as a horizontal progress bar for each topic.

The Learn Mode behaves differently from the Quiz Mode. In the Learn Mode the goal is not to determine the current state of knowledge that the user has but instead to deliberately focus on improving the users current state of knowledge by asking questions about locations that the user has not yet proven to have mastered sooner and more often than questions about locations that the user has shown to have learned.

The flashcard scheduling is implemented with two separate algorithms where one algorithm is used for scheduling locations for short term memory and then after a location was revised correctly 4

times, it is classified as long term memory and is then asked again after 24 hours. For short-term memory scheduling a customized version of the FSRS algorithm that the mobile app Anki uses was used and for long-term memory scheduling the renowned SM2 algorithm was used, which was developed in 1990 by P.A. Wozniak [92].

3.4 Literature Review

We follow the methodology suggested by Webster and Watson [93] for conducting a rigorous literature review, ensuring a comprehensive understanding of gamification's theoretical and practical implications in learning environments.

3.4.1 Overview of deriving gamification features from the literature review into our experiment design

Based on insights gained from the gamification literature review, the goal is to further develop and refine the GeoChamp application. Initially developed as a prototype GeoChamp allows users to study diverse topics of geographic locations from all over the world. Currently, the app already includes sophisticated learning features and many diverse content categories and very little gamification elements like progress bars. Additional gamification features which will be implemented for the user experiment include social elements (following friends, achievement feed) and daily streaks and their corresponding notification-reminders for not breaking the streak. Those features will be ensured to align with the identified best practices in gamification like the ones formulated by Mitchell et. al in [94]. Examples of those features can be seen in Figure 3.8, Figure 3.9 & Figure 3.10 where Duolingo's implementation of those features is shown.

Based on the literature we have identified the five gamification feature dimensions of performance, ecological, social, personal and fictional [34]. Each of those dimensions has two to five subcategories.

After our gamification literature we tried to build out features in as many of the existing dimensions as possible, to provide an overarching deep gamification experience. We redefined the GeoChamp application which in prior did not have many gamification features to have a strongly diverse suite of gamification features.

3.4.2 Planned Literature Review Procedure

Our literature review process was structured into three parts, namely researching about gamification with a focus on learning, researching about learning theories and lastly researching about A/B testing and User Engagement and User Retention.

We conducted our systematic literature search for the topic of gamification using the following keywords: "gamification", "gamification in education", "gamification in learning","Game-Based Learning", "does gamification work", "gamification downsides", and "gamification in X", where X represented one of the domains of "Digital Business", "E-Commerce", "the workplace", "tourism", "health", "exercise", "sustainable consumption", "work/crowd-sourcing", "data gathering", "industrial work processes" and "software engineering".

When researching literature for learning theories we used the keywords "Learning Theories", "Constructivist Learning Theory", "Behaviorism Learning Theory", "Self Determination Theory". Due to the small amount of scientific literature on the connection of the underlying theories of

learning to learning and education itself we also had to research on Google Search for a small amount of supporting articles.

Lastly, when searching for literature on A/B Testing we used the keywords: "A/B testing", "A/B testing in software", "User Retention", "User Engagement", "User Engagement A/B Testing", "User Retention A/B Testing" and "Mobile App A/B Testing".

Adhering to Webster and Watson these keywords were applied across multiple academic databases including Web of Science, IEEE Xplore, ACM Digital Library, and Scopus [93]. After we received our results from our searchers we included and excluded literature in our review based on the title and abstract information selecting only papers that were directly relevant to our research.

3.4.3 Literature Review

After the two new feature sets were implemented and before the development of an A/B testing framework can begin, we will review the newest scientific literature on A/B testing. This will include researching about state of the art best practices in A/B testing, which will include both the technicalities of the data collection and deciding which metrics to consider all the way to analyzing and interpreting the gathered data.

With the gained knowledge from the literature review we will then proceed to connect the GeoChamp application to an analytics collecting service to collect metrics that were found crucial in the literature review. Then we will also change the application to randomly assign users into one of three groups upon first app opening. Users will be split into a control group (A) with the basic feature set, and two experimental groups (B and C) with two different sets where Group B receives Social Gamification Elements and Group C receives the learning streak feature and its linked notifications. This test design falls into the category of multi-armed A/B tests since since with groups A, B and C more than two variants of the product are being compared [82].

Contrary to A/B tests at products with an enormous amount of users like Google's Search Page, where the treatment group often only receives very small changes compared to the control group [83], in our A/B testing experiment we intentionally make the differences between the variants very significant, adding for each a major feature as treatment for variant B and C. This is because we do not expect to be able to get a large userbase until we can start with the testing trial, so we compensate for this aspect by having less granularity, i.e. more drastic differences in our variants, as recommended for smaller user groups by [95].

To measure the impact and difference between certain feature sets on the user groups we will select metrics primarily from the engagement metrics category which is the most commonly chosen metric according to Quin et al. [82]. The user engagement metrics category will include metrics such as active users, engaged sessions per user, average engagement time per month, average engagement time per session, selected event count per session, user retention rate and churn rate. The latter metrics can all be viewed in Table 3.2. If closely before the start of our A/B test the application has gained significant customer attention then monetary metrics such as number of subscriptions or subscription value will also be collected for the A/B experiment.

A/B Testing Measures
• Active users • Engaged sessions per user • Average engagement time per month • Average engagement time per session • Selected event count per session • User retention rate • User stickiness ratios • Churn rate

Table 3.2: Engagement metrics used to measure the impact of different test groups in our A/B test

Once the final implementation step of the A/B testing framework has been done the study trial can begin, where for one month all users will be analytically tracked while they use the application. During this time period no new features will be released for more accurate results to reflect the initial feature set.

Group A	Group B	Group C
• Progress Bars • Learning Path • Game-Like User Interface Design	• Social News Feed • Social News Celebrations	• Daily streak • Streak widget • Notification reminders of not loosing your streak

Table 3.3: Features of the testing groups, where the features of Group A are the base set also available for Group B & C

Gamification elements included in Duolingo: Streaks & Social Elements (2.1.2)

Figure 3.8: Daily Streak

Figure 3.9: Friend List

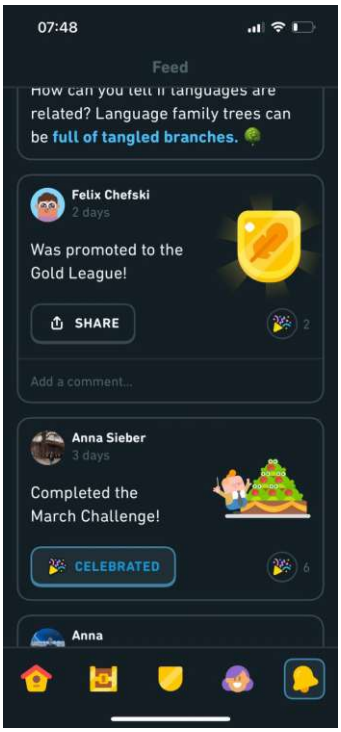

Figure 3.10: Celebrations

In summary, this thesis employs a rigorous methodology encompassing a systematic literature review, application development, A/B testing with the case study subject of a geography learning app. The process starts with a thorough literature review, followed by the enhancement of the GeoChamp app with specific gamification features. An A/B testing framework will then empirically evaluate the effectiveness of these features. An overview of the process of the method can be seen in the Figure of the data collection process, Figure 3.11.

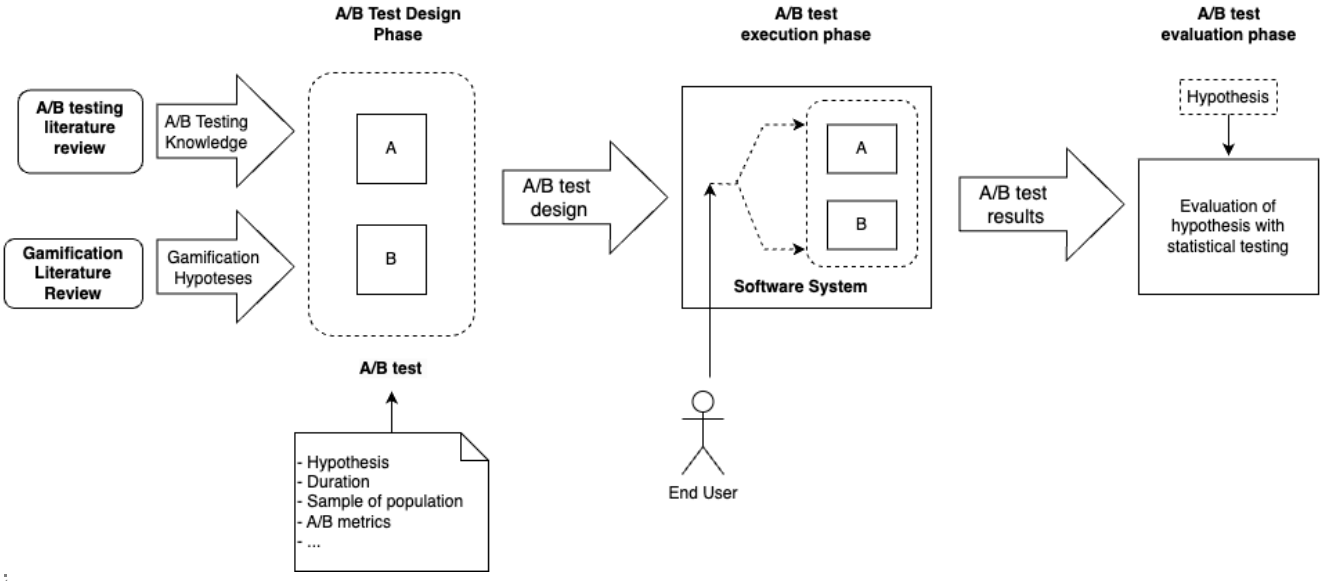

Figure 3.11: Data Collection Process

Formulated based on the described literature review, we now want to test the following two hypotheses using A/B testing:

- *Users in the "Social Gamification Elements" group (Variant B) will show higher user retention compared to the control group (Variant A).*
- Users in the "Learning Streak and Notifications" group (Variant C) will show higher user retention compared to the control group (Variant A).

3.5 A/B Testing

We are employing A/B testing as part of our method. A/B testing being a specific form of standard statistical hypothesis testing is applied primarily in software to determine the difference between two variants of a software, where in most cases variant A serves as the control group and variant B serves as the treatment group which receives the changed version of the software.

3.5.1 A/B Testing Terminology

To measure the success of an A/B test its designer needs to settle on a suiting **Overall Evaluation Criterion (OEC)** which defines the metric of interest for the experiment. Often a companies key performance indicators (KPIs) are chosen as OEC but OECs are different from KPIs as they are specific to the given experiment [96]. For example, an OEC chosen when performing an experiment with the intent to increase user retention would be monthly active users, daily active users or both combined [97].

Every A/B test randomly assigns users to two or more **variants**, in the most simplest form variant A (control) and variant B (treatment) [98].

For ensuring a randomized experiment the A/B test needs to make a randomized comparison, which is why a **randomization unit** is chosen. Randomization units are mapped to each of the variants included in the A/B test. The most common randomization unit chosen is users but also other units such as web-pages or groups of pages [99]. Ideally, units are assigned to variants in pseudo-random process which makes the assignment to a variant persist across all instances of the experiment.

The primary form of A/B testing is the classic single A/B test, followed by the multi-armed A/B tests and sequences of A/B tests. A multi-armed A/B test differs from the classic version by including multiple treatment groups which all receive a different treatment [100]. Multivariate A/B tests are a form of A/B testing similar to multi-armed A/B tests except that not every group receives a distinct feature but combinations of multiple different features are created and assigned to groups [101]. Both multi-armed and multivariate A/B tests offer the benefit that multiple new features can be tested at once, however multivariate tests are also able to discern the cumulative impact of multiple features [102]. Sequential A/B tests are different in that they can test multiple features but do so sequentially one after the other which has the benefit of a larger treatment group size than multi-armed tests [103].

In the following, with **thick clients** we refer to clients that are not browser-based, so clients where the user interface is downloaded and stored on the end users device. Examples of thick clients are mobile apps and desktop apps [104].

3.5.2 Special guidances when A/B testing with thick clients

Thick clients, which are clients that are not web based, so mostly mobile or desktop apps need to be looked at differently when A/B testing. Compared with web pages the process of changing their user interface is a lot more rigid and is dependent on two additional parties than just the developer namely the store that the client is downloaded from and the user who has control of downloading an update to his cached client code. Furthermore, the way data is processed and transferred on thick client devices such as mobile phones is also still slower due to mobile CPUs and mobile network connections and aggregating data locally consumes battery [83]. For those reasons, the following seven guidelines are recommended by Kohavi et al. when performing A/B tests on thick clients:

Firstly, it is advisable to plan multiple experiments ahead for a scheduled roll-out. For many experiments to be released at scheduled dates without being hindered by the app release process, a common strategy is to bundle many experiments at once within the thick client and to disable them at first, after which enabling later at a set point in time, without having to go through the store release process.

