Department of Psychology and Cognitive Science

Doctoral School in Cognitive Science

XXXVI Cycle

Doctoral thesis

The Significance of Gameful Systems Design: A Journey through the Evolution of GamiDOC, from its Conceptualization to Experimental Validation

Supervisor

Annapaola Marconi

Massimo Zancanaro

Antonio Bucchiarone

PhD Candidate

Simone Bassanelli

Academic year 2022-2023

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Antonio Bucchiarone and Annapaola Marconi, for believing in me and providing guidance along this arduous path. Your support and encouragement have been instrumental in shaping both my academic and personal development. I am also deeply thankful to Lennart Nacke and all the friends of The Games Institute who have contributed to my personal and professional growth. I will always carry with me the time we spent together. I am deeply grateful to my family and friends for their support. Their love, encouragement, and understanding have provided me with the strength I needed throughout this journey. I would also like to address a thought to my grandparents. Thank you for teaching me what determination and dedication are all about. Your legacy continues to inspire me every day. I want also to express my deepest gratitude to my girlfriend Marta. Your trust propelled me forward even when the path seemed most challenging. I will always be thankful for your presence in my life. Last, I would like to extend my deepest gratitude to my three beloved kittens for enriching my life with their presence. Probably don't know how important they were during the writing of this thesis.

Simone Bassanelli

27/03/2024
Trento, Italy

Contents

List of Figures

List of Tables

Abstract

The integration of digital games into diverse aspects of human culture has led to a paradigm shift in their utilization beyond mere entertainment. This doctoral thesis explores the challenges in designing effective gameful systems and presents a novel solution, GamiDOC, to address these challenges. The research begins by identifying elements that interfere with the design of gameful systems and subsequently affect their effectiveness. Scientometric analyses reveal the need for systematic and standardized methodologies in gamification design. Challenges such as users' interpersonal differences, goal disparities, contextual variations, and feedback modalities are thoroughly examined. In response to these challenges, GamiDOC is introduced as a tool to support designers and developers throughout the design, development, and evaluation phases of gameful systems. The tool is built on a Design for Science Research Method (DSRM) approach, providing a systematic and holistic framework. GamiDOC includes features such as a gamification design framework, a peer-review procedure, code generation, an open-access database, and a guideline for data collection and analysis.

The effectiveness of GamiDOC is assessed through empirical studies and real-case applications. User feedback demonstrates the tool's usefulness in addressing challenges during the design phase, supporting decision-making processes, and enhancing overall development and evaluation. Real-case applications include the design and evaluation of a Virtual Reality (VR) gameful system for teaching UML and a serious game promoting environmental awareness. Despite the promising results, the thesis acknowledges the need for further development and evaluation of GamiDOC's components. Future work involves refining features, resolving ethical constraints, and conducting more comparisons with alternative methodologies.

In conclusion, this dissertation contributes to the understanding of challenges in gamification design and offers a practical solution in the form of GamiDOC. The presented tool, still under development, shows potential in guiding designers and developers towards creating more effective and engaging gameful systems, contributing to the ongoing evolution of gamification in various domains.

Chapter 1

Introduction

Games have seamlessly integrated into the fabric of human culture, exerting a continuous influence over our social and recreational spheres. This prominence has catalyzed the incorporation of digital games into endeavors that extend beyond mere amusement (Seaborn & Fels, 2015). The utilization of digital games has been documented across diverse domains and for various objectives (Dye, Green, & Bavelier, 2009; Franceschini et al., 2013; Griffiths, 2003; Mainetti, Sedda, Ronchetti, Bottini, & Borghese, 2013; Primack et al., 2012). It is now firmly established that gaming, functioning as a motivating and immersive endeavor, facilitates the persuasion of individuals to break undesirable habits and transform their conduct (Bassanelli, Vasta, Bucchiarone, & Marconi, 2022). Indeed, game environments are recognized as formidable frameworks that sustain users' motivation (Koivisto & Hamari, 2019b). However, back in the day, people began to think that any kind of video game was fun and engaging and had by default the ability to sell numerous copies. But of course, they were wrong. Just think of the famous case of the video game E.T. or the crisis of 1983 (Ernkvist, 2008). This showed us how involvement and fun are not only elements related to all video games but only to those that follow high-quality design and development.

Over the past decade, the concept of "gamification", denoting the integration of game design elements into contexts outside of games (Deterding, Dixon, Khaled, & Nacke, 2011), or as per an alternative definition, "the application of game-like thinking and mechanics to engage users and address challenges" (Huotari & Hamari, 2011;

Zichermann & Cunningham, 2011), has garnered escalating interest across multiple disciplines (Bassanelli et al., 2022). This surge in interest is due to the ensuing motivational advantages, enhanced creativity, sense of play, engagement, and holistic positive development and well-being (Hamari, 2019). Gamification has increasingly become an adopted methodology, offering an ancillary approach for domains that traditionally struggle to capture users' complete attention and motivation, e.g., training (Armstrong & Landers, 2018), education (Bucchiarone, Cicchetti, Bassanelli, & Marconi, 2021; Koivisto & Hamari, 2019b), behavior modification (Bassanelli & Bucchiarone, 2022; Bucchiarone, Bassanelli, Luca, et al., 2023; Bucchiarone, Bassanelli, & Marconi, 2023). According to Bozkurt and Durak (2018), the fundamental objective of implementing gamification is to elevate users' motivation, thereby fostering experiences that are more effective, efficient, captivating, enduring, and enjoyable.

However, the design of game-inspired systems is intricate, necessitating a plethora of precautions to ensure the development of a functional system. Indeed, as indicated by various sources in the literature (Koivisto & Hamari, 2019b; Seaborn & Fels, 2015), there exists a dearth of consensus regarding the theoretical foundations and scope of gamification, leading to inconsistent outcomes related to the implementation of gameful systems. These outcomes can partially be attributed to the absence of standardized design methodologies and the prevalent utilization of a *one size fits all* approach (Böckle, Micheel, Bick, & Novak, 2018). In essence, gameful systems are frequently fashioned without accounting for the distinct interactions that various user categories have with these systems. Consequently, the design and development of gameful systems seems to be following the same paradigm that led to the crisis of the video game domain in the past: the mere addition of game elements, often without further analysis, is deemed adequate to motivate users, without taking into account additional factors. To address this inconsistency in outcomes, numerous solutions have been proposed. Certain authors recommend that the design of gameful systems ought to acknowledge the disparities and preferences among end-users (Codish & Ravid, 2017; Koivisto & Hamari, 2019b; Oliveira et al., 2022; Tondello et al., 2016). In contrast, others have introduced specific frameworks for gamification design to facilitate the meticulous construction of gameful systems (Deterding, 2015; Klock et al., 2016; Morschheuser, Maedche, &

Walter, 2017), considering specific requisites. However, although there have been numerous proposals in the literature — for example, Morschheuser, Maedche, and Walter (2017), Klock et al. (2016), Hunter and Werbach (2012), and Hunicke, LeBlanc, and Zubek (2004) —, no suitable design methodologies or background articles have been identified to support the design phase of educational systems. Some of the reviews I have conducted — and which are reported in Sections 2.3 and 2.4 — support these findings, loudly pointing to the need for a reference for researchers and practitioners to be able to develop truly effective systems, avoiding an unnecessary waste of time and funds.

1.1 Research questions

Starting from these data, I stated four main research questions for this dissertation:

> ***RQ 1****: What are the elements that interfere with the design of gameful systems and, subsequently, their effectiveness?*
>
> ***RQ 2****: How can we address the challenges that exist in gameful systems design, and create an effective way to support designers and developers?*
>
> ***RQ 3****: Does the identified procedure adequately support the design phase?*
>
> ***RQ 4****: Does the use of GamiDOC enable the creation of effective systems?*

1.2 Outline of the dissertation

Although this dissertation bears my name, I will often refer to the studies with *"we"*, since many aspects have been addressed as a research group and not solely by me. Following a Design for Science Research Method (DSRM) approach (Peffers et al., 2007), I sought to highlight the issues, challenges, and possible solutions related to gameful systems design, development, and evaluation. First, I will analyze the challenges that are still present in gamification design, development, and then effectiveness, following the data presented by Koivisto and Hamari (2019b), Hamari, Koivisto, and Sarsa (2014), and Seaborn and Fels (2015), and referring to what is reported in the papers

"Gamification for Behavior Change: A Scientometric Review" (Bassanelli et al., 2022), *"Lost in Gamification Design: A Scientometric Analysis"* (Bassanelli, Gini, Bonetti, Bucchiarone, & Marconi, 2024), *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024), and *"The role of game modality in the outcomes of gamification: A research agenda"* (Gini, Bassanelli, & Bucchiarone, 2023) — presented in Chapters 2 and 3 (**RQ 1**). Then, I will analyze the alternatives in the literature, noticing that there is no single holistic option that encompasses all the necessary elements. At this point, I will describe the theoretical model we have designed to support gamification design, along with the tool's development and evaluation, which is implemented in an online platform (**RQ 2**). The model is presented in Chapter 4 and refers to the paper *"GamiDOC: A Tool for Designing and Evaluating Gamified Solutions"* (Bassanelli & Bucchiarone, 2022), and partially *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024). Having taken cues from issues in the literature and other reference models in game design and gamification design, the proposed theoretical design model aims to guide users, starting from contextual information, achieving the production of a gamification design document (GDD), necessary to link design and development of tools and games (Baldwin, 2005). Then, the model's flow guides the users through several different aspects: the selection of a different gamification design framework, a validation of the design of the gameful systems, a possible validation of the data collection and analysis methods (if any), the code generation for the game rules through domain-specific languages (DSL), the game elements analysis through an open access database, and a guideline for the data collection and data analysis. Since the project is huge and the development of all components could not be completed during a 3-year doctoral period, most of them are currently under development. Chapter 5 reports several studies related to the perceived usefulness of the tool and initial assessments of the theoretical design model, reporting both usability analyses and real-case scenarios (**RQ 2** and **RQ 3**), described also in *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024). Chapter 6 presents the evaluation of two different tools that used GamiDOC during the design phase (**RQ 3** and **RQ 4**). The studies are described in *"Gamification- and Virtual Reality-Based Learn-*

ing Environment for UML Class Diagram Modeling" (Yigitbas, Schmidt, Bucchiarone, Bassanelli, & Engels, 2024), and *"Untitled Bee Game: Be(e)ing Mean to Learn More about Eco-sustainability"* (Bonetti, Bassanelli, Bucchiarone, Gini, & Marconi, 2024). Last, Chapter 7 concludes the document, summarizing the points presented within, providing a final discussion by answering the various research questions, and providing guidance on future work to achieve a complete, working, and validated tool.