Secondly, one should construct his experiments in a way that for many users one expects a delay at the start of the experiment and the start of the logging for users. Some users may still interact with the infrastructure through the old client since they did not download the updated client, which could cause issues with the experiment.

Thirdly, one should consider cases where users are offline upon opening the app, to make sure that variant assignment takes place correctly in such scenarios when dependant on the server and potentially assign a default variant.

Fourthly, if multiple experiments are taking place and since not all users will participate in all experiments at once, it is beneficial to track which users participate in which experiments and do so not at the start of the app but when users actually engage with a certain feature.

Fifthly, other information about the app such as app download size, app battery usage or app notification allow rate should not be neglected when performing A/B tests. A bigger download size or more battery usage leads to fewer installed instances of the app. Furthermore, when introductions more push notifications it might decrease engagement in the long-term since users are turning off notifications for the whole app at some point, which might not be noticeable within the A/B test duration.

Additionally, consider experiments where users are using two completely separate bundles of your app as valid and insightful experiments. This might be done deliberately by releasing two versions within one app bundle. To avoid a large bundle size, one could monitor which users have not updated their thick client to the new version, and use this natural discrepancy as an A/B test, though adoption bias would need to be accounted for.

Finally, consider that multiple different platforms might be used by the same user to interact with the app. This cross-platform interaction should be factored into experiment design and analysis since a user might wrongly receive different generated ids for each platform or there might be interactions between platforms that influence the experiment results.

3.6 The A/B Testing Framework implementation

The A/B test was conducted using the Google Firebase A/B Testing Framework combined with the Firebase Remote Config Service. Both of the mentioned services provide a library interface for the Angular-Capacitor Framework which was used to program the Thick Client.

The Firebase Remote Config was used to fetch and set the two feature flags within the app which included the flags "is_streak_enabled" and "is_feed_enabled". Since the A/B test was done with a thick client with offline capability, we followed the recommendations from Kohavi et al [83] to ensure that the experiment was also running properly if the user has no internet connection. This was achieved by writing the two flags values into the persistence layer as soon as they were fetched, reading them with an online-first fetching strategy where the value from the persistence layer was only fetched if the device was offline. Given the three chosen variants, the two feature flags can be set to different values to achieve the desired variant, all three used combinations are visible in table 3.4.

Variant	**is_streak_enabled**	**is_feed_enabled**
Variant A (baseline)	false	false
Variant B (treatment)	false	true
Variant C (treatment)	true	false

Table 3.4: Feature flag configurations for the variants.

The user distribution will be split up equally into each variant receiving 33% of the visiting users. To ensure this distribution, the firebase A/B testing framework communicates with the firebase remote config service, in order to try to dynamically achieve an equal distribution among variants, as after the test initiation progressively more users open the application.

Due to the Android and iOS app always being seen as separate entities in the firebase console, which includes the A/B testing framework, it was actually necessary to set up two separate A/B testing experiments, one for Android and one for iOS. This had no implications on the experiment

results as the framework was still able to distribute users evenly across both platforms and the experiment results can easily be merged into a single dataset.

In the Firebase A/B testing framework when choosing the desired evaluation criteria we chose the category of "User Engagement" which collects data about many events which are a sign of user engagement, the primary three events being the events "screen_view" which means that the user has first viewed a screen inside the app, the event "user_engagement" which fires when the user has engaged with the app itself with an interaction and the event "session_start" which means that the user has actively started to engage in a usage session through more than just a single interaction. The precise definitions and implementations of those events are defined by Google, nevertheless those are common events which are measured in many A/B testing frameworks.

For collecting data about those events, our thick client uses the Google Analytics Angular-Capacitor library which operates on the native layer, not on the web layer and is hence more tightly coupled to the operating system specifics. The "screen_view" event fires whenever the apps currently viewed screen changes, but due to the analytics library operating on the native layer and due to the native layer consisting only of the web view screen, the "screen_view" event would only fire a single time when the app opens. This is a trade-off and hence a drawback that we have to accept due to the usage of a web view in a thick client. Hence contrary to what is usually expected, we do not anticipate the "screen_view" event to fire significantly more often than the "user_engagement" event.

CHAPTER 4

Results

In the following we present the resulting gamification feature set that was implemented into the GeoChamp application based on the gamification literature review. Following that we relate this feature set to existing literature on gamification. Thereafter, we present the results of the A/B testing experiment.

4.1 Gamification Features Derived from the Literature Review

Two main Gamification feature dimensions were most closely analyzed in the literature review namely the dimensions of performance elements and social gamification elements. Those two dimensions are touched by two gamification frameworks namely the Hexad scale from Tondello visible in Figure 2.2 where social elements can be attributed to the socialiser and performance elements can be attributed to the achiever player type and by the gamification taxonomy of Toda et al visible in figure 2.1.

Performance Elements in Gamification

Performance elements in gamification often include the PBL triad consisting of the elements of Points, Badges and Leaderboards however literature stated that it is crucial when implementing gamification to look beyond the PBL triad and also try to implement further performance elements since the PBL triad itself does not show conclusive results [64].

When looking at the taxonomy of Toda et al. peformance elements include the categories of progression, level, points, stats and acknowledgment [34]. Progression elements allow the user to be aware of their own progress whilst completing actions and level serves as a specific milestone for a player after having progressed a certain amount. Points are basic in that they provide simple feedback to the learner that they have achieved something with value. Stats are more general and reflect the overall multitude of information of the users performance in the environment in a visual manner for overview and acknowledgment rewards the user for a task with a specific acknowledgment that the user gains as recognition.

Related to but taking a different approach than the gamification taxonomy, the Hexad Scale mentions Challenges, Certificates, Learning/New Skills, Quests, Levels/Progression and Boss Battles as elements that appeal to the achiever player type.

Social Elements in Gamification

Social elements in gamification describe the features which can enable and foster social interactions and relations that can be made while interacting with an environment with multiple users. Humans as innately social beings can benefit from such elements as such interactions come as a fundamental natural desire of us.

The taxonomy of Toda et al. includes 4 distinct categories within the social dimension namely competition, cooperation, reputation and social pressure. Competition Elements create a field for users to compete against each other reflecting different achievements of users while cooperation elements give users the opportunity to work together to achieve a common goal. Reputation reflects a users status within the environment which allows differentiation and possibly a the formation of a hierarchy among users. This status is however not strictly based on their performance.

The hexad scale describes one player type who flourishes when he socializes with other players in the system through relating to others, forming guilds/teams, social networks, gaining social status and experiencing social discovery, social pressure and competition.

The Onboarding Process

Upon installing and first opening our developed GeoChamp app the user is still greeted by the mascot Geo as can be seen in Figure 4.1. However this greeting now only appears as part of the onboarding process, where the user is familiarized with the different content that GeoChamp has to offer, starting with the different game modes as visible in 4.2, then ranging from many different courses as can be seen in Figure 4.3, to many different topics for each feature as visible in Figure 4.4 all the way to playing your first session as visible in Figure 4.5.

The User-Onboarding process of GeoChamp

Figure 4.1: Globe Greeting

Figure 4.2: Game Mode

Figure 4.3: Select Course

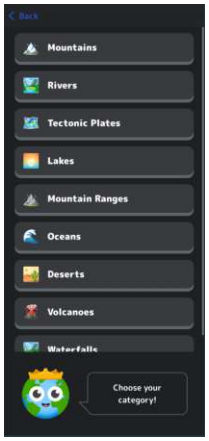

Figure 4.4: Select Topic

Figure 4.5: First Lesson

The Gamified User Interface Components

When comparing GeoChamp's Version 1 to Version 2, what strikes out the most is the reworked visual design of the individual user interface elements. This redesign includes a very colorful color palette widely employed, the redesign of the apps button component and new 2D graphics used. All of those changes and additions are meant to make the user interface of GeoChamp look more like a game, which is an essential part of gamification.

The new color palette of GeoChamp includes 8 color hex codes which are employed widely across the app. Figure 4.6 shows the 8 new colors and Figure 4.16 shows three instances of the training tab where those colors are used for the topic header card and for the color of the path and the path pins themselves. The color palette is also used in the achievements section visible in Figure 4.10.

GeoChamps new button is a central and widely used user interface component. It can be seen as the blue "CONTINUE" button in Figure 4.1 or as all grey interactive buttons in the onboarding process 4.2 4.3 and it is also used in green in the profile pages "Add Friends" button 4.17 or in red in the follower list 4.19. The novelty about the button is that it is designed to appear as a 3D element and additionally upon a user presses it, it gives in the same way as such a button would in the real world. This type of button mechanic is similar to the one that appears commonly for buttons in real world arcade game machines.

Lastly, compared to the native iOS/Android emojis used in Version 1, in Version 2 GeoChamp uses a different set of 2D graphics which appear consistent across platforms and are more two dimensional and cartoon-like with fewer gradients which further make GeoChamp appear more as a game. The 2D graphics used are taken either directly or are modified from the Google Noto Color Emoji font which is an open source font license. Some example usages of the Noto Color Emojis are in the onboarding process where the multiplayer trophy in Figure 4.2 is a modified version of the trophy emoji or in Figure 4.3 and Figure 4.4 where almost every course and topic icon is a Noto Color emoji.

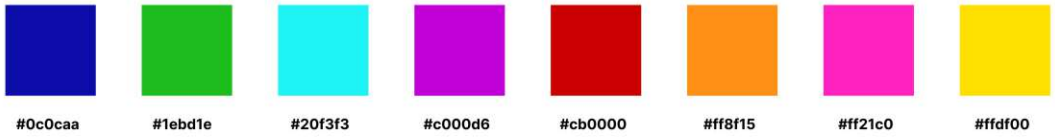

Figure 4.6: GeoChamp's New Color Palette

The Lesson Done and Quiz Done Screen

In general, what a user completes when he engages in the training or quiz game mode is called a session. A lesson is a type of session that is initiated when choosing to engage in the training mode and a quiz is a type if session that is started when choosing to play in quiz mode.

A lesson consists of 20 questions, regardless of the amount of locations inside the topic, even if the topic has less than 20 locations the lesson will still consist of 20 questions. If more than 5 of those questions were answered correctly, then the message in the lesson done screen is rewardingly phrased as visible as in the example message "You're proving what you are capable of" in Figure 4.12. Otherwise, if the user scores less than 5 correct then the lesson done message is phrased in a more encouraging manner as visible in 4.13. A list of all the different messages that can be received in the lesson done screen is shown in Table 1 of the appendix.

Figure 4.7: The World Countries Path

Figure 4.8: The World Nature Path

Figure 4.9: The Germany Path

Figure 4.10: Achievements in GeoChamp

Figure 4.11: The New Color Palette being used in GeoChamp

A quiz consists of as many questions as the user chooses in the quiz configuration modal visible in Figure 4.14 but it can only be at most as many questions as the number of locations contained inside the topic. After the user has finished the quiz they are presented with the Quiz Done Screen where they can see a column chart of how many locations they guessed wrong and how many correctly, as visible in Figure 4.15.

The Profile, Avatars and Followers

After playing his first geography session, the user can now create his own avatar which will appear in his profile page 4.17. There is a wide range of avatars available in GeoChamp and they can be freely customized, as visible in Figure 4.18.

Creating an avatar is not just useful for the user himself but also for other users since in GeoChamp users can add each other as friends and recognize each others avatars. In the profile page there are two numbers that indicate the followers and followings of a user, respectively 4.17. One user can follow another user which makes them receive posts about the updates and achievements of the other user, into their news feed.

The News Feed

The news feed is its own tab in the app and an example news feed can be seen in Figure 4.20. In the news feed users can create reactions to other users achievements, reactions in the form of emojis where five emojis are available as can be seen in Figure 4.20.

The Multiplayer Game Mode

Another feature where GeoChamp users can interact with other users is the newly added multiplayer game mode, which was added in GeoChamp Version 2, in addition to the already present "Learn" and "Quiz" game modes.

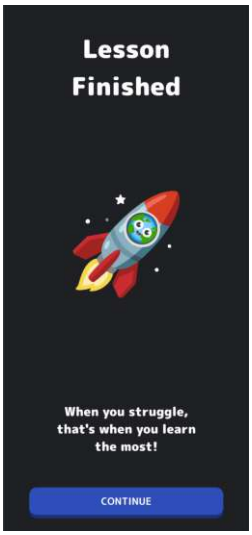

Figure 4.12: Lesson Done with <= 5 correct

Figure 4.13: Lesson Done with > 5 correct

Figure 4.14: The Quiz Configuration Modal

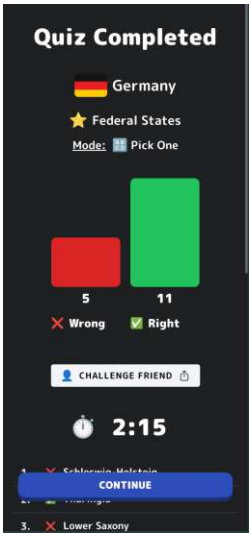

Figure 4.15: The Quiz Done Screen

Figure 4.16: The New Color Palette being used in GeoChamp

The multiplayer game mode exists as its own tab 4.21 and here a specific quiz can be selected and then this quiz is hosted online, with a 4 character alphanumeric code showing up which other users can enter to also join the game.