Chapter 2

A scientometric approach to the background literature

To provide a comprehensive analysis of the reference literature, in this chapter two different scientometric reviews are described. These reviews provide us with a detailed analysis of the structure of the literature related to gamification used in behavior change and, subsequently, on the design methods used, giving us insights into the most important papers, authors, and keywords and how these have changed over time. Related to this data, I will next present (in Chapter 3) a detailed analysis of all the elements that have contributed — and are still contributing to this change — and that have highly influenced the work of GamiDOC and its components. We performed two scientometric analyses: the first one — Section 2.3 — is related to the use of gamification to pursue behavior change and it's extrapolated from the paper *"Gamification for Behavior Change: A Scientometric Review"* (Bassanelli et al., 2022), while the second one — Section 2.4 — is an analysis of gamification design, and it's extrapolated from the paper *"Lost in Gamification Design: A Scientometric Analysis"* (Bassanelli, Gini, Bonetti, et al., 2024).

The literature on gamification in the behavioral change field is chosen for analysis due to the ultimate goal of gameful systems: supporting and encouraging behavior change in users across various domains, including education, environmental sustainability, and other topics. It is therefore logical to think that gamification, by its very principles, is not something to be used by people who are already motivated and enacting a certain behavior. The analysis of drop-outs in gameful systems is an element to be critically analyzed, as it is considered a symptom of both failure — in cases where the user abandons it due to lack of motivation — and success in using the system — in all those cases where users find no utility in using it because they already enact the selected behavior or are motivated a priori. The rationale behind the analysis of gamification design stems from the results of the first analysis, which show that most articles and authors continue to question current design procedures.

2.1 Towards the scientometric analyses

To further analyze the current situation regarding gamification and the design process, Scientometrics can be described as the study of how the literature is structured around a topic, mapping the scientific citation, and understanding the impact of publications in a specific research field (Leydesdorff & Milojević, 2012). Given that gamification is a relatively novel subject and is interrelated across diverse domains, it becomes imperative to examine the structural makeup within each gamification field. In these studies, a compilation of documents was sourced from the Scopus[1] database. These documents encompassed articles, references, authors, and keywords. Employing co-citation techniques, which gauge the occurrence of multiple publications, authors, or keywords being referenced together in other publications (C. Chen, 2016), the material was categorized. The analysis transpired through the utilization of the CiteSpace software[2] (C. Chen, 2014; C. Chen & Morris, 2003), integrating network and timeline assessments. The CiteSpace software visually depicted documents through interactive maps. The software employed various parameters and metrics to gauge the influence of documents, authors, and keywords within particular clusters or across the entire network. This facilitated the identification of paramount documents, authors, and keywords in the realm of gamification literature pertinent to behavior change and positive behaviors over time, and then to gamification design.

[1] https://www.scopus.com/.
[2] http://cluster.cis.drexel.edu/~cchen/citespace/.

2.2 Data analysis and visualization

In both studies, we used the same analyses, metrics, and visualization techniques. To explore the interconnectedness of scholarly documents, a method known as Document Co-citation Analysis (DCA) was undertaken. This involved assessing the frequency with which multiple documents were jointly cited in subsequent publications (Aryadoust, 2020; Carollo, Bonassi, et al., 2021; C. Chen, Song, Yuan, & Zhang, 2008). This analytical approach delves into co-citation networks, focusing on deciphering the patterns of document clusters that emerge from shared citations (C. Chen, Ibekwe-SanJuan, & Hou, 2010). When two documents garner substantial co-citations, it suggests a thematic affinity between them (Bar-Ilan, 2008). Author Co-citation Analysis (ACA) was employed to determine instances where authors were cited in conjunction with one another. ACA unveils higher-order connectivity patterns among authors, shedding light on their academic relationships (C. Chen et al., 2008). To unearth influential keywords and their evolution over time, a keywords analysis was conducted. This method serves to uncover the central themes of the articles (X. Chen & Liu, 2020). By scrutinizing keyword usage and co-occurrence, emerging and cutting-edge subjects can be identified (Xie, 2015). The utilization of the CiteSpace software not only yields cluster perspectives but also offers a chronological overview. This software generates co-citation insights aligned with the temporal progression, enriching the analysis process (Xie, 2015).

2.2.1 Metrics

To explore the characteristics of the networks and clusters, a range of temporal and structural metrics derived from co-citation were employed. In evaluating the structural attributes of the network, specific parameters were taken into consideration, including betweenness centrality, modularity Q index, and average silhouette. Additionally, temporal and hybrid metrics, such as citation burstness and sigma (Σ), were also integrated into the analysis (Carollo, Bonassi, et al., 2021; C. Chen, 2014; C. Chen et al., 2009, 2010).

Betweenness Centrality: This metric pertains to each node within the network and gauges the extent to which a node serves as a linkage between other nodes in the network (C. Chen et al., 2010; Freeman, 1977). In essence, it quantifies how effectively a single node functions as a bridge, connecting otherwise isolated nodes (Carollo, Bonassi, et al., 2021). With a scale ranging from 0 to 1, higher values (close to 1) signify a node that connects multiple sizable groups of nodes, thus indicating influential documents or journals in the network (Gaggero et al., 2020).

Modularity Q Index: This index measures a network's capacity to be divided into modules or clusters. With values ranging between 0 and 1, lower values (near 0) suggest that the network lacks clearly defined clusters, while higher values (above 0.7) indicate a well-structured network divided into distinct groups. Values nearing one suggest components that remain largely isolated from one another (Aryadoust, Tan, & Ng, 2019; Carollo, Bonassi, et al., 2021; C. Chen et al., 2010; Gaggero et al., 2020).

Silhouette: Represented by a score that varies between -1 and 1, the silhouette score gauges the homogeneity of a cluster. A high score (above 0.7) signifies a cluster's internal consistency and distinctness from other clusters. A moderate score (around 0.5) indicates a reasonable clustering outcome. Scores near zero indicate objects in a cluster bordering neighboring clusters, while negative values suggest incorrect cluster assignments (Carollo, Bonassi, et al., 2021; C. Chen et al., 2010; Rousseeuw, 1987; Zhou, Chen, & Meng, 2019). Silhouette has two applications: measuring cluster homogeneity and estimating network partitioning (average silhouette score) (Aryadoust, Zakaria, Lim, & Chen, 2021).

Burstness: It relates to a sudden surge in citations for a node within a short time frame within the overall time period (C. Chen et al., 2010; Kleinberg, 2003). This metric reflects an abrupt upswing in research attention directed at a publication within a specific timeframe (Aryadoust et al., 2019).

Sigma (Σ): Sigma serves as an indicator of scientific novelty and is calculated by combining betweenness centrality and citation burstness, represented as $\Sigma = (centrality + 1)^{burstness}$. Elevated sigma values signify works with greater influential potential, characterized not only by their strategically vital structural attributes but also their distinctive temporal implications (C. Chen, 2016; C. Chen et al., 2009, 2010; Gaggero et al., 2020).

2.2.2 Clustering

To identify document clusters within the framework of Document Co-citation Analysis (DCA), we employed the clustering feature within CiteSpace. This algorithm forms clusters of publications by evaluating the strength of connections between documents that reference and those being referenced. The labels assigned to these clusters are derived from noun phrases and index terms, utilizing three distinct algorithms: Log-Likelihood Ratio (LLR), Mutual Information (MI), and Latent Semantic Indexing (LSI) (C. Chen, 2014). These algorithms adopt varying approaches to discern the themes of the clusters. For instance, LSI operates on document matrices, while LLR and MI deduce cluster themes by analyzing noun phrases found within the abstracts of citing articles (C. Chen et al., 2010). All three algorithms were implemented to automatically assign labels to the clusters, and after evaluation, the LLR algorithm was selected in both studies for comparison due to its capacity to provide distinct labels with satisfactory coverage. The clusters that emerged through the application of the LLR algorithm were subsequently numbered in a descending sequence based on their cluster size, following the recommendation of the software's creator (C. Chen, 2014).

2.3 Gamification for behavior change: A scientometric review

Our aim in this analysis is to provide an accurate overview of the literature's structure and to describe in a structured and systematic fashion the developments and trends behind the **gamification-related literature in the domain of behavior change**, reporting the most influential documents, authors, and keywords. Considering the aim of our research, we state the following research questions:

> ***RQ A.1****: What are the most influential documents in the gamification for behavior change field?*

> ***RQ A.2****: Who are the most influential authors contributing to the research of behavior change?*

RQ A.3*: How have research trends in gamification applied to behavior modification changed over time?*

2.3.1 Methods

In this section, we present the method adopted in our study. It is mainly composed of two macro-steps: (i) the **Literature search** (Figure 2.1) and, (ii) the **Data analysis and visualization**. The following sections present all the details needed to understand the study protocol used and possibly replicate it.

2.3.1.1 Literature search

The data used for the analyses include 984 publications on gamification and its application in the field of behavior change and positive behavior published between January 1st, 2012 and February 24th, 2022, with 46,609 unique references downloaded from the Scopus database. The time range of publications depended uniquely on Scopus' availability and no a-priori temporal exclusion criteria was applied. From an initial pool of 1,001 documents, we excluded those that were written in languages other than English and duplicates, thus arriving at a final sample of 984 (Figure 2.1). The search code used was `"( TITLE-ABS-KEY ( gamif* ) AND TITLE-ABS-KEY ( "behav* change" ) OR TITLE-ABS-KEY ( "positive behav*" ) )"`. The search terms gamif* and behav* were chosen as they take into account all possible forms derived from the root (i.e. gamif* covers also gamification, and the verb gamify in all its forms; behav* covers behavior, behavioral, behaviour, and behavioural). Gamification is a recent discipline in large part because its literature is made up of book chapters and conference papers. The reason for choosing Scopus over other databases was its coverage of books, book chapters, reference books, and scientific publications (C.-K. Huang et al., 2020; Pranckutė, 2021).

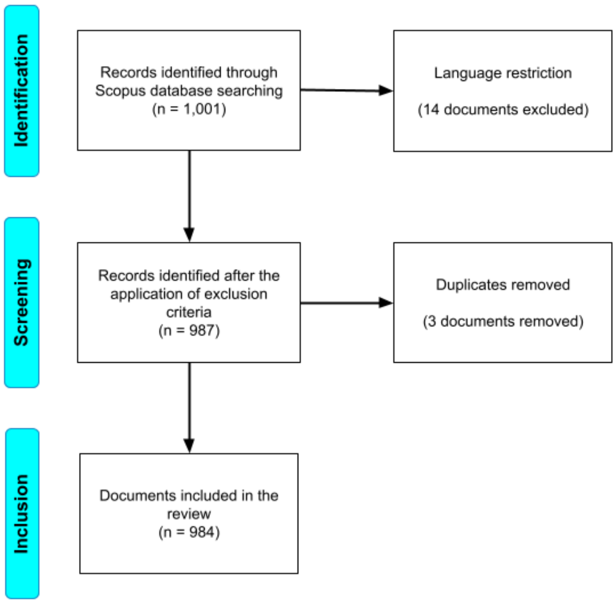
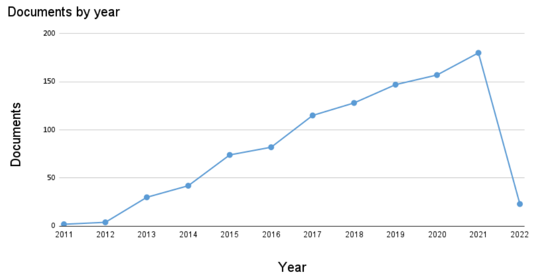


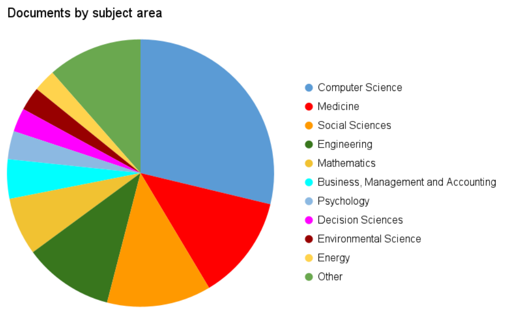
Figure 2.1: Prisma diagram of the literature search.