In the multiplayer game mode users compete to see who can get the highest score of correct answers and after the quiz is ended a leaderboard is displayed where the top three users are placed onto a podium 4.22.

Daily quests and GeoCoins

Furthermore in GeoChamp users get three new quests to complete daily. The quests involve randomly generated tasks ranging from playing a certain number of lessons to guessing locations correctly or to just playing a certain category. Progress made towards the quests is shown after each lesson or quiz, as can be seen in Figure 4.23. Upon completing a quest, the user gets to retrieve the reward of the quest which is in the form of GeoCoins 4.24 4.25.

The Learning Streak, Streak-Notifications and the Streak Widget

To keep the user learning daily, the GeoChamp application uses a learning streak, which among other elements includes a widget which displays the streak and device notifications to remind the user of not loosing their learning streak.

The learning streak reflects the number of consecutive days in a row that the user has learned geography. A learning streak is initiated as soon as the user finishes a lesson or a quiz, where he is then presented with the screen visible in Figure 4.26. As frequent revision is key to long term memory [105] it is the goal of GeoChamp to make the user value and praise their streak. Hence, the current streak number is visible in the top right corner of all three tabs where users can engage in learning or quizzing themselves, namely in the "Training" tab, the "Quiz" tab and the

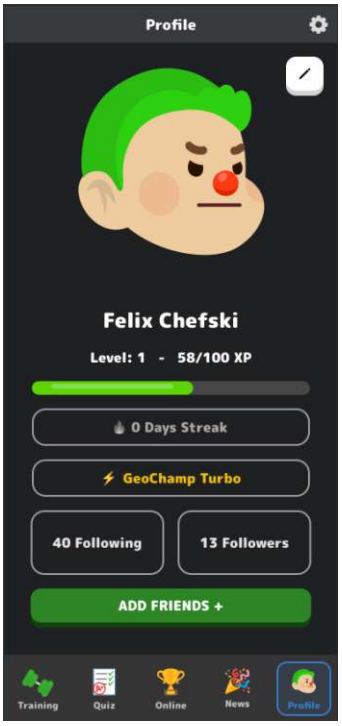

Figure 4.17: The Profile Page

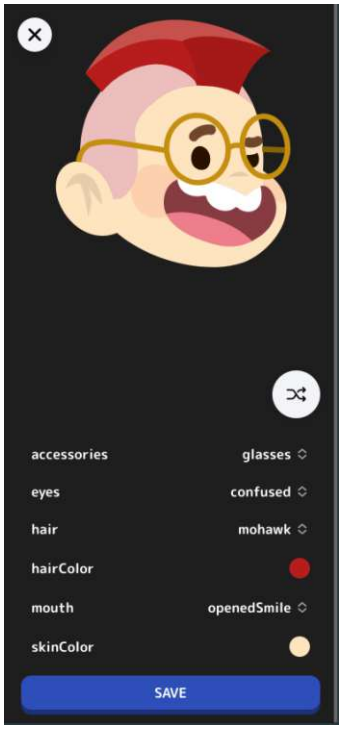

Figure 4.18: Creation of an avatar

"Multiplayer" tab, as visible in Figure 4.27. The streak of friends is also visible in the followers and following tab and in the profile page upon visiting the profile of a friend, as can be seen in Figure 4.19.

The GeoChamp application also has a widget which displays the users current streak. As a reinforced reminder it can be placed centrally onto the users home screen as visible as in Figure 4.28. A display of all the different widgets can be seen in Figure 4.30. The design of the widgets is themed specifically to geography, featuring mostly geographical topics which can be found in the World Nature course, such as a Polar or Saharan Desert widget but the design also features two widgets representing courses namely the Asia Widget and the Africa widget, which display traditional elements of Asia and Africa respectively. Lastly, two widgets such as the Christmas widget and the school widget display non-geographical topics, where the Christmas widget is the only widget that is only present during a specific time of the year, namely from December 1 to December 25.

If a user forgets to learn geography for one day, then their streak gets reset to zero. So to remind users of not loosing their streak GeoChamp employs notifications that appear on the users home screen or as a banner. Contrary to most applications, the notifications in the GeoChamp application are not Push Notifications which are pushed onto the mobile device from a remote server but instead they are Local Notifications which are scheduled and activated locally by the devices operating system and would hence also appear when the users device is offline.

Users receive streak notifications at six fixed times per day and at one special variable time. The

Social Network Features of GeoChamp

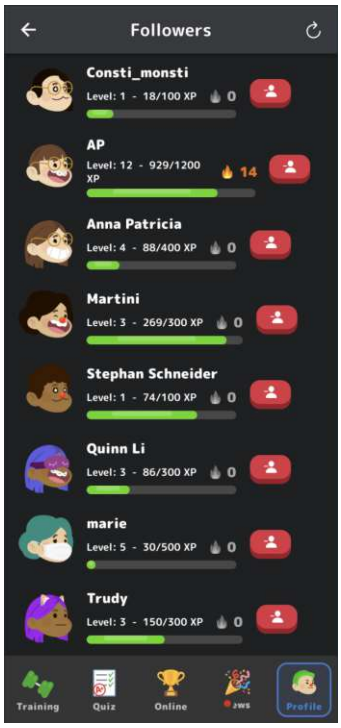

Figure 4.19: Followers List

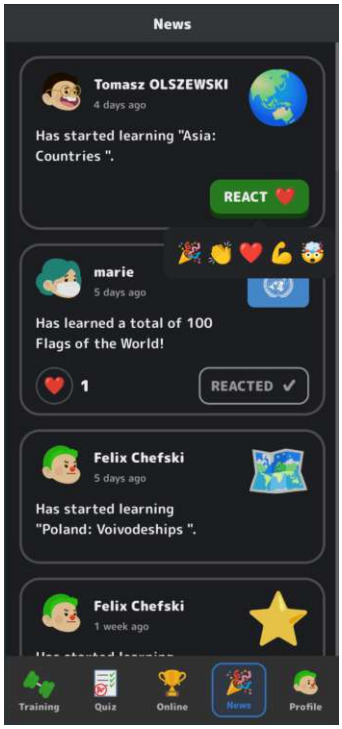

Figure 4.20: News Feed

notifications use funny and creative texts related to learning geography, specifically tailored to the time of the day too. All notification times and texts are displayed in GeoChamp can be found for English in Table 4.1 and for German in the appendix in Table 2. The only time-slot of notifications that is not present in the two tables is the single time-slot which is dynamically scheduled, scheduled based on the time that the user prolonged his streak. This is based on the study from Duolingo where they found out that this is the most effective time to send learning reminders, with the explanation that if a user was able to learn the day before at a certain time, then he would also be able to learn the next day at the same time [106].

Figure 4.21: Multiplayer Lobby

Figure 4.22: Multiplayer Podium

Daily Quests Elements included in GeoChamp

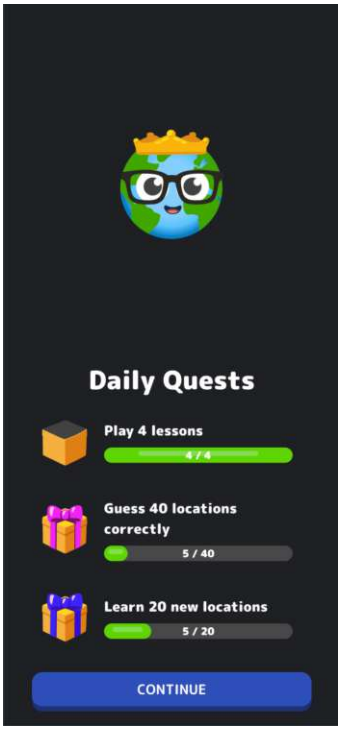

Figure 4.23: Daily Quests

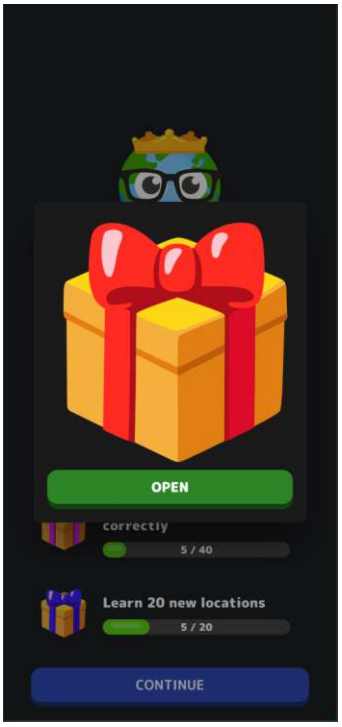

Figure 4.24: Daily Quests Gift

Figure 4.25: Daily Quests Gift Opened

Various Gamification Elements inside the GeoChamp Application

Figure 4.26: A Streak was initiated after playing a lesson

Figure 4.27: The Streak Flame appearing in the Training Tab

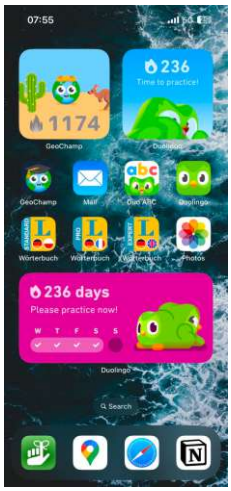

Figure 4.28: The Streak Widget on a users home screen

Figure 4.29: A notification from GeoChamp to practice Geography

Figure 4.30: The Streak Widgets of GeoChamp

Title	Body
LUNCHTIME - 12:00	
Lunchtime Geography Break!	Take a quick mental trip around the world. Perfect with your meal!
Midday Map Challenge!	Spice up your lunch with some exciting world facts!
Global Lunch Quiz!	Feed your mind and belly with some geographical nuggets!
Tasty Geography Treat!	Savor some geographical knowledge with your lunch today!
World Trivia Luncheon!	A side of global knowledge makes any meal better!
Geo-Brunch Challenge!	Start your meal with a serving of global knowledge!
Culinary World Tour!	Explore global cuisines and geography in one sitting!
AFTERNOON - 15:00	
Afternoon World Tour!	Take an exciting geography challenge to boost your energy!
Global Energy Boost!	Recharge with some thrilling world facts and quizzes!
Atlas Adventure Break!	Explore new territories in your afternoon downtime!
Geo-Quiz Power Hour!	Energize your afternoon with a quick geography session!
World Wisdom Break!	Expand your horizons with a brief geography challenge!
Afternoon Atlas Escape!	Take a mental vacation with our geography quiz!
Global Fact Fiesta!	Liven up your afternoon with world trivia!
NEWDAY - 18:00	
Don't Let Geography Escape!	Quick! The countries are trying to hide from your streak!
Your Geography Streak is Hungry!	Feed it some capital cities before it gets hangry!
Geo the Globe Misses You!	Keep your streak flying high with today's lesson!
Save Your Streak!	The continents are plotting to break your streak! Stop them!
Your Maps Are Getting Dusty!	Time to sweep through some geography facts!
Geography Streak Alert!	Don't let your geography skills go on vacation!
Protect Your Geo-Streak!	Countries are waiting to test your knowledge!
EVENING - 20:00	
Evening Globe Trot!	End your day on a high note with some world exploration!
Nighttime Geography Quest!	Your evening adventure through the world awaits!
Twilight World Tour!	Take a mental trip around the globe before bedtime!
Dusk Geography Challenge!	Spice up your evening with geographical brain teasers!
Starlit Geography Session!	Explore the world under the stars!
Moonlight Map Marathon!	Navigate the globe by the light of the moon!
Nocturnal Knowledge Quest!	Wise up on world facts as the night falls!
LATENIGHT - 22:00	
Late Night Geo-Challenge!	Squeeze in some quick geography before the day ends!
Eleventh Hour Geo-Quiz!	It's not too late for a quick world tour!
Midnight Geography Rush!	Race against time for some late-night learning!
Twilight Geography Sprint!	A quick dash through world facts before bed!
Nightcap of Knowledge!	Top off your day with some world trivia!
Bedtime Brain Boost	One last geography challenge before sleep!
Drowsy Geography Dive!	Count countries instead of sheep tonight
FINALCALL - 23:50	
Final Geography Frontier!	Last chance to keep your streak alive! Don't miss out!
Ultimate Geography Showdown!	It's now or never for your geography fix!

Title	Body
Last-Minute World Dash!	Sprint to the finish line with some rapid-fire geography!
Geographical Hail Mary!	Your final shot at maintaining your streak! Go for it!
Eleventh Hour Globe Trot!	Your last chance to circle the world today!
Final Countdown Quiz!	Quick, save your streak with some world facts!
Last Call for World Explorers!	Final boarding call for today's geography journey!