Figure 2.2 presents the results of the frequency analysis performed on the sample, revealing the number of documents by year, the most productive institutions, authors, and countries, and the subject areas. Figure 2.2a presents the total number of documents by year. Overall, exponential growth can be observed in the research domain, except for the last year, because it refers only to January and February. The Scopus database presents only two and four publications in the first and second year respectively, reaching 30 publications after two years; they reach their highest point in 2021 with 180 publications, which corresponds to 20.1% of the total production in the Scopus database. Figure 2.2b shows the different areas of application according to the Scopus database division. The biggest area corresponds to the computer science domain (28.9%); this is understandable as most of the gamification is implemented in software and mobile applications. The second most frequent applications of domains are medicine, and social sciences (12.7% each), followed by engineering (10.9%), and mathematics (7%). Other areas of application are business, management, and accounting (4.8%), psychology (3.4%), decision sciences, and environmental science (2.9% each), and energy (2.7%). Figure 2.2c presents the 10 most productive authors. J. Hamari is the most prolific with 15 publications, followed by R. Orji and A. Marconi with 14 and 13 publications respectively. Other prolific authors are M. S. Patel

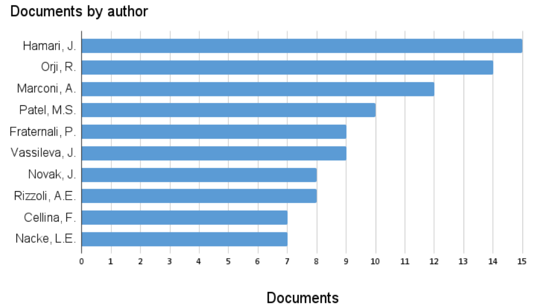
(10 documents), P. Fraternali (9 documents), J. Vassileva (9 documents), J. Novak (8 documents), A. E. Rizzoli (8 documents), F. Celina (7 documents), and L. E. Nacke (7 documents). Figure 2.2d presents the most prolific countries. The first to appear was the United States (167 documents), followed by the United Kingdom (119 documents), Germany (96 documents), and Australia (69 documents). Other prolific countries are Spain (65 documents), Italy (61 documents), Canada (56 documents), Netherlands (45 documents), Switzerland (45 documents), and Finland (42 documents). Figure 2.2e presents the 10 most productive institutions. Tampere University is the most prolific with 16 documents, followed by the University of Pennsylvania, VA Medical Center, and Fondazione Bruno Kessler with 15 documents each. Other prolific institutions are the Queensland University of Technology and The University of Oulu (14 documents each), Politecnico di Milano (13 documents), the University of Pennsylvania Perelman School of Medicine, The University of Auckland, and The University of Waterloo (with 12 documents each).



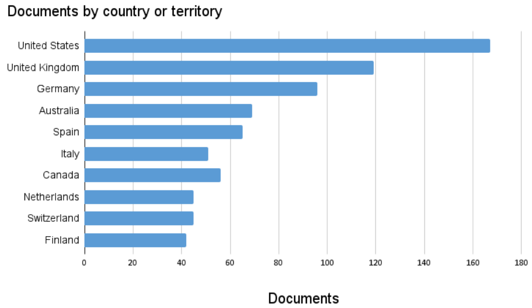
(a) Number of documents per year from January 1st 2012 to February 24th 2022.



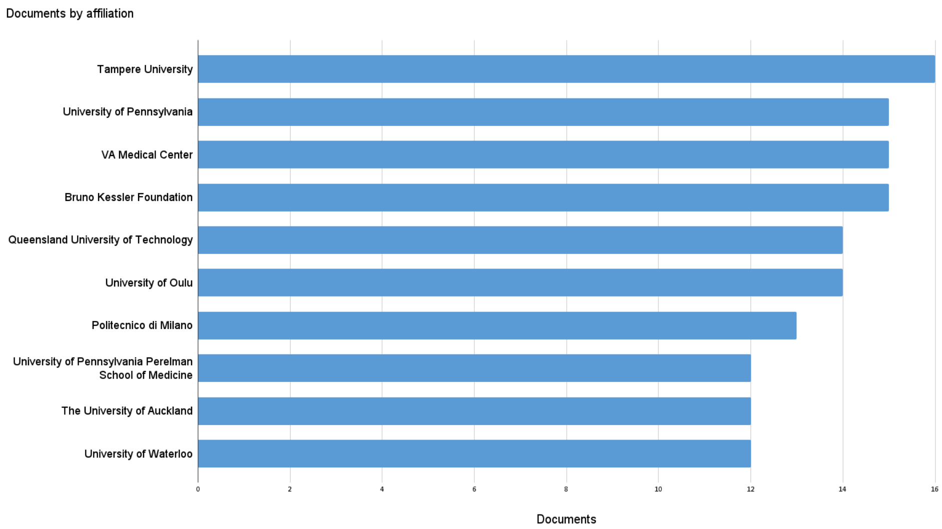
(b) Subject area.



(c) Number of documents by author.



(d) Number of documents by country.



(e) Number of documents by affiliation.

Figure 2.2: Analysis of the Scopus search.

2.3.1.2 Data Analysis and Visualization

The data from the Scopus database were converted to a CiteSpace-friendly format (C. Chen, 2014) with the information related to each of the 984 publications retrieved.

At this point, we used the CiteSpace software (version 5.8.R3) to analyze the data. Of the total references cited, 44,682 of the 46,609 (95%) were considered valid. A small loss of references is due to data irregularities that cannot be processed by CiteSpace. This percentage of unprocessed references can be considered as a negligible loss of data (C. Chen, 2016).

2.3.1.3 Settings

To generate and analyze the networks with CiteSpace, we set no time span, with the time slicing outline at one per year. We compared three criteria for node selection to identify the optimal DCA, ACA, and keyword analysis networks: Top N, Top N%, and g-index. Top N function picks up the N most cited articles and uses information from them to form the network for each time slice. Top N% includes the Top N% most cited articles in each time slice to construct the network. G-index is an improvement of the h-index that allows to measure the global citation performance of a set of articles. It is the *"(unique) largest number such that the top g articles received (together) at least g^2 citations"* (Egghe, 2006). The networks built with Top N with N at 50 and 25, Top N% with N at 5 and 10, and g-index with a scaling factor at 10 and 25 were compared.

Overall, we selected the networks with Top N at 25 for DCA and keyword analysis, and Top N at 50 for ACA, since they provided better overall effects on the network's structural metrics, number of nodes and links, and a major consistency in the cluster structure for DCA. Furthermore, to obtain the best network possible, we set CiteSpace parameters *"Link Retaining Factor"* and *"Maximum Links per node"* as unlimited. After a first check, we decided to set *"Look back years"* as *"100"* to remove the few outlier values related to a few internet site references with wrong temporal information, leading to alterations in the timeline representation. The selected network for each analysis refers to the largest connected component, that is the largest subnetwork in which you can start from any node and reach any other node (C. Chen, 2016).

2.3.1.4 Analysis

Document co-citation analysis (DCA) was performed to examine the frequency in which multiple documents have been cited together in later publications (Aryadoust, 2020;

Carollo, Bonassi, et al., 2021; C. Chen et al., 2008). The study of co-citation networks focuses on interpreting the nature of clusters of co-cited documents (C. Chen et al., 2010). If two documents receive high co-citations, they can be thematically connected with each other (Bar-Ilan, 2008). Author co-citation analysis (ACA) was performed to identify the times authors were cited together. It allows the identification of higher-order connectivity patterns between authors (C. Chen et al., 2008). Keyword analysis was carried out to detect the most influential keywords and their change over time. It can provide information about the core content of the articles (X. Chen & Liu, 2020); analysis of keywords and their co-occurrence can help us find hot and cutting-edge topics (Xie, 2015). Besides producing a cluster view, CiteSpace software can also generate a timeline view. For all the analyses mentioned before, this provides co-citation information as a function of the time sequence (Xie, 2015).

2.3.1.5 Clustering

Cluster labeling was conducted automatically using all three algorithms. After a first check, we decided to use the LLR algorithm to compare the occurrences of terms in the citing articles. The clusters obtained through the LLR algorithm were numbered in descending order according to their cluster size. This approach is supported by the software creator (C. Chen, 2014), since the cluster labeling LLR provides the best results in unique labeling with sufficient coverage. The labels obtained were checked by experts to modify duplicate or unsuitable labels (i.e. labels that did not match the content of the cluster). A detailed description of this renaming process can be found in the "results" section, within the "Document Co-citation Analysis" subsection. In the cluster view, which displays a spatial representation of the diagram (Figure 2.5), the thickness of the node reflects the amount of cited references inside the clusters. The passage of time is represented by the color shading from the oldest (purplish) to the newest (yellowish). In addition, multi-colored rings reflect the burstness (red) and betweenneess centrality (purple).

2.3.2 Results

In this section, we provide a set of results according to the adopted metrics for each CiteSpace analysis used. Hence, we describe each cluster found through cluster analysis.

2.3.2.1 Document Co-citation Analysis (DCA)

The DCA provided a network with 922 nodes and 18,602 links, showing a modularity Q index of 0.8561 and an average silhouette metric of 0.9649, suggesting that the network was sufficiently divisible into clusters (according to the Q index) and that each cluster was highly consistent (according to the silhouette index) (Figure 2.5).

DCA resulted in the identification of 18 co-citation clusters (Table 2.7) sorted from the largest in size (cluster #0 = "Evaluating behavior change intervention", size = 142, silhouette = 0.861, mean year = 2010) to the smallest (cluster #12 = "Persuasive mobile application", size = 6, silhouette = 0.991, mean year = 2010).