Table 4.1: The English Streak Notifications in GeoChamp

4.2 Analysis of the feature set of GeoChamp in regards to Gamification

4.2.1 Classification of Gamification Elements in GeoChamp

Following the gamification taxonomy proposed by Toda et al [34] the vast array of features that GeoChamp has can be categorized into each of the five dimensions. Each dimension is represented to a different degree in GeoChamp, which also holds for each subcategory of each dimension.

Personal

The personal dimension is by far the strongest represented dimension in GeoChamp, in absolute terms. Puzzle, novelty and sensation are strongly represented, objectives are represented and renovation is weakly represented.

In GeoChamp "Puzzle" is the strongest represented category of all the categories proposed by Toda et al. even if it is not the most typical gamification feature. Puzzle is the core of GeoChamp since the user is constantly solving the challenge and cognitive task of answering geographical questions, which takes place in the guessing page visible in Figure 4.5. This puzzling effect is present in the quiz game-mode as well as in the training game-mode.

Novelty, which is also known as update, surprise and changes manifests itself in GeoChamp through daily quest box openings with the reward being undetermined, visible in Figure 4.24 & Figure 4.25 and it also manifests through the differences in the design and appearance of daily quests changing on a day to day basis, visible in Figure 4.23.

Social

The social dimension of gamification is strongly represented in GeoChamp. It consists of four categories namely reputation, cooperation, competition and social pressure.

The most represented social category in GeoChamp is reputation. Users of GeoChamp can earn reputation by curating a long learning streak and by gaining a lot of experience, which is then visible in their profile 4.17, in the friend list of users that they have added as friends4.19. Additionally, through reaching impressive achievements which show up in ones friends news feed one can earn reputation in GeoChamp 4.20. The reputation reached from an impressive news feed post can be further increased by reaching lots of celebrations from your friends for a post of yours 3.10. Lastly, if a user reaches a high rank on the podium after a multiplayer quiz it gives them a high reputation 4.22.

The second most represented category from the social dimension is competition. Direct competition among users is present in the multiplayer game mode where users compete to see who can achieve

the highest score which gets finally determined in the game-done screen visible in Figure 4.22. Additionally, all of the elements mentioned in the reputation category can be seen as competition elements too if users view the reputation of other users as something they would like to compete against, to also gain a high reputation.

GeoChamp also weakly employs cooperation elements. The only cooperation gamification element in GeoChamp is the reaction that you can make to a users post, visible in Figure 3.10. This acts as cooperation as you are cooperating towards the goal of learning geography and motivating each other along the way.

Finally, social pressure is implicitly employed as a gamification strategy in GeoChamp since the learning streak acts as a central social element in GeoChamp. Having a long learning streak with having added your friends as followers in GeoChamp implicitly builds up social pressure to use the app daily, since if you and your friends are all cultivating a long learning streak then social pressure builds up to keep the streak, to preserve the fundamental need of relatedness [107]. With the widget also displaying your current streak when the app is not opened, this can also make for situations where your streak is visible when someone sees your phones home screen, making the streaks social pressure extend into real life friend encounters and hence being reinforced. Currently however, there are no elements that explicitly create social pressure like for example the option to send a notification to a friend in case if they were seeming to forget to play their daily session.

Performance

Despite it not being the strongest represented dimension, the performance dimension of gamification is certainly represented within GeoChamp. According to the definition of Toda et Al, the performance dimensions consist of five sub-categories, namely Acknowledgment, Level, Progression, Points, Stats.

Acknowledgment features are meant to acknowledge the users performance which in GeoChamp is done through achievements which can be seen in Figure 4.10 and motivational messages at the end of a lesson, which reflect how well a user completed a lesson, visible in Figure 4.12 and Figure 4.13.

The Level category is weakly yet still represented in GeoChamp. It is implemented through the players levels which are thresholded by experience points. The players level can be seen in the profile page 4.17 and in the follower list 4.19. However, the aspect of a new level "unlocking" a new way of gameplay as mentioned by Toda et al. is not present in the app, which is why we classify it as only weakly present.

Progression elements are strongly present in GeoChamp. One of the key features of GeoChamp is the light green shiny progress bars which appear in many places throughout the app, like in the cards for each topic in the learning path 4.7. They also appear in the profile page where they allow the user to view his XP progress within the current level 4.17 or in the achievements page where the progress for each achievement star can be seen 4.10. The most central part of its appearance however, is the guessing page, where it represents the progress throughout a lesson, which can be seen in Figure 4.5.

The category of Points is represented by experience points in GeoChamp which can be seen on the profile page, as shown in Figure 4.17. After completing a quiz, users also receive a score, which according to the definition of Toda et al. also counts towards the Point category. A displayed high score in the quiz done screen can be seen in Figure 4.15, where the score of 11 serves as the points.

The Stats category is represented in GeoChamp through the vertical stack of progress bars within the training and quiz mode, where each topic has its own progress bar, visible in Figure 4.7. However, this is just a rudimentary representation of this category.

Fictional

The fictional dimension is underrepresented in GeoChamp. This category is defined by narrative and storytelling. Storytelling is weakly represented by the mascot of Geo the globe (visible in Figure 4.1), the text from notifications mimicking Geo the globe talking to you, to not loose your streak 4.29 and the message after having completed a lesson 4.12 4.13. This way the app tells the story that the mascot Geo the globe guides the user along his journey of learning and hence incorporates the category of storytelling.

According to the definition of Toda et al. [34], for the narrative category to be present, the interactions and choices made would need to influence a long-term path of the learner. In GeoChamp, since the user has no long term effect on the story that is told by Geo the globe, since notifications and the lesson done text is only based upon the most recent actions that that the user has taken, for this reason the category of narrative is not present in GeoChamp.

Ecological

The ecological dimension, along with the fictional dimension, is in absolute terms the least represented dimension in GeoChamp, despite consisting of the five diverse categories of rarity, economy, imposed choice, chance and time pressure.

Rarity and economy are not present in the GeoChamp app. Despite the user being able to earn GeoCoins an economy is not present in GeoChamp, since GeoCoins currently are not able to be spent on anything. And rarity is missing since currently there are no through chance rare items or rewards that could be reached or earned.

The category of "Imposed choice" is only weakly present since while there is a lot of choice while making guesses in GeoChamp, this choice does not block the advance of the user if he makes the wrong choice, but this is mentioned as a trait in the definition of Toda et al [34].

Time pressure is weakly represented in GeoChamp, only within the quiz game mode the time you take for a quiz is visible in the screen after the quiz is done 4.15. Time is also used as a discriminator between equal high scores so if the user wants to keep improving their high score when already having guessed all locations correctly, then the user could set time as a self-set pressure element, to reach an even better score.

Lastly, chance is represented in GeoChamp but only to a minimal degree with the selection of the next location to be guessed in the guessing page being random for the quiz mode or seemingly random in training mode, where the scheduling algorithm decides the next location to be asked.

4.3 Results of the A/B Testing Experiment

The format of the A/B test that we carried out can be classified as a multi-armed A/B test since there is more than one treatment group, namely two treatment groups of variant B and variant C, which provide treatment on top of the feature set of the control variant A. We conducted this hypothesis test to measure the user engagement effects

The A/B Testing Experiment ran from December 4, 2024 until January 2, 2025 which accounts for a total of 30 days. In total 137 GeoChamp users opened the app during the testing trial

period. It was 48 users on Android and 89 users on iOS. The number of retained users however, so the number of users who returned to the app after their first opening during this period was much lower, where on Android we managed to retain 7 users and on iOS we managed to retain 15 users, for a total of 22 users which equates to roughly 16% of the 137 total users. A table depicting an overview of the experiment can be viewed in Table 4.2.

Table 4.2: Number of Users and Experiment Dates

Experiment	**Number of Users**	**Start Date – End Date**
iOS Experiment	89	December 4, 2024 – January 2, 2025
Android Experiment	48	December 4, 2024 – January 2, 2025
Whole Experiment	137	December 4, 2024 – January 2, 2025

4.3.1 Data collected during the A/B test

The results of the separately collected data for Android and iOS in its rawest form can be seen in Table 4.3 and Table 4.4.

After merging the two tables results and after dropping the two unnecessary and rare data points "os_update" and "app_remove", we receive a new dataset visible in the Table 4.5 and visualized in the Figure 4.31. The remaining events of interest to determine user engagement are the events "screen_view", "user_engagement" and "session_start".

4.3.2 The concluding data from the A/B test

Due to the low number of only 137 users opening the app during the A/B test and due to the low rate of retained users of only 16% in the A/B test we were not able to reach enough user data to attain statistical significance. Nevertheless, we would still like to present our data and draw conclusions based on it.

By combining the Android User Retention Table 4.6 and the iOS User Retention Table 4.7, by weighting the retention percentages according to the number of users having opened the app during the A/B test in each respective experiment, we present a new table showing the key results

Table 4.3: Raw iOS Experiment Data

experimentVariant	eventName	count
Variant A	session_start	18
Variant A	screen_view	119
Variant A	user_engagement	114
Variant B	session_start	26
Variant B	screen_view	85
Variant B	user_engagement	77
Variant C	session_start	15
Variant C	user_engagement	54
Variant C	screen_view	39

Table 4.4: Raw Android Experiment Data

experimentVariant	eventName	count
Variant A	screen_view	54
Variant A	user_engagement	52
Variant A	session_start	15
Variant A	app_remove	1
Variant A	os_update	1
Variant B	user_engagement	20
Variant B	screen_view	12
Variant B	session_start	5
Variant B	app_remove	1
Variant C	user_engagement	36
Variant C	screen_view	16
Variant C	session_start	10

Table 4.5: Android & iOS Experiment Data merged into one Table

experimentVariant	eventName	count
Variant A	screen_view	173
Variant A	user_engagement	166
Variant A	session_start	33
Variant B	screen_view	97
Variant B	user_engagement	97
Variant B	session_start	31
Variant C	user_engagement	90
Variant C	session_start	25
Variant C	screen_view	55

from our A/B test experiment. The weighting used was the fraction 89/137 for iOS and the fraction 48/137 for Android. The combined data is visible in Table 4.8.

The table included five results columns, namely "Ret. Users", "Retention Rate", "% Baseline-Diff", "95% CI" and "P-value". "Ret. Users" is the absolute number of retained users with the variant, "Retention Rate" is the respective relative measure and "% Baseline-Diff" is the difference of the respective variant when comparing with the control group A. In the "% Baseline-Diff" we can see that Variant B with the news feed and celebrations feature performed worse in terms of user retention than its control group.

Looking at the table we notice that out of all the tables Variant C which included the learning streak had the most positive effect on the relative user retention, managing to retain 20.30% of users. Surprisingly, Variant C is followed right by the control group (Variant A) with 17.00% retained users, where Variant A included neither of the two gamification feature sets. Lastly, Variant B which only included the news feed and related celebrations performed the worst overall, only retaining 10.93% of users.

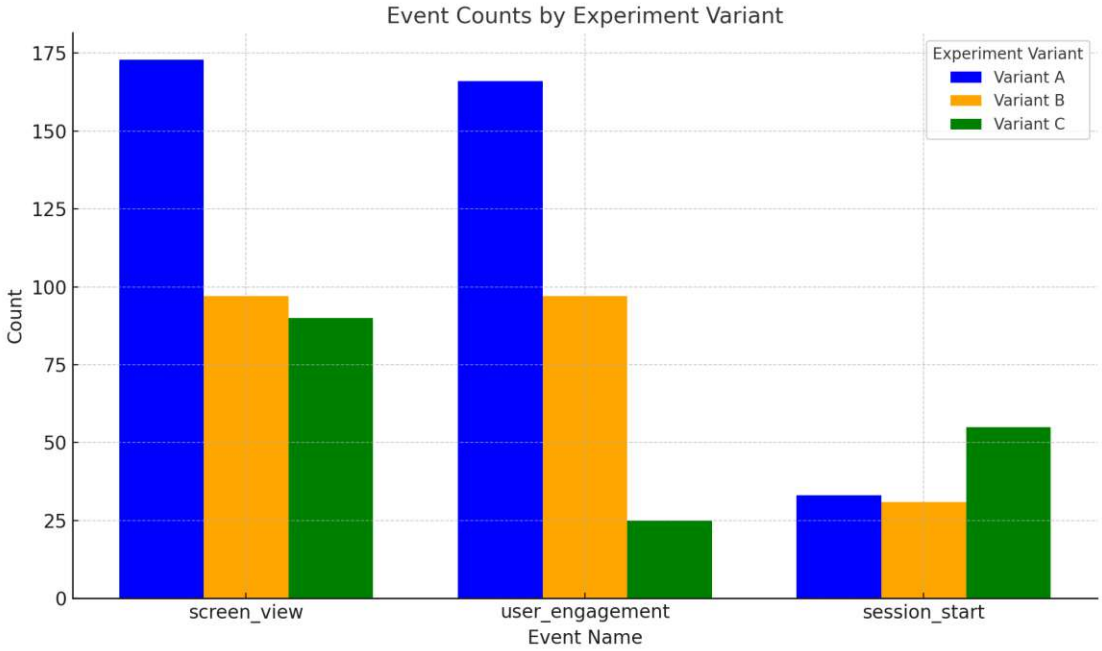

Figure 4.31: Event Counts by Experiment Variant

Table 4.6: Retention Data from the Android Experiment

Variant	Ret. Users	Retention Rate	% Baseline-Diff	95% CI	P-value
Variant A	3	17%	Baseline	Baseline	Baseline
Variant B	1	7.1%	-57%	-28% to 8.8%	0.790
Variant C	3	19%	+13%	-20% to 24%	0.437

The next column "% Baseline-Diff" is derived from the "Retention Rate" column. It shows the same result as the "Retention Rate" column, where "Variant C" retained 19.41% more users than without this feature being implemented and Variant B, where the news feed feature set was added to the application, that decreased the user retention rate by 35.69%.