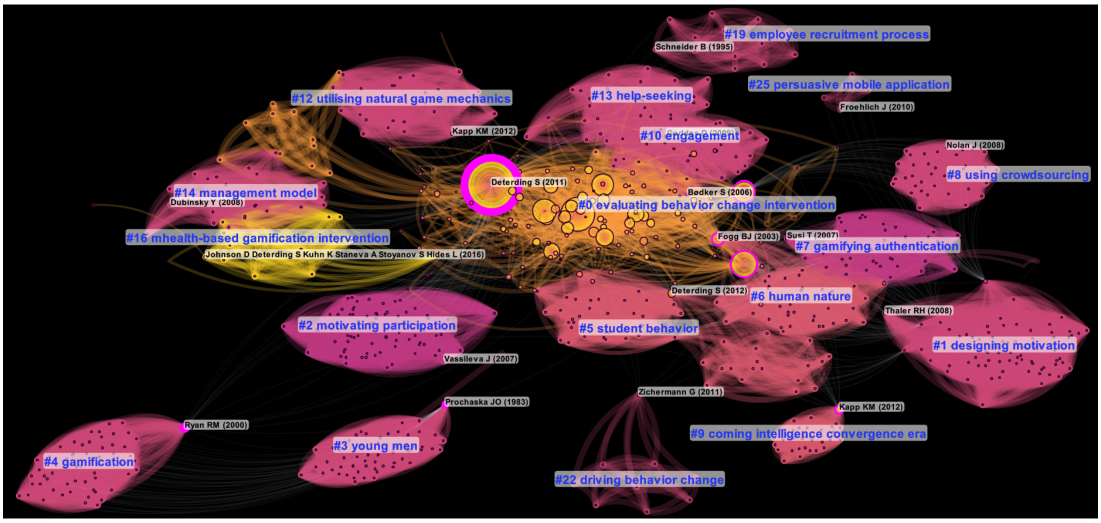


Figure 2.3: Cluster view of the document co-citation analysis (DCA) generated using CiteSpace Version 5.8.R3. Modularity Q = 0.8561; average silhouette = 0.9649. Colored shades indicate the passage of time, from past (purplish) to the present time (yellowish). Colored tree rings refer to the nodes with high betweenness centrality (purple tree rings) and burstness (red tree rings).

Since some of the clusters identified through the DCA were not substantial enough, we chose to present in detail only the 7 major clusters generated through the "generate narrative" command. In addition, following Aryadoust et al. (2021), emphasizing the importance of considering clusters counter-label based on the evaluation of the documents within each cluster, we chose to rename cluster #5 "Evaluating behavior change intervention" in "Student behavior" and cluster #6 "Human nature" in "Fun belief" (a detailed description of the rationale for these changes will be performed below, during the description of the individual clusters).

Among the 7 major clusters, the duration ranged from 25 to 58 years, presenting several overlaps. Cluster #2 = "Motivating participation" has the higher duration over time (58 years), followed by cluster #6 = "Fun belief" (previously "Human nature") (64 years) and cluster #3 = "Young men" (51 years). Cluster #4 = "Gamification" has the smallest duration over time (25 years). By looking at publication mean year, cluster #2 seems to be the oldest one (mean year = 2003), while cluster #0 = "Evaluating behavior change intervention" (mean year = 2010), cluster #4 (mean year = 2011) and cluster #6 (mean year = 2010) the most recent ones.

However, it is worth noting that the mean year of publication of some clusters may have been largely biased by older publications. For example, cluster #2's mean year of publication is 2003, but this cluster is among those with a longer duration (58 years, from 1954 to 2012) and bigger size (79). Since mean is extremely affected by extreme values, older papers, even if few, may have drastically lowered the publication mean year reported by CiteSpace. For this reason, we chose to group and describe clusters by sorting them by size (i.e., by the number of cited documents in each cluster) (Table 2.7), rather than by year.

Below, the 7 major clusters found through the "generate narratives" command are presented and described.

Cluster #0, "Evaluating behavior change intervention" is the biggest one in size, but also the less homogenous (silhouette = 0.861) among the 7 major clusters. It contains 142 cited documents written between 1977 and 2019 (mean year = 2010), some of which contributed to the definition and development of the gamification domain. This cluster focuses mostly on describing the strengths and the weaknesses of gamification applied to behavioral change. It also collects early and recently cited documents on gamification and its application to different domains (mainly education, health and environmental awareness), from both a theoretical and applied point of view. The cluster name is in line with the scope of interest of this review and suggests that the citing documents within this cluster refer mainly to the evaluation of behavior change intervention programs. Although the citing documents do not always directly address gamification, it is interesting to note that the three cited documents with the highest citation frequency explicitly refer to gamification. The citing document with the greatest coverage (i.e., the one that cited more references in the cluster) is Trinidad, Ruiz, and Calderón (2021) with 18 citations, while the cited documents with the highest citation frequency are Deterding, Dixon, et al. (2011) with a frequency of 210, Hamari et al. (2014) with a frequency of 144, Seaborn and Fels (2015) with a frequency of 73 and McGonigal (2011) with a frequency of 61.

Cluster #1, "Designing motivation" and **Cluster #2**, "Motivating participation" both have a size of 79 cited references, with the former ranging from 1968 to 2011 and the latter ranging from 1954 to 2012. Both clusters contain cited documents that attempt to explore the importance of motivation in behavior change interventions, with more theoretical and design-oriented documents in cluster #1 and more applied documents in cluster #2. It is interesting to note that most of the cited documents in these clusters are before 2011 (when gamification was mentioned for the first time) and do not refer explicitly to gamification. Moreover, these two clusters draw their name from a few citing documents (4 citing documents for cluster #1 and only one for cluster #2) suggesting that the literature tends to remain anchored to few theoretical papers. In cluster #1 the citing document with the greatest coverage is Nakajima and Lehdonvirta (2013) with 79 citations, while the three most influential cited references are Leonard (2008) with a frequency of 5 citations, Fogg (2002) with a frequency of 4 and Reeves and Read (2009) with a frequency of 4. In cluster #2, all cited documents show a citation frequency equal to one, as the cluster consists of only one citing document, namely Vassileva (2012).

Cluster #3, "Young men", contains 61 cited references written between 1961 and 2012 (mean year = 2006). It is mainly composed of cited documents that aim to promote gamified physical activity programs in young participants. Like cluster #2, cluster #3 consists of a single citing document, namely Ahola et al. (2013), which therefore contains all the cited documents with a citation frequency of one. The fact that only one citing document makes up the cluster suggests that Ahola et al. (2013) is a document of relevance in the literature, while the cluster name suggests that gamified programs often have young participants as their target audience.

Cluster #4, “Gamification”, has a size of 57 cited documents from 1988 to 2013 and it is the most recent one according to the mean year (2011). Among the other clusters, it is also the shortest in terms of duration (25 years). It collects several cited documents involving behavioral change protocols, gamification, and gamification applied to behavioral changes. The cluster consists of only two citing documents, namely Schoech, Boyas, Black, and Elias-Lambert (2013) and Putz and Treiblmaier (2015), both of which are theoretical documents on the application of gamification to different domains. The only cited document that emerges in this cluster is (Ryan & Deci, 2000) with a citation frequency of 5, while the remaining show a frequency of one.

Cluster #5, “Student behavior” (previously “Evaluating behavior change intervention”), has a size of 55 cited references, written from 1980 to 2012 (mean year = 2007). Cited documents included in this cluster involve mostly gamification applied to education programs or theoretical works on gamification in general. By looking at the citing and cited documents, it was noted that the cluster name “Evaluating behavior change intervention” was redundant (as it was identical to cluster #0 name) and not precise, while the name “Student behavior” allowed for more accurate capture of the cluster core focus, as several documents involved exploring students behavior. Thus, the cluster name was changed. Here, the citing document with the greatest coverage is Rao (2013) with 36 citations, while the cited documents with the highest citation frequency are Deterding (2012) with a frequency of 17 and Baranowski, Buday, Thompson, and Baranowski (2008) with a frequency of 10.

Cluster #6, “Fun belief” (previously “Human nature”), contains 44 cited references written between 1955 and 2012 (mean year = 2010). Most of its cited documents are related to persuasive techniques, gaming, and gamification. The cluster name was changed to “Fun belief” since “Human nature” was vague and not really informative of the cluster content, while “Fun belief” represented the cluster more accurately. Its citing document with the greatest coverage is Whitson (2013) with 42 citations, while the most influential cited documents are Fogg (2002) with a citation frequency of 22, and Tekinbas and Zimmerman (2003) with a citation frequency of 5.

Cluster ID	Cluster label	Size	Silhouette	Mean (year)	Begin	End	Duration
0	Evaluating behavior change intervention	142	0.861	2010	1977	2019	42
1	Designing motivation	79	0.995	2008	1968	2011	43
2	Motivating participation	79	0.995	2003	1954	2012	58
3	Young men	61	1	2006	1961	2012	51
4	Gamification	57	0.988	2011	1988	2013	25
5	Student behavior	55	0.955	2007	1980	2012	32
6	Fun belief	44	0.968	2010	1955	2012	57

Table 2.1: Cluster labels computed via document co-citation analysis (DCA).

Through DCA, we computed the major 25 citation bursts; Table 2.2 reports the strongest 10. The publication of Deterding, Sicart, Nacke, O'Hara, and Dixon (2011) has the strongest burst of the network, with a strength of 16.41, and it was the burst with the longest duration over time (4 years) along with the publication of Zichermann and Cunningham (2011). The oldest bursts in the network started in 2014 (McGonigal, 2011; Zichermann & Cunningham, 2011), while the newest started in 2020 (Koivisto & Hamari, 2019b; Sailer, Hense, Mayr, & Mandl, 2017). Interestingly, all the main citation bursts are contained within the cluster #0, suggesting that this cluster collects all the documents that have attracted research attention the most.

Reference	Burst strength	Burst begin	Burst end	Centrality	Sigma	Cluster ID
Deterding, Dixon, et al. (2011)	16.41	2015	2019	0.72	7055.67	0
Koivisto and Hamari (2019b)	14.91	2020	2022	0.00	1.01	0
Hamari et al. (2014)	12.83	2016	2019	0.04	1.64	0
Seaborn and Fels (2015)	12.21	2019	2022	0.01	1.14	0
Zichermann and Cunningham (2011)	11.05	2014	2018	0.17	5.90	0
McGonigal (2011)	9.67	2014	2017	0.17	4.63	0
D. Johnson et al. (2016)	9.66	2019	2022	0.00	1.02	0
Hamari (2017)	8.61	2019	2022	0.00	1.03	0
Sailer et al. (2017)	8.14	2020	2022	0.00	1.02	0
Huotari and Hamari (2017)	8.02	2019	2022	0.00	1.02	0

Table 2.2: List of the top 10 documents for burst strength, estimated via document co-citation analysis (DCA).

Among our network, the publication of Deterding, Dixon, et al. (2011) has a sigma value higher than the other publications (7055.67), followed by Zichermann and Cunningham (2011) (5.90), McGonigal (2011) (4.63), and Fogg (2002) (2.30). The other values do not differ so much from 1. Instead, regarding the values for the betweenness centrality, publications range from 0 to 0.72 (Table 2.3). The highest value is the publication of Deterding, Dixon, et al. (2011).

Reference	Centrality	Cluster ID
Deterding, Dixon, et al. (2011)	0.72	0
Ryan and Deci (2000)	0.37	0
Prochaska and DiClemente (1983)	0.21	3
McGonigal (2011)	0.17	0
Zichermann and Cunningham (2011)	0.17	0

Table 2.3: Top 5 documents for betweenness centrality via document co-citation analysis (DCA).

2.3.2.2 Author Co-citation Analysis (ACA)

By analyzing author co-citation analysis, we can find influential authors in the field of gamification applied to behavior change. The magnitude of each node represents the author's citation counts and the length between two nodes represents the two author co-citation frequency. A bigger node suggests an important author for the network; a smaller distance between two nodes detect a high authors' co-citation frequency, and a closer research topic and direction (X. Chen & Liu, 2020). The network obtained through the ACA contains 857 authors and 20,433 collaboration links (Figure 2.4), showing a modularity Q index of 0.7988, and an average silhouette metric of 0.936. The network has a wide range of collaborations, which reflects the interdisciplinary nature of gamification and the several domains in which behavior change can be utilized.

Table 2.9 shows the top 10 authors according to citation frequency. The largest node represents the author Deterding S with a citation frequency of 401 and a centrality value of 0.61, followed by Hamari J with a citation frequency of 289 and a centrality of 0.01. The third author ordered by citation frequency is [Anonymous] which is not of interest because this might be due to missing names of the authors of some publications in the Scopus dataset or due to loss of information during the conversion of the Scopus files. Since it is considered an outlier, we decided to ignore it.

Authors	Frequency	Centrality
Deterding S	401	0.61
Hamari J	289	0.01
[Anonymous]	166	0.13
Ryan RM	131	0.04
Deci EL	119	0.05
Zichermann G	118	0.17
Werbach K	102	0.01
Fogg BJ	96	0.17
Seaborn K	95	0.01
Huotari K	89	0.03

Table 2.4: Top 10 cited authors ordered by citation frequency via author co-citation analysis (ACA).

Table 2.10 presents the top 10 ranking authors according to burstness computed via ACA. The author with the highest burst strength was Koivisto J (strength = 14.42, centrality = 0.01), whose burstness started in 2020 and ended in 2022, followed by Morschheuser B (strength = 10.29, centrality = 0.00, 2019–2022) and Johnson D (strength = 10.05, centrality 0.00, 2019–2022). However, the increasing trend in citations for the cited authors listed in Table 2.10 ended in 2022 (except Oinas-Kukkonen H, which burst ended in 2019), which is the year this review was written, suggesting that their burst strength is likely to continue for the foreseeable future.

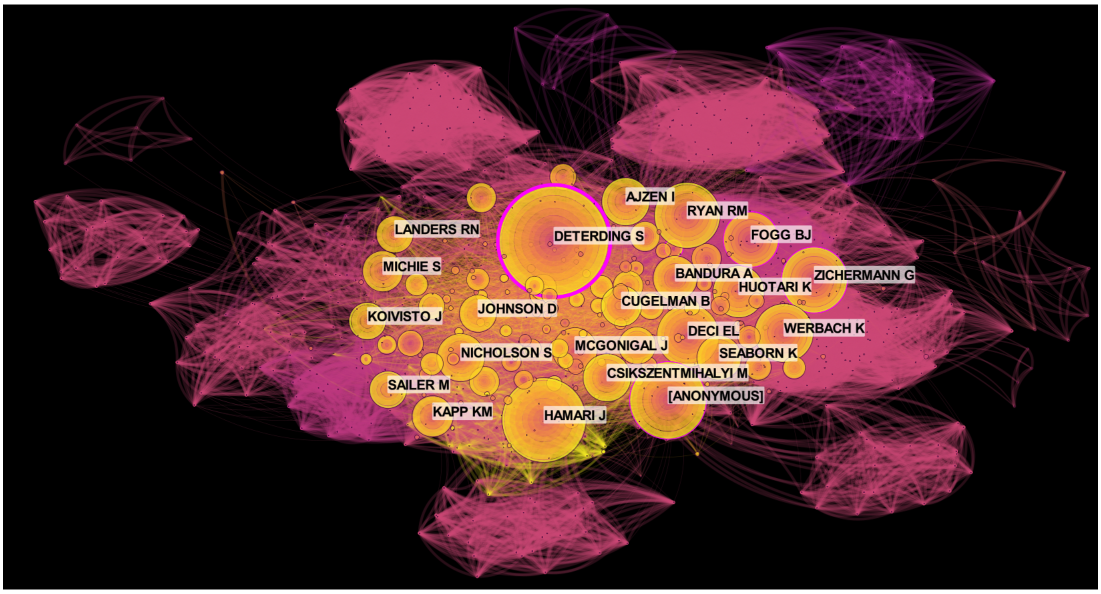


Figure 2.4: A visualization of the author co-citation analysis generated using CiteSpace Version 5.8.R3. Modularity Q = 0.7988; average silhouette = 0.936.

Cited authors	Burst strength	Begin	End	Span	Centrality	Frequency
Koivisto J	14.42	2020	2022	2	0.00	83
Morschheuser B	10.29	2019	2022	3	0.00	36
Johnson D	10.05	2019	2022	3	0.00	74
Landers RN	10.04	2020	2022	2	0.00	67
Sailer M	9.24	2020	2022	2	0.00	62
Sardi L	8.84	2019	2022	3	0.00	35
Dichev C	8.34	2020	2022	2	0.00	19
Oinas-Kukkonen H	7.82	2017	2019	2	0.00	41
Cohen J	7.20	2020	2022	2	0.01	28
Edwards EA	7.09	2018	2022	4	0.00	30

Table 2.5: Top 10 author bursts computed via author co-citation analysis (ACA).

2.3.2.3 Keyword Co-occurrence Analysis

The keyword co-occurrence analysis is an important aid in explaining the structure of scientific knowledge and discovering research trends (Su, Li, & Kang, 2019). The detection of keywords refers to the words that are frequently used or that are used in a shorter period. The keyword co-occurrence analysis provided a network with 325 nodes and 3,965 links, showing a modularity Q index of 0.5087 and a weighted mean silhouette of 0.8698.

Table 2.11 lists the top 10 keywords with the strongest bursts. In terms of burst strength, the top ranked keyword is "education" with a burst of 9.13, followed by "health" with a burst of 8.12, "behavioural change" with a burst of 7.42, "behavioral change" with a burst of 6.93 and "serious game" with a burst of 6.70. "Persuasive technology", "design", "sustainable development", and "intrinsic motivation" have the earliest burst begin, while "major clinical study", "controlled study", "randomized controlled trial", "sustainability" and "article" have the latest burst begin, which is over in 2022 because it was the date of our search. It is legitimate to think that it could continue in the future years, increasing the duration time.

According to the beginning and the end of the burst, we can discover the change over time for the topics in the field. In the early stages, "Persuasive technology", "design", "sustainable development", and "intrinsic motivation" are the mainstream trends, followed by "health", "health promotion", "game based learning", "education", "human computer interaction", and "video game". After them, "behavioral change", "behavioural change", "serious game", "energy conservation", "health behavior", and "psychology" have become the trends in the literature. However, "major clinical study", "controlled study", "randomized controlled trial", and "sustainability" have become the research frontier in recent years.

Keywords	Strength	Begin	End	Duration
education	9.13	2015	2017	2
health	8.12	2014	2017	3
behavioural change	7.42	2016	2019	3
behavioral change	6.93	2017	2019	2
serious game	6.70	2016	2018	2
major clinical study	6.54	2019	2022	3
energy conservation	6.35	2017	2018	1
sustainability	6.06	2020	2022	2
human computer interaction	5.99	2015	2017	2
design	5.79	2013	2016	3

Table 2.6: Top 10 keyword bursts computed via keyword analysis.

2.3.3 Discussion

In this section we answer the research questions we initially defined. Our aim is to provide a structured and systematic description of gamification's literature applied to behavior change. Thus, we outlined the main outcomes we found during the analysis and we propose some directions for future studies. In each section, a single research question is discussed based on the findings described in the results section.

2.3.3.1 What are the most influential documents in the gamification for behavior change field?

To answer this question, we focused on DCA only, since it contains all the information needed to respond. Here, to extrapolate the most influential documents we followed two different ways: (i) on the one hand we looked at documents with higher burst strength, betweenness centrality and sigma values (Tables 2.2, 2.3), as burstness reflect a sudden research interest during a limited period of time, betweenness centrality reflect the influence on the network and sigma is a combination of these two measures (see the Methods section for a detailed description); (ii) on the other hand we took the top cited and citing documents contained in the clusters with higher size (Table 2.7).

Considering burst strength, betweenness centrality, and sigma values, the paper that has attracted the most research attention and that has influenced the literature network is definitely Deterding, Dixon, et al. (2011), with the first place in both burst strength (16.41) and betweenness centrality (centrality = 0.72), and with a sigma value significantly higher than the other documents (sigma = 7055.67). This document is the first paper that defined the concept of "gamification", describing the design of a typical gamified paradigm and focusing on gamification's historical origins and applications. According to our review, it has been a cornerstone paper in the field of gamification applied to behavioral change and it stands as a guideline for subsequent gamification works. However, its citation peak ended in 2019 (burst started in 2015 and ended in 2019), suggesting that Deterding, Dixon, et al. (2011) has been a popular document in this field for some years, but has recently been overlooked.

In contrast, citations bursts of Koivisto and Hamari (2019b) and Seaborn and Fels (2015), ranked second (14.91) and fourth (12.21) in terms of burst strength respectively, began in a relatively more recent year (2020 and 2019) and may not have ended yet (bursts ended in 2022, which is the year this review was written). In detail, Koivisto and Hamari (2019b) consists of a large systematic review on 819 empirical studies that employed gamification, while Seaborn and Fels (2015) aims to conduct an impressive review outlining the theoretical understandings of gamification and comparing gamification with other methodologies (such as alternate reality games, games with a purpose, and gameful design). Finally, at the third place in terms of burst strength we found Hamari et al. (2014), with a burst of 12.83 (burst began in 2016 and ended in 2019). This document consists of a large review on the effectiveness of gamification when applied to different domains.

Looking at betweenness centrality, in the first place we find (as already reported) Deterding, Dixon, et al. (2011) (centrality = 0.72), followed by Ryan and Deci (2000) (centrality = 0.37), next Prochaska and DiClemente (1983) (centrality = 0.21), McGonigal (2011) (centrality = 0.17), and finally Zichermann and Cunningham (2011) (centrality = 0.17). Thus, Deterding, Dixon, et al. (2011) is also the document with the highest influence on the network of documents selected for this review. Interestingly, only Prochaska and DiClemente (1983) describes an empirical study, while Ryan and Deci (2000) is a review, and both McGonigal (2011) and Zichermann and Cunningham (2011) are books. This might have affected their top rankings in betweenness centrality: since books and reviews (generally) contain more information than scientific papers, they are very likely to be cited more and in more domains. Moreover, Zichermann and Cunningham (2011) and McGonigal (2011) show the second (5.90) and third (4.63) highest sigma values respectively, suggesting that they are documents with a high influential potential on the topic (C. Chen et al. (2009, 2010)).