The last column of the table is the P-Value. It also relates to the "% Baseline Diff" column, where it shows with what probability the difference between the control group and the treatment group is due to chance. With a P-Value of less than 0.05 the difference is unlikely to be purely due to chance. However, in our experiment the P-values for the treatment groups greatly exceed 0.05 with 0.706 for Variant B and 0.362 for Variant C 4.8, showing that the results of Variant B are to be taken with even greater caution than Variant C, but even the difference that Variant C demonstrates is still 36.2% and hence likely to only exist due to chance.

Table 4.7: Retention Data from the iOS Experiment

Variant	Ret. Users	Retention Rate	% Baseline-Diff	95% CI	P-value
Variant A	5	17%	Baseline	Baseline	Baseline
Variant B	4	13%	-23%	-19% to 11%	0.661
Variant C	6	21%	+29%	-12% to 22%	0.322

Table 4.8: Weighted Combined Retention Data

Variant	Ret. Users	Retention Rate	% Baseline-Diff	95% CI	P-value
Variant A	8	17.00%	Baseline	Baseline	Baseline
Variant B	5	10.93%	-35.69%	-22.15% to 10.23%	0.706
Variant C	9	20.30%	19.41%	-14.80% to 22.70%	0.362

CHAPTER 5

Discussion

With the research question of wanting to find out how social and performance elements of gamification influence user retention in learning apps, we performed a literature review on gamification, drew inferences from our findings and with those developed a whole gamified mobile app with which we performed a quantitative gamification evaluation experiment using A/B testing. We would now like to discuss our findings on gamification and how specifically it can help with increasing user retention.

5.1 The potential of gamification in learning

Scientific literature has shown that Gamification can have various positive effects on the learner and the learning process, if applied correctly as we discussed in section 2.1.6. On the one side gamification can serve as a tool to increase user engagement and user retention, on the other side gamification has the potential to even directly increase learning outcomes.

With cultivating high user engagement and user retention often being a major target of today's digital platforms, developers and designers are seeking for new ways to help achieve those goals. Multiple studies like the one from Barrio et al have shown that gamification offers itself as a promising means to elevate user engagement and user retention by attracting users more with game-like elements (Section 2.1.6). Beyond user engagement also other benefits such as improved collaboration skills, problem-solving and communication skills have been shown to be initiated by gamification. However, it is crucial to look beyond just user engagement and user retention and also measure gamifications direct effect on the learning outcome.

In learning apps, retaining the user is not enough since we would also like to make sure that the main goal of learning apps, which is that the user is reaching target learning outcomes, is actually enhanced with gamification. Studies have shown that gamification has the potential to directly improve the learning process with the user being more motivated to perform the intended actions in the applied context (Section 2.1.6). Nonetheless it is crucial to have a nuanced understanding of the benefits of gamification and to make sure to avoid potential pitfalls.

While evaluating research on gamification in learning we noticed several conflicts within the field of gamification science. While since its inception lots of research on gamification was published,

the study results differ a lot and lots and new theoretical claims have been made over the years which shifted the academic view on gamification from just trying to describe and understand gamification superficially to understanding its implications on a deeper psychological level all the way to recently viewing gamification more critically in terms of its effectiveness for learning applications (Section 2.1.8).

Critics mention many biases and effects that led to gamification gaining significance in scientific literature and in the industry including the novelty bias where newly added elements tend to create excitement for users which does not last for a long time but is only initiated by the initial novelty through the change of adding an element (Section 2.1.7).

Over time criticism of gamification became more widespread with new research discrediting prior research and finding gamification to be able to have adverse effects on user retention and even on the learning outcomes (Section 2.1.7). This can also be explained with a co-evolution of gamification and how companies and users make use of it, where interleavedly over time new gamification inventions influenced the usage and user's gamification usage patterns influenced the development of gamification.

The most common mistake made that we found when exploring the literature on gamification was that gamification was often not applied to actually create deep game-like experiences. This includes but it is not limited to the term shallow gamification and it manifests itself in the addition of game elements without considering the gamification as a whole (Section 2.1.9).

Other mistakes that we recognized after reviewing gamification literature was that many developers and designers should be vary that gamification can have many types of adverse effects. Those effects go further than shallow gamification, with the gamification not only not yielding any benefit to the learning application but even having a negative effect on user engagement where users become less likely to use the app or sometimes the detriment is not user engagement but the actual learning effect being hurt by gamification.

To achieve gamification which improves the learning outcomes of your application, developers and designers should create gamification that increases user engagement while simultaneously making sure that the learning outcomes are not negatively impacted by the gamification. To make sure this happens, the focus should be to incorporate overarching game-design principles into the application while avoiding the most common reasons for a negative impact on learning outcomes referenced to as "gamification misuse" (Section 2.1.9).

Following our thorough Literature Review we advise developers and industry designers who want to use gamification to improve a learning apps outcomes to closely consider the implications that gamification can have on your learning product. There are many famous learning apps that suggest that gamification can have a big impact on your apps learning experience including Duolingo, Codecademy [108] and Khan Academy [109]. However, research shows that the task of applying gamification effectively is not trivial and that there are many potential pitfalls that one can fall into when applying gamification.

With our gamified learning app GeoChamp we tried to implement many gamification features to create a deeper complete game-like experience for the user while focusing especially on implementing performance and social elements. The performance features that we implemented target the achiever from the HEXAD scale and the social features attract the socialiser from the same scale.

Performance elements that we included are multiple places where we added the simple PBL element of the progress bar including the guessing page, the learning path cards and the quizzes

tab cards. We also show scores after a quiz with a chart, which falls into the dimension of stats.

To achieve similar engagement effects as happen in a social network we implemented various features that make learning social in GeoChamp. On the one side, users are able to create personal avatars and add each other to their friends list in the app and on the other side there is a news feed where users can see the learning and quizzing activity of other players including being able to react to with emojis, to other users achievements reached by engaging with the app.

Where user engagement refers to the quality of the user experience that emphasizes the positive aspects of interacting with an online application and, in particular, the desire to use that application longer and repeatedly user retention is often seen as a part of user engagement but it places a distinct focus on retaining the engagement of users over a period of time, with the user coming back multiple times to reengage with the product (Section 2.2).

5.2 Practical Evaluation of Gamification

We used user retention as a metric to measure the effectiveness of our gamified learning app GeoChamp. To perform the measurement we implemented an A/B test which measures the change in user retention between three evenly large groups of GeoChamp users which each got a specific set of features activated, with one group having the performance element of the learning streak present in their app, the other having the social element of the friend activity news feed present and the third group acting as a control group or baseline, having neither of the two features.

The feature group which received the learning streak shows an increase in user retention more than the baseline which was expected by us since a learning streak surpasses the shallow element of plain progress bars and makes for an overarching feature which is present in multiple places inside the app and even outside the app in notification reminders.

We can see that the social news feed feature underperformed the baseline user group. This possibly being a case where gamification led to adverse outcomes could theoretically be explained in a similar manner as prior research where the negative impact from social features was caused by a decreased autonomy of users to choose what, when and how often to learn.

The stochastic results of this study are to be taken with caution since we did not reach statistical significance in our experiment. Despite having managed to recruit 137 users to use the application at least once during the A/B testing trial, only a total of 22 users were retained over this period of 30 days. This comparatively small sample of users also resulting in a too low number of retained users makes it hard to rule out that other factors led to the users being retained and hence makes us unable to conclude that the difference between the variants results was due to the differences in the feature sets.

A follow up study might look at recruiting more people over a longer period of time to get significant results. Additionally, a future study might be done with an application that has already gained enough users for the results to become significant. Furthermore, a direct linking of the participants engagement results together with a questionnaire would be useful to determine the underlying causes of the high or low engagement of individual users and would allow us to test and validate proposed psychological patterns and frameworks that aim to explain the effect of gamification better.

Furthermore, this study only looked at gamification in learning apps used in a consumer-facing application with the learning domain not being one used in a professional work education context.

Therefore, future studies could explore the usage of social elements and performance elements of gamification in an industry context, where under different circumstances such as different social interaction structures and more structured application usage schedules the effect of such categories elements could differ a strong amount.

Additionally, while this study only focused on engagement-related metrics such as user engagement and user retention, we did not measure whether the users that had a higher retention also had higher learning outcomes. Therefore, future studies that evaluate gamification in learning apps quantitatively could choose a different KPI such as "newly learned items per week" for their hypothesis tests. This could be especially interesting for our winning variant group which received the daily learning streak, which could possibly enhance retention only since users are using GeoChamp daily to keep their streak but this low daily bar for achievement of a single lesson could also have lead to gamification misuse taking place where users only return to the app daily for a single lesson to prolong their streak or where users prefer to do easier non-challenging lessons which do not extend their range of knowledge, instead of actually engaging with the app for a longer amount of time with challenging learning material.

CHAPTER 6

Conclusion and Future Work

In our research we explored the topic of gamification in learning apps by following a mixed methods approach [11]. We performed a thorough literature review of gamification as a whole with a focus on learning and thereafter we further explored gamification by programming a gamified mobile learning application and measuring gamification directly on this application, with the help of software hypothesis testing.

From our literature review on gamification we concluded that the state of the art of the view on gamification changed significantly over the years after its first mainstream appearance in 2011. Further, we learned about feature and user classifications proposed by researchers and we recognized the criticism on gamification ranging from gamification not having any significant effect all the way to gamification having adverse effects on learning outcomes and being unethical.

It was then our aim to test the hypothesis to compare the effects of two specific gamification feature sets on user engagement and user retention, namely:

> *the feature set including social features the friend activity news feed, including being able to react to posts of your friends*
>
> *the feature set including a learning streak and notifications corresponding to the learning streak*

Unfortunately, the sample size of our experiment was too small and hence we were not able to attain statistical significance with the A/B test that we carried out. Despite not being statistically significant, the results suggest that the **streak and notification feature** as expected had a **positive** effect on user retention while the **news feed feature** unexpectedly resulted in **lower** user retention.

We believe that future research on gamification should stick with the trend of viewing gamification critically and trying to properly understand with which underlying frameworks gamification affects which kind of goals that can be chosen when deciding to apply gamification to an application.

We also suggest that studies that investigate gamification in the future should make sure that when performing a quantitative study they should assure that the sample size of users is big enough and that many of those users have already over a longer period of time shown engagement towards the application that you are using as a subject.

Overview of Generative AI Tools Used

During the writing of this thesis the use of tools using artificial intelligence was strongly limited. The authors only used artificial intelligence in very rare cases when the rephrasing of a sentence proved difficult and in such rare cases the Large Language Model GPT4o was used.wwawA

List of Figures

2.1 Gamification Taxonomy by Toda et al. [34] . . . 6
2.2 HEXAD Player Type Framework [38] . . . 9
2.3 Gamification Framework developed by Schöbel et al . . . 11
2.4 A StackOverflow post . . . 14
2.5 The StackOverflow Profile Page . . . 14
2.6 Gamification Elements in Duolingo (2024) . . . 19
2.7 General A/B testing process from Quin et al. [82] . . . 23

3.1 An architectural overview of the technologies used in the GeoChamp application . 30
3.2 UML diagram showing the entity relations within GeoChamp . . . 30
3.3 UML diagram showing the category-topic relationship. . . . 31
3.4 The screen that appears if the user chooses to learn about the Course Vienna, the Topic Districts and the Category Map. . . . 32
3.5 The Home Screen . . . 34
3.6 The Topic Selection . . . 34
3.7 World Flags . . . 34
3.8 Daily Streak . . . 38
3.9 Friend List . . . 38
3.10 Celebrations . . . 38
3.11 Data Collection Process . . . 39

4.1 Globe Greeting . . . 44
4.2 Game Mode . . . 44
4.3 Select Course . . . 44
4.4 Select Topic . . . 44
4.5 First Lesson . . . 44
4.6 GeoChamp's New Color Palette . . . 45
4.7 The World Countries Path . . . 46
4.8 The World Nature Path . . . 46
4.9 The Germany Path . . . 46
4.10 Achievements in GeoChamp . . . 46
4.11 The New Color Palette being used in GeoChamp . . . 46
4.12 Lesson Done with <= 5 correct . . . 47
4.13 Lesson Done with > 5 correct . . . 47
4.14 The Quiz Configuration Modal . . . 47
4.15 The Quiz Done Screen . . . 47
4.16 The New Color Palette being used in GeoChamp . . . 47
4.17 The Profile Page . . . 48

4.18 Creation of an avatar . . . 48
4.19 Followers List . . . 49
4.20 News Feed . . . 49
4.21 Multiplayer Lobby . . . 50
4.22 Multiplayer Podium . . . 50
4.23 Daily Quests . . . 51
4.24 Daily Quests Gift . . . 51
4.25 Daily Quests Gift Opened . . . 51
4.26 A Streak was initiated after playing a lesson . . . 52
4.27 The Streak Flame appearing in the Training Tab . . . 52
4.28 The Streak Widget on a users home screen . . . 52
4.29 A notification from GeoChamp to practice Geography . . . 52
4.30 The Streak Widgets of GeoChamp . . . 52
4.31 Event Counts by Experiment Variant . . . 59

1 UML Diagram of GeoChamp . . . 80

List of Tables

3.1 Overview of Courses, Topics, and Categories 33
3.2 Engagement metrics used to measure the impact of different test groups in our A/B test 37
3.3 Features of the testing groups, where the features of Group A are the base set also available for Group B & C 37
3.4 Feature flag configurations for the variants. 41

4.1 The English Streak Notifications in GeoChamp 54
4.2 Number of Users and Experiment Dates 57
4.3 Raw iOS Experiment Data 57
4.4 Raw Android Experiment Data 58
4.5 Android & iOS Experiment Data merged into one Table 58
4.6 Retention Data from the Android Experiment 59
4.7 Retention Data from the iOS Experiment 60
4.8 Weighted Combined Retention Data 60

1 English and German Lesson Done Messages 77
2 The German Streak Notifications in GeoChamp 79

List of Algorithms

Listing 6.1: GeoChamp's Spaced Repetition Algorithm Reading Function, used for next item selection. A lower return number means select sooner.

```
1 WHEN learningStage = '${LearningStage.LEARNING}' AND dueDateTime <=
    strftime('%s','now') THEN 1
2 WHEN learningStage = '${LearningStage.REVIEW}' AND dueDateTime <=
    strftime('%s','now') THEN 2
3 WHEN learningStage IS NULL THEN 3
4 WHEN learningStage = '${LearningStage.LEARNING}' AND dueDateTime >
    strftime('%s','now') THEN 4
5 WHEN learningStage = '${LearningStage.REVIEW}' AND dueDateTime >
    strftime('%s','now') THEN 5
```

Listing 6.2: GeoChamp's Spaced Repetition Algorithm Writing Function

```
private static updateSRS(oldItem: AppItemDetail, successful:
    boolean): AppItemDetail {
    const updatedItem = compatibleStructuredClone(oldItem);
    if (!updatedItem.learningStage) { // this is the case where this
        is a new item without progress so far
      updatedItem.learningStage = LearningStage.LEARNING;
      updatedItem.correctStreak = 0;
      updatedItem.eFactor = DEFAULT_E_FACTOR;
    }

    if (successful) {
      updatedItem.learningStage = LearningStage.REVIEW;
      /*** START ANKI ALGORITHM SECTION*/
      if (updatedItem.correctStreak === 0) {
        updatedItem.dueDateTime = dateInNMinutes(5); // 5 minutes
      } else if (updatedItem.correctStreak === 1) {
        updatedItem.dueDateTime = dateInNMinutes(60); // 1 hour
      } else if (updatedItem.correctStreak === 2) {
        updatedItem.dueDateTime = dateInNMinutes(60 * 12); // 12
            hours (1/2 day)
        updatedItem.intervalDays = 1; // NOTE: this is actually just
            useless side info about anki values, -> see next line
        // it is not used by my algo, since mine starts with 4.
```

```
        // Just showing that theoretically i could apply sm2 here
            already.
      } else if (updatedItem.correctStreak === 3) {
        const days = 2;
        updatedItem.intervalDays = days;
        updatedItem.dueDateTime = dateInNMinutes(60 * 24 * days); //
            currently: 2 days
      } else {
        if (!updatedItem.intervalDays) {
          console.error(compatibleStructuredClone(updatedItem));
          console.error('0: Assertion failed, need to handle this!');
          updatedItem.intervalDays = 4; // fallback...
        }
        if (!updatedItem.learningStage) {
          console.error(compatibleStructuredClone(updatedItem));
          throw new Error('1: Assertion failed, need to handle
              this!');
        }
        let newIntervalDays = updatedItem.intervalDays *
            updatedItem.eFactor;
        if (newIntervalDays > 365 * 100) {
          newIntervalDays = 365 * 100; // cap at 100 years, max
              human lifespan which would use this app
        }
        updatedItem.intervalDays = newIntervalDays;
        const intervalInMilliseconds = newIntervalDays * 24 * 60 *
            60 * 1000;
        if (intervalInMilliseconds <= 0) {
          console.error('Assertion failed', intervalInMilliseconds,
              newIntervalDays, updatedItem);
        }
        updatedItem.dueDateTime = new Date(Date.now() +
            intervalInMilliseconds); // custom
      }
      updatedItem.correctStreak++;
      const qFactorWhenCorrect = 5;
      updatedItem.eFactor = updatedItem.eFactor + (0.1 - (5 -
          qFactorWhenCorrect) * (0.08 + (5 - qFactorWhenCorrect) *
          0.02));
    } else {
      /** START AGAIN ANKI ALGORITHM SECTION */
      updatedItem.correctStreak = 0;
      updatedItem.learningStage = LearningStage.LEARNING;
      updatedItem.dueDateTime = dateInNMinutes(0.25); // ~15 seconds
          is 0.25 -> SKIP 1 GUESS FOR SURE. 60 seconds is 1.0
      /** END AGAIN ANKI ALGORITHM SECTION */
      /** START AGAIN SM2 ALGORITHM SECTION */
      const qFactorWhenIncorrect = 0;
      updatedItem.eFactor = updatedItem.eFactor + (0.1 - (5 -
```

```
            qFactorWhenIncorrect) * (0.08 + (5 - qFactorWhenIncorrect)
            * 0.02));

        /**
         * NOTE: we are not using the formula: "let newIntervalDays =
             updatedItem.intervalDays * updatedItem.eFactor;",
         * where eFactor is the poor one, since the sm2 actually has
             hardcoded intervalDays when the correctStreak is set to
             zero,
         * since it starts with the first if-clause again, if
             successful. If unsuccessful, it does not even use read the
             intervalDays at all.
         * So intervalDays actually only gets used in a single case,
             the 5th one.
         */
        updatedItem.intervalDays = null;
        /** END AGAIN SM2 ALGORITHM SECTION */
      }

      if (updatedItem.eFactor < 1.3) { // truncate as in SM2
        updatedItem.eFactor = 1.3;
      }

      if (successful && updatedItem.dueDateTime.getTime() <=
          Date.now()) {
        console.error('after successful guess, dueDateTime CANNOT be
            in the past', updatedItem.dueDateTime);
      }

      return updatedItem;
    }
```

Appendix

English	German
Bad Lesson	
When you struggle, that's when you learn the most!	Wenn es schwer ist, lernst du am meisten!
Every mistake is an opportunity to learn!	Jeder Fehler bietet eine Chance um zu lernen!
Don't give up! Learning takes time...	Gib nicht auf! Lernen braucht Zeit...
Success is built on countless tiny victories!	Erfolg baut auf unzähligen kleinen Siegen auf!
Every expert started as a beginner!	Jeder Experte hat als Anfänger begonnen!
Small steps today, big achievements tomorrow!	Kleine Schritte heute, große Erfolge morgen!
Good Lesson	
You're on fire! Keep that momentum going!	Du bist in Topform! Bleib so am Ball!
Look how far you've come already!	Schau, wie weit du schon gekommen bist!
You're making it look easy now!	Du machst das jetzt schon mit Links!
Each victory makes you stronger!	Jeder Erfolg macht dich stärker!
You're proving what you're capable of!	Du zeigst, was in dir steckt!
Your hard work is paying off beautifully!	Deine harte Arbeit zahlt sich aus!

Table 1: English and German Lesson Done Messages

Titel	Text
LUNCHTIME - 12:00	
Mittagspause Geografie-Pause	Mach eine kurze geistige Reise um die Welt. Perfekt zu Deiner Mahlzeit.
Mittags-Karten-Challenge	Würze Dein Mittagessen mit spannenden Weltfakten.
Globales Mittagsquiz	Füttere Deinen Geist und Bauch mit geografischen Leckerbissen.
Leckerer Geografie-Leckerbissen	Genieße heute etwas geografisches Wissen zu Deinem Mittagessen.
Welt-Trivia-Mittagessen	Eine Portion globales Wissen macht jede Mahlzeit besser.
Kulinarische Weltreise	Erkunde globale Küchen und Geografie in einer Sitzung.
AFTERNOON - 15:00	
Nachmittags-Weltreise	Nimm an einer spannenden Geografie-Challenge teil, um Deine Energie zu steigern.

Titel	Text
Globaler Energieschub	Lade Dich mit aufregenden Weltfakten und Quizzen auf.
Atlas-Abenteuer-Pause	Erkunde in Deiner Nachmittagspause neue Gebiete.
Geo-Quiz Power Hour	Belebe Deinen Nachmittag mit einer schnellen Geografie-Sitzung.
Welt-Weisheits-Pause	Erweitere Deinen Horizont mit einer kurzen Geografie-Challenge.
Nachmittags-Atlas-Flucht	Mach einen geistigen Urlaub mit unserem Geografie-Quiz.
Globale Fakten-Fiesta	Belebe Deinen Nachmittag mit Welt-Trivia.
NEWDAY - 18:00	
Lass die Geografie nicht entkommen	Schnell! Die Länder versuchen sich vor deinem Streak zu verstecken.
Dein Geografie-Streak hat Hunger	Füttere ihn mit Hauptstädten, bevor er hangry wird.
Geo der Globus vermisst Dich	Lass deinen Streak mit der heutigen Lektion hochfliegen.
Rette Deinen Streak	Die Kontinente planen Deinen Streak zu brechen. Halte sie auf.
Deine Karten werden staubig	Zeit für eine Runde Geografie-Fakten.
Geografie-Streak Alarm	Lass Deine Geografie-Kenntnisse nicht in Urlaub gehen.
Beschütze Deinen Geo-Streak	Länder warten darauf, Dein Wissen zu testen.
EVENING - 20:00	
Abendlicher Globus-Trab	Beende Deinen Tag mit etwas Welterkundung auf hohem Niveau.
Nächtliche Geografie-Quest	Dein Abendabenteuer durch die Welt wartet.
Dämmerungs-Welttour	Mach vor dem Schlafengehen eine geistige Reise um den Globus.
Abenddämmerungs-Geografie-Challenge	Würze Deinen Abend mit geografischen Denksportaufgaben.
Sternenlicht-Geografie-Sitzung	Erkunde die Welt unter den Sternen.
Mondschein-Karten-Marathon	Navigiere den Globus im Mondlicht.
Nächtliche Wissenssuche	Werde klüger in Weltfakten, während die Nacht hereinbricht.
LATENIGHT - 22:00	
Spätnacht-Geo-Challenge	Quetsche noch etwas Geografie ein, bevor der Tag endet.
Elfte Stunde Geo-Quiz	Es ist nicht zu spät für eine schnelle Weltreise.
Mitternachts-Geografie-Eile	Wettlauf gegen die Zeit für etwas spätnächtliches Lernen.
Zwielicht-Geografie-Sprint	Ein schneller Dash durch Weltfakten vor dem Schlafengehen.
Schlummertrunk des Wissens	Kröne Deinen Tag mit etwas Welt-Trivia.
Schlafenszeit-Gehirnboost	Eine letzte Geografie-Challenge vor dem Schlaf.
Schläfrige Geografie-Tauchgang	Zähle heute Nacht Länder statt Schafe.
FINALCALL - 23:50	
Letzte Geografie-Grenze	Letzte Chance, Deine Serie am Leben zu erhalten. Verpasse es nicht.
Ultimativer Geografie-Showdown	Jetzt oder nie für Deine Geografie-Dosis.

Titel	Text
Last-Minute-Welt-Sprint	Sprinte zur Ziellinie mit etwas Schnellfeuer-Geografie.
Geografischer Hail Mary	Dein letzter Versuch, Deine Serie aufrechtzuerhalten. Los geht's.
Elfte Stunde Globus-Trab	Deine letzte Chance, die Welt heute zu umrunden.
Finaler Countdown-Quiz	Schnell, rette Deine Serie mit einigen Weltfakten.
Letzter Aufruf für Weltentdecker	Letzter Aufruf für die heutige Geografie-Reise.