Considering the top cited and citing documents within the four largest size clusters, we find Deterding, Dixon, et al. (2011) and Hamari et al. (2014) as the most cited documents in cluster #0 (i.e., the cluster with the largest size), with 210 citations the former and 114 the latter, and Trinidad, Ruiz, and Calderón (2021) as the citing document with the largest coverage (18). These documents are respectively a conference paper, a review, and a bibliometric analysis, focusing on both theoretical and applied aspects of gamification. Therefore, it is not surprising that they have been grouped in cluster #0, which collects the most important documents concerning general theoretical and applicative information on the gamification domain applied to behavior change. In cluster #1, Leonard (2008) and Fogg (2002) are the cited documents with the higher citation frequency, 5 and 4 respectively, while Nakajima and Lehdonvirta (2013) is the citing document with the greatest coverage (79). Leonard (2008) and Fogg (2002) are both before 2011 and they don't address gamification directly. They are respectively a comment on a book about nudging and a book chapter on persuasive techniques. On the other hand, Nakajima and Lehdonvirta (2013) describes four case studies that employed gamified persuasive technologies for behavior change.

Cluster #2 and #3 are both composed of a single citing document, thus each cited document within each cluster has a citation frequency equal to one. This fact, in our opinion, points out that the literature in the domain of gamification applied to behavior change sometimes tends to remain anchored to few theoretical papers and struggle to build a comprehensive network. In cluster #2, the citing document that collects all the cited documents of the cluster is Vassileva (2012). This paper describes several different approaches to motivate people engaging in behavioral change programs. In cluster #3, Ahola et al. (2013) is the only citing document: it consists of a paper describing a massive study in which 1,500 young men undergo a 6-months online gamified activation method to change their behavior.

Interestingly, the most important documents in the field of gamification applied to behavior change are rarely papers about original experimental studies (apart from a few studies, such as Ahola et al. (2013) and Prochaska and DiClemente (1983)). This seems to suggest that this field possesses some strong theoretical works (mainly books and reviews), but lacks corroborated experimental support. Future studies should focus more on this second aspect.

2.3.3.2 Who are the most influential authors contributing to the research of behavior change?

To address this research question, we rely on the results of the ACA. Tables 2.9 and 2.10 give us an overview of the most influential authors according to citation frequency and burst strength.

Considering burst strength, the author who attracted the most research attention over a period of time is Koivisto J (burst strength = 14.42). This author's burst is probably linked to the review Koivisto and Hamari (2019b), which is also the second document for burst strength. Since the burst ends in 2022 (date of the review), it is legitimate to think that it could continue in the future years, increasing the duration time. This can mean Koivisto J is helping in shaping the recent and future parts of the literature. In the second place in terms of burst strength, we find Morschheuser B (burst strength = 10.29). His documents deal with gamification design (Morschheuser, Hamari, Koivisto, & Maedche, 2017; Morschheuser, Hamari, Werder, & Abe, 2017; Morschheuser, Hassan, Werder, & Hamari, 2018). His burst strength started in 2019 and ended in 2022. Also in this case we can think that the burst continues beyond the date. In the third and fourth place in terms of burst strength, we find Johnson D (burst strength = 10.05), whose most cited documents (C. Johnson et al., 2016; D. Johnson et al., 2016; D. Johnson, Horton, Mulcahy, & Foth, 2017) deals with the use of gamified solutions to motivate users to adopt behaviors related to health and well-being, and to the reduction of domestic energy consumption, and Landers RN (burst strength = 10.04), whose publications deal with gamification theory (Landers, 2014; Landers, Auer, Collmus, & Armstrong, 2018), gamification use (Armstrong & Landers, 2018), and several analysis on gamification elements (Landers, Bauer, & Callan, 2017; Landers & Landers, 2014). Looking at the timeline of the most influential documents, the authors with the biggest burst strength are the most recent.

Interestingly, exploring the research fields of the most influential authors for burst strength, within the ones with the higher burst value (Koivisto J, Morschheuser B, Johnson D , and Landers RN), only Johnson D directly applied gamification to produce a behavioral change. Sorting the burst strength by the beginning year of burst, most of the authors with an old burst (Bartle L, Marczewski A, Zichermann G, McGonigal J, and Farzan R) deal with user motivation and participation, and gamification definition. Hence, more recent authors for burst (Oinas-Kukkonen H, Nacke LE, and Edwards EA) deal with gamification definition, personalization, and application in the health domain. The most recent authors according to burst begin (Morschheuser B, Johnson D, Sardi L, Koivisto J, Landers RN, Dichev C, and Cohen J) deal with the issues in gamification development, the need of novel designing methods, and the application of gamified solutions to produce behavioral changes in users.

Considering the citation frequency, the most influential authors are Deterding S (citation frequency = 401) and Hamari J (citation frequency = 289). This result is not surprising since a great amount of the most important documents in the gamification domains are written by these two authors (Deterding, 2012; Deterding, Dixon, et al., 2011; Deterding, Sicart, et al., 2011; Hamari, 2013, 2017; Hamari & Koivisto, 2013; Hamari et al., 2014; Hamari & Tuunanen, 2014; Koivisto & Hamari, 2019b).

Overall, the ACA results suggest that initially, the structure of gamification's literature applied to behavior change has been guided by the documents of Deterding S, Hamari J, Marczewski A, Zichermann G, McGonigal J, Farzan R, and Bartle L, resulting in an initial cohesive structure composed of theoretical documents dealing with persuasion, design and gamification definition. Next, according to the newest bursts, the current structure of gamification's literature applied to behavior change is divided into two parts depending on the main subject of the authors: (i) Morschheuser B, and

Koivisto J deal with the need to question the current application and design methodologies, hence finding new solutions (Koivisto & Hamari, 2019b; Morschheuser, Hamari, Werder, & Abe, 2017; Morschheuser et al., 2018); (ii) Johnson D, Sardi L, Cohen J, and Landers RN deal with a practical application of gamification to promote behavioral changes in users (D. Johnson et al., 2016, 2017; Sardi, Idri, & Fernández-Alemán, 2017).

2.3.3.3 How have research trends in gamification applied to behavior modification changed over time?

To answer this question, we rely on an overview of the keywords change over time and on the DCA analysis. Examining the keywords' burst strength begin year (Table 2.11) and the major clusters' mean years (Table 2.7), we managed to extrapolate a timeline of the research trends in the gamification for behavior change domain. From this analysis, it seems clear that the researchers' interest has changed over time.

According to keywords, the first trends that appeared in the field of gamification applied to behavioral change are "design" (begin year = 2013), "health" (2014), "human computer interaction" (2015), and "education" (2015). This seems to suggest that the first research trends were linked to a general design stage, mainly involving health and education domains. Hence, the trend has changed, showing interest in "serious games" (2016) and "behavioral change" - which appears twice: "behavioural change" (2016) and "behavioral change" (2017) -, thus suggesting that research attention shifted from a general theoretical design stage toward the study of gamified procedures for behavior change. Finally, the last trends are related to environmental awareness ("energy conservation", begin year = 2017, and "sustainability", begin year = 2020) and clinical disorders ("major clinical study", begin year = 2019), suggesting that these trends are the most recent in the field of gamification for behavior change. Interestingly, the burst's ending year of "major clinical study" and "sustainability" is 2022, reflecting the fact that the bursts may still be ongoing.

According to the DCA analysis, the oldest clusters are "motivating participation" (mean year = 2003), "young men" (2006), "student behavior" (2007), and "designing motivation" (2008). Considering the labels of these clusters and their content (in terms of cited and citing documents), it seems that the research interest in the behavior change domain has been initially focused on the study of motivational intervention designs targeting students or young subjects. On the other hand, the most recent clusters (according to our review) are "evaluating behavior change intervention" (2010), "fun belief" (2010), and "gamification" (2011), thus clusters that collect documents on behavior change interventions (mostly gamified) and both theoretical and applied studies on gamification. This suggests that, only in recent times, research has focused on studying proper gamified interventions.

Overall, these results seem to suggest that trends have changed considerably over time, first describing broad motivational intervention designs and then leading more and more resources in the direction of a unitary concept of gamification based on gamified interventions. Finally, according to the keywords analysis, it seems that the most recent trends involve gamified intervention for environmental awareness.

2.3.3.4 Limitations

When interpreting the results of this scientometric review, it is worth noting that there are some limitations to consider. First, only data from the Scopus database were used in this study; data from other databases such as WoS, PubMed, and PsyInfo were not used. Future studies could compare other databases to decide on the more comprehensive database to use. Second, as already observed by other authors (Carollo, Balagtas, Neoh, & Esposito, 2021; Carollo, Bonassi, et al., 2021), the scientometric approach of DCA depends on the quantitative patterns of citations and co-citations: hence, all the citations are treated the same way, leaving out the reason behind each citation. Third, the impact of recent influential documents might have been underestimated, causing a bias towards the old ones due to their longer lifetime. Lastly, only the names of the first authors were used in the co-citation analyses performed in this study; hence, the co-citation analysis may yield different results if all the author names were made available.

2.3.4 Study conclusion

Gamification is facing continuous growth in disparate application contexts (e.g. education, training, health, and so forth), especially in those that promote a positive behavior change (Adrián & Elena, 2019). Indeed, gaming, as a motivating and engaging activity, makes it easier to convince people to break their bad habits and change their behavior.

This study analyzed research works on gamification to promote behavior change or positive behaviors, based on publications from 2011 to 2022 available in the Scopus database. It reveals that from a small number of publications that first appeared in 2011 and 2012, the number of works related to behavior change has exponentially grown and that the application areas are many. We performed co-citation analysis to identify the most influential documents, authors, and keywords, and how the documents are gathered in clusters to represent the scientific domains within the available literature, and we investigated the trends change over time. Overall, what emerges most is that the research interest has changed slightly over time. In the beginning, it has been anchored to those keywords, authors, and documents related to the self-determination theory, and methods for designing gamification as a persuasive and motivational tool. According to several recent reviews in the literature (Koivisto & Hamari, 2019b; Seaborn & Fels, 2015), the failure to promote a standard guideline, and the lack in employing adequate methodological rigor (such as sample size selection and controlled experimental research methods) has led to numerous inconsistent results with the gamification use. Hence, the research interest appears to have spread into two main areas to solve these problems, moving away from the first research topics: (i) the research for new solutions and new design methods, and (ii) the application of gamification for promoting environmental awareness, sustainability and well-being behaviors with greater methodological rigor.

In conclusion, the results of this study suggest that, as in other scientific areas (Chambers & Tzavella, 2022; Foster & Deardorff, 2017), and in line with other gamification domains (Trinidad, Ruiz, & Calderón, 2021), we expect that the use of gamification for behavior change will be supported by documents aimed at suggesting new and standardized procedures for the gamification design, and documents promoting an adequate methodological rigor. It may be useful in the future to conduct scientometric studies in specific fields related to behavior change (i.e. health and well-being, environmental awareness, and sustainability). This may provide in-depth information regarding the status of gamification for providing behavior change in various fields. We hope that the findings of the present study will lead to a better understanding of the topic we presented.