Table 2: The German Streak Notifications in GeoChamp

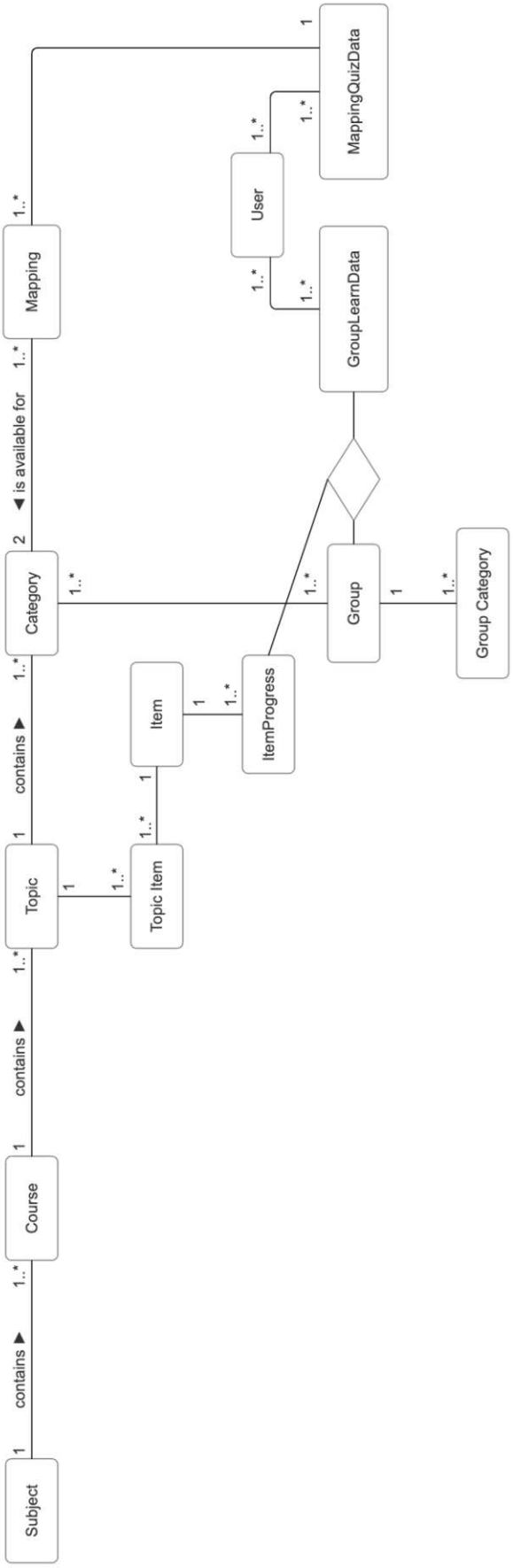

Figure 1: UML Diagram of GeoChamp

Bibliography

[1] C. Blanco, "Changes in duolingo usage during the covid-19 pandemic," Duolingo Blog, 2020, accessed: 2024-05-16. [Online]. Available: https://blog.duolingo.com/changes-in-duolingo-usage-during-the-covid-19-pandemic/

[2] B. Redaktion. (2025) Mimo: Wiener coding-lern-startup verzeichnet 32 mio. user. [Online]. Available: https://brutkasten.com/artikel/mimo-wiener-coding-lern-startup-verzeichnet-32-mio-user/3

[3] Statista, "Top language learning apps worldwide by downloads," 2023, accessed: 2024-05-16. [Online]. Available: https://www.statista.com/statistics/1239522/top-language-learning-apps-downloads/

[4] P. Bitrián, I. Buil, and S. Catalán, "Enhancing user engagement: The role of gamification in mobile apps," *Journal of Business Research*, vol. 132, pp. 170–185, 2021.

[5] C. W. Tan, L. Ling, P.-D. Yu, C. N. Hang, and M. F. Wong, "Mathematics gamification in mobile app software for personalized learning at scale," in *2020 IEEE Integrated STEM Education Conference (ISEC)*, 2020, pp. 1–5.

[6] Z. Wei, J. Zhang, X. Huang, and H. Qiu, "Can gamification improve the virtual reality tourism experience? analyzing the mediating role of tourism fatigue," *Tourism Management*, vol. 96, p. 104715, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026151772200228X

[7] B. J. Morris, C. Dragovich, R. Todaro, S. Balci, and E. Dalton, "Comparing badges and learning goals in low-and high-stakes learning contexts," *Journal of Computing in Higher Education*, 2019.

[8] N. Zaric, V. Lukarov, and U. Schroeder, "A fundamental study for gamification design: Exploring learning tendencies' effects," *International Journal of Serious Games*, vol. 7, no. 4, 2020.

[9] A. Marinho, W. Oliveira, I. Bittencourt, and D. Dermeval, "Does gamification improve flow experience in classroom? an analysis of gamer types in collaborative and competitive settings," *Brazilian Journal of Computers in Education*, vol. 27, no. 2, 2019.

[10] A. N. Saleem, N. M. Noori, and F. Ozdamli, "Gamification applications in e-learning: A literature review," *Technology, Knowledge and Learning*, vol. 27, pp. 139–159, 2022.

[11] A. Tashakkori and J. W. Creswell, "The new era of mixed methods," pp. 3–7, 2007.

[12] P. Dubey, R. L. Pradhan, and K. K. Sahu, "Underlying factors of student engagement to e-learning," *Journal of Research in Innovative Teaching & Learning*, vol. 16, no. 1, pp. 17–36, 2023.

[13] A. Toda, A. Klock, W. Oliveira *et al.*, "Analysing gamification elements in educational environments using an existing gamification taxonomy," *Smart Learning Environments*, vol. 6, no. 16, 2019, received: August 25, 2019; Accepted: November 26, 2019; Published: December 04, 2019. [Online]. Available: https://doi.org/10.1186/s40561-019-0106-1

[14] E. Meşe and Ç. Sevilen, "Factors influencing efl students' motivation in online learning: A qualitative case study," *Journal of Educational Technology and Online Learning*, vol. 4, no. 1, pp. 11–22, 2021.

[15] A. E. Clark, H. Nong, H. Zhu, and R. Zhu, "Compensating for academic loss: Online learning and student performance during the covid-19 pandemic," *China Economic Review*, vol. 68, p. 101629, 2021.

[16] R. Gross, *Psychology: The science of mind and behaviour 6th Edition*. Hodder Education, 2012.

[17] G. E. Hein, "Constructivist learning theory," *Institute for Inquiry. Available at:/http://www. exploratorium. edu/ifi/resources/constructivistlearning. htmlS*, 1991.

[18] N. A. N. Burhanuddin, N. A. Ahmad, R. R. Said, and S. Asimiran, "Learning theories: Views from behaviourism theory and constructivism theory," *International Journal of Academic Research in Progressive Education and Development*, vol. 10, no. 1, pp. 85–98, 2021.

[19] K.-C. Chen and S.-J. Jang, "Motivation in online learning: Testing a model of self-determination theory," *Computers in human behavior*, vol. 26, no. 4, pp. 741–752, 2010.

[20] D. Lamprinou and F. Paraskeva, "Gamification design framework based on sdt for student motivation," in *2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)*. IEEE, 2015, pp. 406–410.

[21] A. Reimann, "Behaviorist learning theory," *The TESOL encyclopedia of English language teaching*, pp. 1–6, 2018.

[22] Educasciences, "Learning theories: Behaviorism," 2024, comprehensive overview of behaviorism learning theory, including its principles, history, applications in education, and criticisms. [Online]. Available: https://www.educasciences.org/learning-theories-behaviorism

[23] ELM Learning, "What is behavioral learning theory?" 2024, comprehensive guide on behavioral learning theory, including core principles, historical background, and practical applications in workplace training. [Online]. Available: https://elmlearning.com/hub/learning-theories/what-is-behavioral-learning-theory/

[24] Western Governors University, "Understanding behavioral theory," 2023, online blog article discussing behavioral learning theory, its key concepts, founders, applications in education, and limitations. [Online]. Available: https://www.wgu.edu/blog/what-behavioral-learning-theory2005.html

[25] M. Gagné and E. L. Deci, "Self-determination theory and work motivation," *Journal of Organizational behavior*, vol. 26, no. 4, pp. 331–362, 2005.

[26] E. L. Deci and R. M. Ryan, *Self-Determination Theory*. Oklahoma State University Library, 2025. [Online]. Available: https://open.library.okstate.edu/foundationsofeducationaltechnology/chapter/4-self-determination-theory/

[27] J. D. Gobert, R. S. Baker, and M. B. Wixon, "Operationalizing and detecting disengagement within online science microworlds," *Educational Psychologist*, vol. 50, no. 1, pp. 43–57, 2015.

[28] R. S. Alsawaier, "The effect of gamification on motivation and engagement," *The International Journal of Information and Learning Technology*, vol. 35, no. 1, pp. 56–79, 2018.

[29] C. Hofacker, K. de Ruyter, N. Lurie, P. Manchanda, and J. Donaldson, "Gamification and mobile marketing effectiveness," *Journal of Interactive Marketing*, vol. 34, 2016.

[30] N. Barz, M. Benick, L. Dörrenbächer-Ulrich, and F. Perels, "The effect of digital game-based learning interventions on cognitive, metacognitive, and affective-motivational learning outcomes in school: A meta-analysis," *Review of Educational Research*, vol. 94, no. 2, pp. 193–227, 2024. [Online]. Available: https://doi.org/10.3102/00346543231167795

[31] N. F. Khaitova, "History of gamification and its role in the educational process," *International Journal of Multicultural and Multireligious Understanding*, vol. 8, no. 5, pp. 212–216, 2021.

[32] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: defining" gamification"," in *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, 2011, pp. 9–15.

[33] N. Zeybek and E. Saygı, "Gamification in education: Why, where, when, and how?—a systematic review," *Games and Culture*, vol. 19, no. 2, pp. 237–264, 2024. [Online]. Available: https://doi.org/10.1177/15554120231158625

[34] A. M. Toda, W. Oliveira, A. C. Klock, P. T. Palomino, M. Pimenta, I. Gasparini, L. Shi, I. Bittencourt, S. Isotani, and A. I. Cristea, "A taxonomy of game elements for gamification in educational contexts: Proposal and evaluation," in *2019 IEEE 19th international conference on advanced learning technologies (ICALT)*, vol. 2161. IEEE, 2019, pp. 84–88.

[35] N. Xi and J. Hamari, "Does gamification affect brand engagement and equity? a study in online brand communities," *Journal of Business Research*, vol. 109, pp. 449–460, 2020.

[36] X. Wu and S. Santana, "Impact of intrinsic and extrinsic gaming elements on online purchase intention," *Frontiers in psychology*, vol. 13, p. 885619, 2022.

[37] F. Xu, D. Buhalis, and J. Weber, "Serious games and the gamification of tourism," *Tourism management*, vol. 60, pp. 244–256, 2017.

[38] G. F. Tondello, A. Mora, A. Marczewski, and L. E. Nacke, "Empirical validation of the gamification user types hexad scale in english and spanish," *International Journal of Human-Computer Studies*, vol. 127, pp. 95–111, 2019.

[39] S. M. Schöbel, A. Janson, and M. Söllner, "Capturing the complexity of gamification elements: a holistic approach for analysing existing and deriving novel gamification designs," *European Journal of Information Systems*, vol. 29, no. 6, pp. 641–668, 2020.

[40] Microsoft, "Minecraft education," Microsoft Education, 2025, a game-based learning platform that inspires creative, inclusive learning through play. [Online]. Available: https://education.minecraft.net/en-us

[41] D. Newnham, "The toca boca story: From sweden to san francisco," *The Startup*, December 2015, interview with Björn Jeffery, CEO and co-founder of Toca Boca. [Online]. Available: https://medium.com/swlh/the-toca-boca-story-dae33acb0adc

[42] R. Al-Azawi, F. Al-Faliti, and M. Al-Blushi, "Educational gamification vs. game based learning: Comparative study," *International journal of innovation, management and technology*, vol. 7, no. 4, pp. 132–136, 2016.

[43] M. Sailer and L. Homner, "The gamification of learning: a meta-analysis," *Educational Psychology Review*, vol. 32, no. 1, pp. 77–112, 2020. [Online]. Available: https://link.springer.com/article/10.1007/s10648-019-09498-w

[44] I. Bogost, "Why gamification is bullshit," 2015.

[45] J. Hamari, J. Koivisto, and H. Sarsa, "Does gamification work?–a literature review of empirical studies on gamification," in *2014 47th Hawaii international conference on system sciences*. Ieee, 2014, pp. 3025–3034.

[46] A. Jacob, A. Faatz, L. Knüppe, and F. Teuteberg, "Understanding the effectiveness of gamification in an industrial work process: an experimental approach," *Business Process Management Journal*, vol. 28, no. 3, pp. 784–806, 2022.

[47] O. Pedreira, F. García, N. Brisaboa, and M. Piattini, "Gamification in software engineering–a systematic mapping," *Information and software technology*, vol. 57, pp. 157–168, 2015.

[48] H. Cavusoglu, Z. Li, and K.-W. Huang, "Can gamification motivate voluntary contributions? the case of stackoverflow q&a community," in *Proceedings of the 18th ACM conference companion on computer supported cooperative work & social computing*, 2015, pp. 171–174.

[49] M. Aparicio, C. J. Costa, and R. Moises, "Gamification and reputation: key determinants of e-commerce usage and repurchase intention," *Heliyon*, vol. 7, no. 3, 2021.

[50] S. Dale, "Gamification: Making work fun, or making fun of work?" *Business information review*, vol. 31, no. 2, pp. 82–90, 2014.

[51] B. Morschheuser, J. Hamari, and J. Koivisto, "Gamification in crowdsourcing: a review," in *2016 49th Hawaii international conference on system sciences (HICSS)*. IEEE, 2016, pp. 4375–4384.

[52] B. D. Douglas and M. Brauer, "Gamification to prevent climate change: A review of games and apps for sustainability," *Current opinion in psychology*, vol. 42, pp. 89–94, 2021.

[53] P. Barratt, "Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling," *Health & place*, vol. 46, pp. 328–336, 2017.

[54] J. Dah, N. Hussin, M. K. Zaini, L. I. Helda, D. S. Ametefe, and A. A. Aliu, "Gamification is not working: Why?" *Games and Culture*, vol. 0, no. 0, p. 15554120241228125, 0. [Online]. Available: https://doi.org/10.1177/15554120241228125

[55] A. García-Jurado, M. Torres-Jiménez, A. L. Leal-Rodríguez, and P. Castro-González, "Does gamification engage users in online shopping?" *Electronic Commerce Research and Applications*, vol. 48, p. 101076, 2021.

[56] W. Hammedi, T. Leclercq, I. Poncin, and L. Alkire, "Uncovering the dark side of gamification at work: Impacts on engagement and well-being," *Journal of Business Research*, vol. 122, pp. 256–269, 2021.

[57] D. Pradhan, G. Malik, and P. Vishwakarma, "Gamification in tourism research: A systematic review, current insights, and future research avenues," *Journal of Vacation Marketing*, p. 13567667231188879, 2023.

[58] A. Mazeas, M. Duclos, B. Pereira, and A. Chalabaev, "Evaluating the effectiveness of gamification on physical activity: systematic review and meta-analysis of randomized controlled trials," *Journal of medical Internet research*, vol. 24, no. 1, p. e26779, 2022.

[59] M. F. Shahzad, S. Xu, O. u. Rehman, and I. Javed, "Impact of gamification on green consumption behavior integrating technological awareness, motivation, enjoyment and virtual csr," *Scientific Reports*, vol. 13, no. 1, p. 21751, 2023.

[60] Y. Lu, X. Mao, M. Zhou, Y. Zhang, Z. Li, T. Wang, G. Yin, and H. Wang, "Motivation under gamification: An empirical study of developers' motivations and contributions in stack overflow," *IEEE transactions on software engineering*, vol. 48, no. 12, pp. 4947–4963, 2021.

[61] C. M. Barrio, M. Muñoz-Organero, and J. S. Soriano, "Can gamification improve the benefits of student response systems in learning? an experimental study," *IEEE Transactions on Emerging Topics in Computing*, vol. 4, no. 3, pp. 429–438, 2015.

[62] J. Rabah, R. Cassidy, and R. Beauchemin, "Gamification in education: Real benefits or edutainment," in *17th European Conference on e-Learning, Athens, Greece*, 2018, pp. 489–497.

[63] C. Almeida, M. Kalinowski, A. Uchôa, and B. Feijó, "Negative effects of gamification in education software: Systematic mapping and practitioner perceptions," *Information and Software Technology*, vol. 156, p. 107142, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0950584922002518

[64] L. Rodrigues, F. D. Pereira, A. M. Toda, P. T. Palomino, M. Pessoa, L. S. G. Carvalho, D. Fernandes, E. H. Oliveira, A. I. Cristea, and S. Isotani, "Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 13, 2022.

[65] M. R. Lepper, D. Greene, and R. E. Nisbett, "Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis," *Journal of Personality and Social Psychology*, vol. 28, no. 1, pp. 129–137, 1973. [Online]. Available: https://psycnet.apa.org/record/1974-10497-001

[66] R. M. Ryan, "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory," *Journal of Personality and Social Psychology*, vol. 43, no. 3, pp. 450–461, 1982. [Online]. Available: https://selfdeterminationtheory.org/SDT/documents/1982_Ryan_ControlandInfo_JPSP.pdf

[67] R. N. Landers, K. N. Bauer, R. C. Callan, and M. B. Armstrong, *Psychological Theory and the Gamification of Learning*. Cham: Springer International Publishing, 2015, pp. 165–186. [Online]. Available: https://doi.org/10.1007/978-3-319-10208-5_9

[68] M. Bitterman, "Classical conditioning since pavlov," *Review of general psychology*, vol. 10, no. 4, pp. 365–376, 2006.

[69] E. A. Locke and G. P. Latham, "New directions in goal-setting theory," *Current directions in psychological science*, vol. 15, no. 5, pp. 265–268, 2006.

[70] B. of Apps, "Duolingo statistics," https://www.businessofapps.com/data/duolingo-statistics/, 2025, accessed: 2025-01-04.

[71] A. Lieberoth, "Shallow gamification: Testing psychological effects of framing an activity as a game," *Games and Culture*, vol. 10, no. 3, pp. 229–248, 2015.

[72] C. Languages, "Why you can't rely solely on duolingo to learn a language and what you should do instead," https://www.corelanguages.com/why-you-cant-rely-solely-on-duolingo-to-learn-a-language-and-what-you-should-do-instead/, 2024, accessed: 2025-01-04.

[73] R. Hadi Mogavi, B. Guo, Y. Zhang, E.-U. Haq, P. Hui, and X. Ma, "When gamification spoils your learning: A qualitative case study of gamification misuse in a language-learning app," in *Proceedings of the Ninth ACM Conference on Learning@ Scale*, 2022, pp. 175–188.

[74] G. Trends, "Duolingo is bad - explore - google trends," https://trends.google.com/trends/explore?date=all&q=duolingo%20is%20bad&hl=en, 2025, accessed: 2025-01-04.

[75] M. Lalmas, H. O'Brien, and E. Yom-Tov, *Measuring user engagement*. Springer Nature, 2022.

[76] H. L. O'Brien, P. Cairns, and M. Hall, "A practical approach to measuring user engagement with the refined user engagement scale (ues) and new ues short form," *International Journal of Human-Computer Studies*, vol. 112, pp. 28–39, 2018.

[77] A. Q. Andrade, A. M. Beleigoli, T. Silva, M. de Fátima H Diniz, and A. L. Ribeiro, "Exploring the user engagement scale short form as a determinant of adherence in digital health interventions," in *MEDINFO 2019: Health and Wellbeing e-Networks for All*. IOS Press, 2019, pp. 1901–1902.

[78] Y. Xie, M. Xu, E. Chow, and X. Shi, "How to measure your app: a couple of pitfalls and remedies in measuring app performance in online controlled experiments," in *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2021, pp. 949–957.

[79] C. Oltra, A. Boso, J. Espluga, and A. Prades, "A qualitative study of users' engagement with real-time feedback from in-house energy consumption displays," *Energy Policy*, vol. 61, pp. 788–792, 2013.

[80] P. C. Brown, H. L. Roediger III, and M. A. McDaniel, *Make it stick: The science of successful learning*. Harvard University Press, 2014.

[81] Z. Luo, "Gamification for educational purposes: What are the factors contributing to varied effectiveness?" *Education and Information Technologies*, vol. 27, no. 1, pp. 891–915, 2022.

[82] F. Quin, D. Weyns, M. Galster, and C. C. Silva, "A/b testing: A systematic literature review," *Journal of Systems and Software*, p. 112011, 2024.

[83] R. Kohavi, D. Tang, and Y. Xu, *Trustworthy online controlled experiments: A practical guide to a/b testing*. Cambridge University Press, 2020.

[84] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

[85] H. Keselman, A. R. Othman, R. R. Wilcox, and K. Fradette, "The new and improved two-sample t test," *Psychological Science*, vol. 15, no. 1, pp. 47–51, 2004.

[86] H.-Y. Kim, "Statistical notes for clinical researchers: Type i and type ii errors in statistical decision," *Restorative dentistry & endodontics*, vol. 40, no. 3, pp. 249–252, 2015.

[87] R. Kohavi and S. Thomke, "The surprising power of online experiments," *Harvard business review*, vol. 95, no. 5, pp. 74–82, 2017.

[88] V. Venkatesh, S. A. Brown, and H. Bala, "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems," *MIS quarterly*, pp. 21–54, 2013.

[89] A. Tashakkori, C. Teddlie *et al.*, "Quality of inferences in mixed methods research: Calling for an integrative framework," *Advances in mixed methods research*, vol. 53, no. 7, pp. 101–119, 2008.

[90] J. W. Creswell, V. L. P. Clark, M. L. Gutmann, and W. E. Hanson, "Advanced mixed," *Handbook of mixed methods in social & behavioral research*, vol. 209, pp. 209–240, 2003.

[91] D. Riehle, "Composite design patterns," in *Proceedings of the 12th ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*, 1997, pp. 218–228.

[92] P. Wozniak, "Supermemo 2004," *TESL EJ*, vol. 10, no. 4, pp. 1–12, 2007.

[93] J. Webster and R. T. Watson, "Analyzing the past to prepare for the future: Writing a literature review," *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002. [Online]. Available: http://www.jstor.org/stable/4132319

[94] M. Shortt, S. Tilak, I. Kuznetcova, B. Martens, and B. Akinkuolie, "Gamification in mobile-assisted language learning: A systematic review of duolingo literature from public release of 2012 to early 2020," *Computer Assisted Language Learning*, vol. 36, 05 2021.

[95] E. M. Feit and R. Berman, "Test & roll: Profit-maximizing a/b tests," *Marketing Science*, vol. 38, no. 6, pp. 1038–1058, 2019.

[96] M. Mandić, I. Gregurec, and U. Vujović, "Measuring the effectiveness of online sales by conducting a/b testing," *Market-Tržište*, vol. 35, no. 2, pp. 223–249, 2023.

[97] P. Chandar, B. St. Thomas, L. Maystre, V. Pappu, R. Sanchis-Ojeda, T. Wu, B. Carterette, M. Lalmas, and T. Jebara, "Using survival models to estimate user engagement in online experiments," in *Proceedings of the ACM Web Conference 2022*, 2022, pp. 3186–3195.

[98] N. Ju, D. Hu, A. Henderson, and L. Hong, "A sequential test for selecting the better variant: Online a/b testing, adaptive allocation, and continuous monitoring," in *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, 2019, pp. 492–500.

[99] A. Deng, J. Lu, and J. Litz, "Trustworthy analysis of online a/b tests: Pitfalls, challenges and solutions," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017, pp. 641–649.

[100] D. Xiang, R. West, J. Wang, X. Cui, and J. Huang, "Multi armed bandit vs. a/b tests in e-commerce-confidence interval and hypothesis test power perspectives," in *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2022, pp. 4204–4214.

[101] C. Holzmann and P. Hutflesz, "Multivariate testing of native mobile applications," in *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*, 2014, pp. 85–94.

[102] C.-Y. Yang, L. Lei, N. Ho, and W. Fithian, "Bonus: Multiple multivariate testing with a data-adaptivetest statistic," *arXiv preprint arXiv:2106.15743*, 2021.

[103] Y. Bondarenko, "Sequential a/b testing," in *2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT)*. IEEE, 2019, pp. 488–491.

[104] P. P. Hung, T.-A. Bui, M. A. G. Morales, M. Van Nguyen, and E.-N. Huh, "Optimal collaboration of thin–thick clients and resource allocation in cloud computing," *Personal and ubiquitous computing*, vol. 18, pp. 563–572, 2014.

[105] N. J. Cepeda, H. Pashler, E. Vul, J. T. Wixted, and D. Rohrer, "Distributed practice in verbal recall tasks: A review and quantitative synthesis." *Psychological bulletin*, vol. 132, no. 3, p. 354, 2006.

[106] K. P. Yancey and B. Settles, "A sleeping, recovering bandit algorithm for optimizing recurring notifications," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2020, pp. 3008–3016.

[107] J. Tang and P. Zhang, "Exploring the relationships between gamification and motivational needs in technology design," *International Journal of Crowd Science*, vol. 3, no. 1, pp. 87–103, 2019.

[108] P. Fotaris, T. Mastoras, R. Leinfellner, and Y. Rosunally, "From hiscore to high marks: Empirical study of teaching programming through gamification," in *European Conference on Games Based Learning*. Academic Conferences International Limited, 2015, p. 186.

[109] M. M. Tenório, R. P. Lopes, L. A. d. Góis, and G. d. S. Junior, "Influence of gamification on khan academy in brazilian high school," *PUPIL: International Journal of Teaching, Education and Learning*, vol. 2, no. 2, pp. 51–65, 2018.