2.4 Gamification design frameworks: A scientometric analysis

As above-mentioned, this section refers to a study entitled "Lost in Gamification Design: A Scientometric Analysis" (Bassanelli, Gini, Bonetti, et al., 2024). The rationale behind this study is related to the findings of our previous scientometric study and to the recent relevant literature reviews (for example Koivisto and Hamari (2019b); Mora, Riera, González, and Arnedo-Moreno (2017); Sailer et al. (2017)), that present the relevant issues related to the design of gameful systems. As mentioned clearly by Mora et al. (2017), a large number of frameworks for the design of gameful systems can be found in the literature. However, most of these frameworks have been developed ad-hoc for business or a single case study, so no validation of the models has been done; moreover, these frameworks exhibit high heterogeneity, especially for theoretical concepts, and present a lack of personalization components.

With this study, we want to describe the major trends and changes in this literature in a systematic way, focusing on the most important documents, authors, and keywords related to gameful systems design. To achieve this goal, we have formulated the following research questions:

RQ B.1: *What are the most influential documents?*

RQ B.2: *Who are the most influential authors?*

RQ B.3: *How have research trends changed over time?*

2.4.1 Methods

The analyses were based on 1,751 publications about gamification design frameworks and models published between January 1, 2012, and December 16, 2022. The Scopus database was used to download 66,880 unique references. The time range of publications depended uniquely on Scopus' availability, and no a priori temporal exclusion criteria were applied. Out of a group of 1,826 documents, we discarded the results that were not written in English and the duplicates, leaving us with a sample of 1,751. The search code used was `\(TITLE-ABS-KEY (gamif*) AND TITLE-ABS-KEY (design AND framework) OR TITLE-ABS-KEY (design AND model))"`. The data from the Scopus database were converted to a CiteSpace-friendly format with the information related to each of the 1,751 publications retrieved. At this point, we used the CiteSpace software (version 6.1.R4) to analyze the data. Of the total references cited, 64,109 of the 66,880 (95%) were considered valid. Due to mistakes in the data that CiteSpace cannot fix, a small number of references were lost. This percentage of unprocessed references can be considered a negligible loss of data (C. Chen, 2016). Using CiteSpace, we analyzed networks without setting any time limits. We divided the time frame into annual slices and evaluated three different criteria for selecting nodes in the networks: Top N, Top N%, and g-index. We compared networks created using the Top N method with N set to 25, the Top N% method with N set to 5 and 10, and the G-index method with a scaling factor of 10 and 25.

In summary, we chose to use the networks generated with the Top N and a scaling factor of 25 for our Document co-citation analysis (DCA), author co-citation analysis (ACA), and keyword co-occurrence analysis as they produced the best overall results in terms of structural metrics, number of nodes and links, and consistency in cluster structure for DCA. The chosen network for each analysis refers to the largest connected component (LCC), which is the largest sub-network where it is possible to reach any node from any other node (C. Chen, 2016). In this case, the LCC included 890 nodes (71% of the whole network).

2.4.2 Results

2.4.2.1 Document co-citation analysis

The DCA provided a network with 1,243 nodes and 24,911 links, showing a modularity Q index of 0.8614 and an average silhouette metric of 0.9369, suggesting that the network was sufficiently divisible into clusters (according to the Q index) and that each cluster was highly consistent (according to the silhouette index) (Figure 2.5).

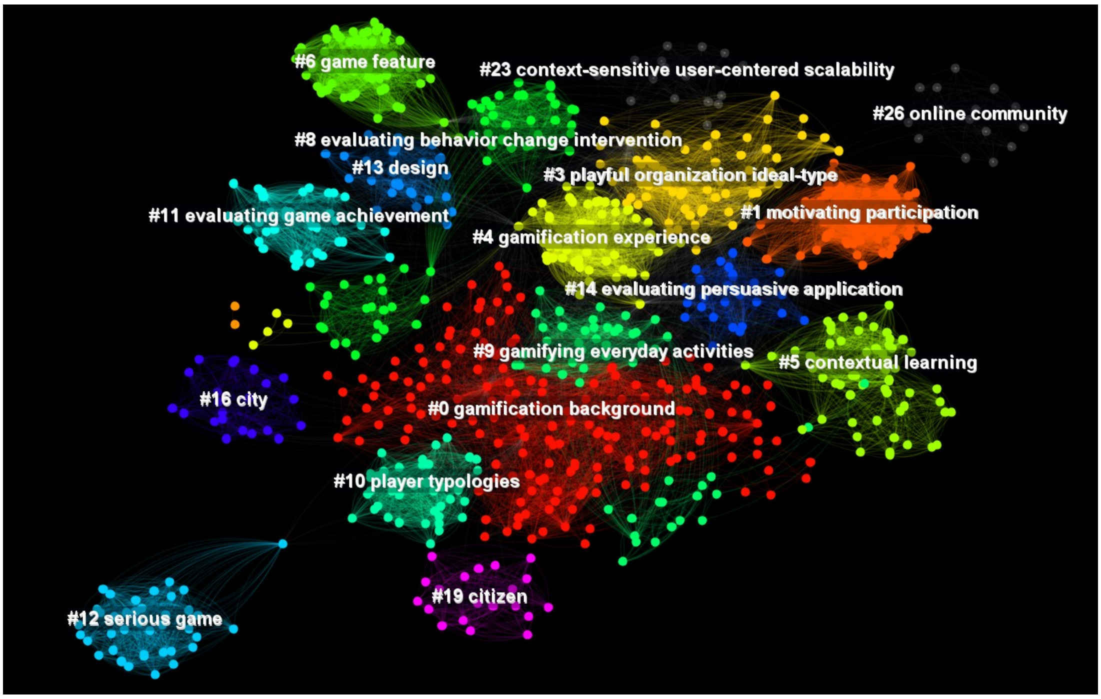


Figure 2.5: Cluster view of the document co-citation analysis (DCA) generated using CiteSpace Version 6.1.R6. Modularity Q = 0.8614; average silhouette = 0.9369. Colored shades indicate the different clusters.

The DCA resulted in 17 co-citation clusters, sorted from the largest in size (cluster #0 = “gamification background”, size = 183, silhouette = 0.759, mean year = 2007) to the smallest (cluster #26 = “online community”, size = 15, silhouette = 0.991, mean year = 2006). Since some of the clusters identified through the DCA were not substantial enough, others were too heterogeneous and others had been created by a few citing papers, we chose to present in detail only 4 of the major clusters generated through the ”generate narrative” command (Tab 2.7). Furthermore, in alignment with the recommendations of Aryadoust et al. (2021), which stress the significance of evaluating the contents within each cluster, and to differentiate them, we have elected to adopt a new nomenclature for cluster #3, replacing the term "gamification background" with "gamification experience" to identify the clusters more accurately.

Cluster ID	Cluster label	Size	Silhouette	Mean (year)	Begin	End	Duration
0	gamification background	227	0.759	2007	1938	2020	82
1	motivating participation	81	0.991	2003	1954	2012	58
3	gamification experience	66	0.955	2003	1967	2012	45
4	contextual learning	63	0.980	2002	1980	2012	32

Table 2.7: Cluster labels computed via document co-citation analysis (DCA).

Among the reported clusters, the duration ranged from 32 to 82 years, presenting a partial overlap. In this section, each of the clusters presented is described qualitatively, in descending order of their size. They are described in terms of citing papers and cited papers. Citing papers that contribute to each cluster can be characterized by their coverage (number of references in the cluster cited by that paper) and global citing score (GCS: the total number of citations of that paper according to Scopus).

Cluster #0, "gamification background", is the biggest one in size (size = 183, silhouette = 0.759). It comprises the following top citing papers: Teh, Schuff, Johnson, and Geddes (2013) (coverage = 29; GCS = 10), Herzig, Jugel, Momm, Ameling, and Schill (2013) (coverage = 22; GCS = 30), and Nah, Telaprolu, Rallapalli, and Venkata (2013) (coverage = 12; GCS = 59). These papers mainly focus on how to use motivational elements to increase user participation in theory, learning, and business. Specifically, Teh et al. (2013) report the need in gameful systems design to tap into intrinsic motivations by framing tasks in ways that evoke positive attitudes through challenge and friendly competition. The LLR label of the cluster, "gamification background", points to the importance of gamification background, hence how can gamification be a useful approach to motivate users in different contexts. As a result, the most cited papers in this cluster mainly present the theoretical aspect of gamification effectiveness (Deterding, Dixon, et al., 2011; McGonigal, 2011; Robson, Plangger, Kietzmann, McCarthy, & Pitt, 2015; Seaborn & Fels, 2015), how to correctly design gamified solutions (Huotari & Hamari, 2017; Werbach, Hunter, & Dixon, 2012; Zichermann & Cunningham, 2011), game design elements effectiveness to foster motivation (Dicheva et al., 2019; Sailer et al., 2017), the difficulty to use game design technique and what issues need to be curbed(Hamari et al., 2014; Koivisto & Hamari, 2019b), and how to implement game-based learning in the classroom (Domínguez et al., 2013; Hanus & Fox, 2015; Kapp, 2012).

Cluster #1, "motivating participation", is the second cluster in size (size = 81, silhouette = 0.991). It is mainly composed of two citing papers that focus on the use of gamification to increase users' motivation through persuasion. Specifically, Vassileva (2012) (coverage = 81; GCS = 145) present a paper with an overview of different approaches to motivate users to participate, and M. Oja and Riekki (2012) (coverage = 31; GCS = 7) present a framework that facilitates creation and evaluation of persuasive systems based on different theories to change users' behavior in a desirable way. The LLR label of the cluster, "motivating participation", refers to how important is to provide the right stimuli to foster users' motivation. As a result, the majority of the cited papers present some background about persuasion techniques (Cheng & Vassileva, 2005; Fogg, 2002), intrinsic motivation and self-determination theory (Deci et al., 1985; Reiss, 2004) and how gamification can help increase user motivation (Deterding, Dixon, et al., 2011; Deterding, Sicart, et al., 2011).

Cluster #3, "gamification experience" (size = 66, silhouette = 0.955), is composed of three main citing papers: Marache-Francisco and Brangier (2014) (coverage = 62; GCS = 16), M. Oja and Riekki (2012) (coverage = 29; GCS = 7), and Holman, Aguilar, and Fishman (2013) (coverage = 5; GCS = 37). The most citing document, "*The Gamification Experience: UXD with a Gamification Background*" (Marache-Francisco & Brangier, 2014), is a book chapter in which the authors provide a categorization of gamification elements through HCI design concepts. The authors refer to three different dimensions: sensory-motor modalities, motivation elements, and cognitive process support elements. The documents cited with the highest citation frequency deal mainly with game design, engagement and flow (Adams, 2014; Csikszentmihalyi & Csikzentmihaly, 1990; Reeves & Read, 2009; Schell, 2008)

Cluster #4, "contextual learning" (size = 63, silhouette = 0.980), comprehends a list of cited papers that refer to game-based learning, contextual learning, student behavior (Brown, Collins, & Duguid, 1989; Gee, 2003; Prensky, 2003; Weisberg, 2011). The citing document with the greatest coverage in this cluster are Li, Grossman, and

Fitzmaurice (2012) (coverage = 41; GCS = 170), and Bidarra, Figueiredo, and Natálio (2015) (coverage = 24; GCS = 20). These papers mainly focus on the use of gamification and a game-based approach for contextual learning.

Through DCA, we computed the major 10 citation bursts (see Table 2.8). The publications of Koivisto and Hamari (2019b), Seaborn and Fels (2015), and Sailer et al. (2017) have the strongest burst of the network, with a strength of 22.50, 22.29, and 20.96 respectively. The publication of Hamari et al. (2014) is the burst with the longest duration over time (6 years). The oldest burst in the network started in 2014 (McGonigal, 2011), while the newest started in 2019 (Koivisto & Hamari, 2019b; Sailer et al., 2017; Werbach et al., 2012). Among our network, the publications of McGonigal (2011), Kapp (2012), and Zichermann and Cunningham (2011) have a sigma value higher than the other publications (10.98, 3.31 and 2.01 respectively), followed by Deterding and colleagues (Deterding, Dixon, et al., 2011; Deterding, Sicart, et al., 2011) (1.86 and 1.68 respectively), Schell (2008) (1.64), and Reeves and Read (2009) (1.50). The other documents present lower values. Instead, regarding the values for the betweenness centrality, publications range from 0 to 0.21. The only document that stands apart is McGonigal (2011) with a betweenness centrality value of 0.21.

Reference	Frequency	Burst strength	Burst begin	Burst end	Centrality	Sigma	Cluster ID
Koivisto and Hamari (2019b)	57	22.50	2019	2022	0.00	1.02	0
Seaborn and Fels (2015)	111	22.29	2017	2022	0.01	1.25	0
Sailer et al. (2017)	62	20.96	2019	2022	0.00	1.07	0
Hamari et al. (2014)	52	13.33	2016	2022	0.01	1.09	0
Deterding, Dixon, et al. (2011)	120	12.66	2016	2020	0.05	1.86	0
McGonigal (2011)	50	12.56	2014	2016	0.21	10.98	3
Dicheva et al. (2019)	55	11.47	2017	2022	0.00	1.02	0
Werbach et al. (2012)	112	10.55	2019	2022	0.03	1.42	0
Kapp (2012)	97	10.38	2016	2020	0.12	3.31	0
Robson et al. (2015)	40	9.38	2017	2022	0.00	1.03	0

Table 2.8: List of the top 10 documents for burst strength, estimated via document co-citation analysis (DCA).

2.4.2.2 Author co-citation analysis

By analyzing author co-citation analysis, we can find influential authors in the field of gamification design frameworks. The network obtained through the ACA contains 344 nodes and 7,940 collaboration links, showing a modularity Q index of 0.5785, and

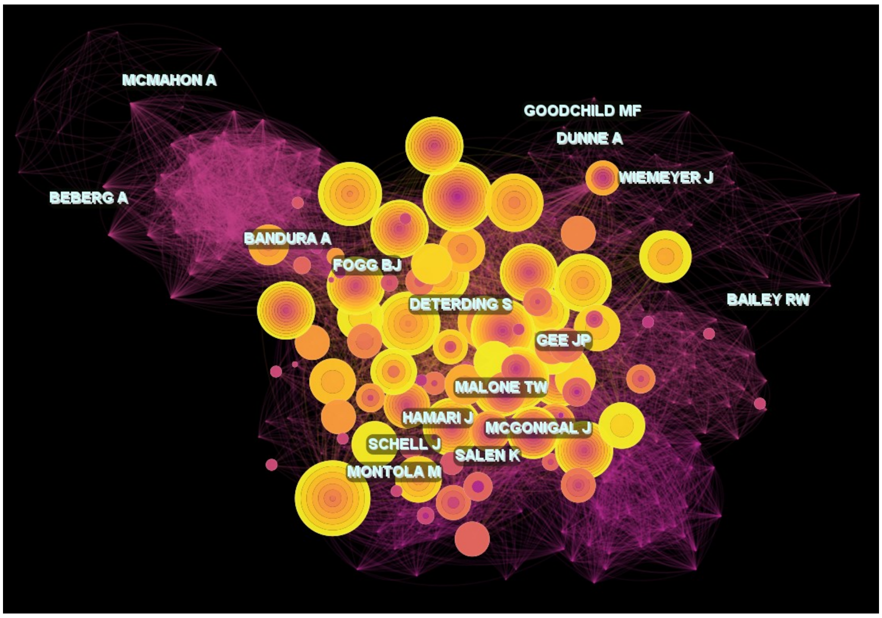


Figure 2.6: Overview of the author co-citation analysis (ACA) generated using CiteSpace Version 6.1.R6. Modularity Q = 0.5785; average silhouette = 0.8943. Colored rings refer to the nodes with high burstness.

an average silhouette metric of 0.8943 (Figure 2.6). The network has a wide range of collaborations, which reflects the interdisciplinary nature of gamification and the several domains in which it is applied.

Table 2.9 presents the top 10 authors according to the citation frequency. The largest node represents Deterding S. with a frequency of 647 and a centrality value of 0.45, followed by Hamari J. with a citation frequency of 399 and a centrality of 0.01, and Werbach K., with a citation frequency of 213 and a centrality value of 0.

Table 2.10 depicts the top 10 authors according to citation burst, or the number of citations for an author over a short time period. The author with the highest burst strength is Koivisto J. (strength = 17.34, centrality = 0.01), whose burstness started in 2020 and ended in 2022, followed by Sailer M. (strength = 16.47, centrality = 0.00) and Landers R. M. (strength = 14.24, centrality = 0.00). However, the citation burst of some authors ended in 2022, which is the year this review was written, suggesting that their burst strength is likely to continue for the foreseeable future.

Authors	Frequency	Centrality
Deterding S.	647	0.45
Hamari J.	399	0.01
Werbach K.	213	0.00
Zichermann G.	189	0.11
Ryan R. M.	189	0.03
Kapp K. M.	175	0.01
Csikszentmihalyi M.	159	0.10
Seaborn K.	155	0.00
Deci E. L.	150	0.07
Huotari K.	145	0.01

Table 2.9: Top 10 cited authors ordered by citation frequency via author co-citation analysis (ACA).

Cited authors	Burst strength	Begin	End	Span	Centrality	Frequency
Koivisto J.	17.34	2020	2022	2	0.01	111
Sailer M.	16.47	2020	2022	2	0.00	113
Landers R. N.	14.24	2019	2022	3	0.00	79
Gee J. P.	14.20	2014	2016	2	0.01	32
Dicehv C.	14.03	2020	2022	2	0.00	32
Malone T. W.	14.01	2011	2017	6	0.09	46
McGonigal J.	13.29	2013	2017	4	0.14	100
Mora A.	13.21	2018	2022	4	0.00	63
Schell J.	12.97	2013	2016	3	0.01	30
Mekler E. D.	12.21	2017	2019	2	0.00	32

Table 2.10: Top 10 author bursts computed via author co-citation analysis (ACA).

2.4.2.3 Keyword co-occurrence analysis

The keyword co-occurrence analysis is an important aid in explaining the structure of scientific knowledge and discovering research trends (Su et al., 2019). The detection of keywords refers to the words that are frequently used or that are used in a shorter period. The keyword co-occurrence analysis provided a network with 218 nodes and 2,365 links, showing a modularity Q index of 0.5305 and a weighted mean silhouette of 0.87.

Table 2.11 lists the top 10 keywords with the strongest bursts. In terms of burst strength, the top-ranked keyword is “education” with a burst of 15.64, followed by “software engineering” with a burst of 10.07, “teaching” with a burst of 8.15, “design” with a burst of 8.06 and "application program" with a burst of 7.87. "Design" has the earliest burst begin, while "application program", "survey", and "artificial intelligence" have the latest burst begin. The "survey" burst ends in 2022 because it was the date of our search. It is legitimate to think that it could continue in the future years, increasing the duration time.

Keywords	Strength	Begin	End	Duration
education	15.64	2015	2017	2
software engineering	10.07	2015	2017	2
teaching	8.15	2016	2018	2
design	8.06	2013	2015	2
application program	7.87	2019	2020	1
information use	7.78	2018	2019	1
survey	7.06	2019	2022	1
artificial intelligence	6.84	2019	2020	1
article	6.59	2018	2020	2
computer game	6.51	2017	2019	2

Table 2.11: Top 10 keyword bursts computed via keyword analysis.

2.4.3 Discussion

In this section, we answer the research questions that we defined at the beginning of Section 2.4. Our final aim is to provide a structured and systematic description of the literature related to the design frameworks developed and used in the gamification domain.

2.4.3.1 What are the most influential documents?

To answer this question, we focused on DCA only, since it contains all the information needed to respond. Our analysis is mainly based on two different approaches: (i) we looked at the documents with the higher sigma values, burst strength, and betweenness centrality since burstness reflects sudden research interest in a limited period of time, betweenness centrality reflects the influence on the network and sigma is a combination of these two measures (Bassanelli et al., 2022), and on the other hand, (ii) we considered the top cited and citing documents contained in the presented clusters.

Considering the three metrics and the frequency, the papers that have attracted the most research attention and that have influenced the literature network are definitely Seaborn and Fels (2015) (burst strength = 22.29, sigma = 1.25), Koivisto and Hamari (2019b) (burst strength = 22.50, sigma = 1.02), Sailer et al. (2017) (burst strength = 20.96, sigma = 1.07), and McGonigal (2011) (burst strength = 12.56, sigma = 10.98). In detail, Seaborn and Fels (2015) aim to examine the current theoretical underpinnings of gamification, as well as to compare and contrast gamification with related methodologies such as alternate reality games, games with a purpose, and gameful design. *"The rise of motivational information systems: A review of gamification research"* (Koivisto & Hamari, 2019b) consists of a large systematic review of 819 empirical studies that employed gamification which reports a critical overview of gamification application and a detailed future research agenda that suggests future directions for gamification research; *"How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction"* (Sailer et al., 2017) consists of a randomized controlled study in an online simulation environment with different configurations of game design elements. In *"Reality Is Broken: Why Games Make Us Better and How They Can Change the World"* (McGonigal, 2011), the author argues that games have the power to make people happier, more productive, more creative and that they can be used to solve real-world problems. Moreover, the author suggests that games are inherently more engaging and motivating than other forms of media and that people are naturally drawn to the kind of challenges and feedback that games provide. According to the sigma value, this work can be considered a milestone that significantly impacted shaping the literature on gamification design frameworks (C. Chen et al., 2009, 2010). However, the citation peak (or the burstness) of the cited papers ended in 2016 (burst started in 2014 and ended in 2016), suggesting that this document had been influential for some years for design frameworks in gamification, but has recently been overlooked. The burst of the other above-mentioned papers began in a relatively more recent year (2017 and 2019) and may not have ended yet (bursts ended in 2022, which is the year this review was written). Considering the citation frequencies, the most important documents are Deterding, Dixon, et al. (2011) (citation frequency = 120), Werbach et al. (2012) (citation frequency = 112) and Seaborn and Fels (2015) (citation frequency = 111). Interestingly, these documents are also considered among the most important according to the metrics.

All the named documents, except McGonigal (2011), are representative of cluster #0, which collects the most important documents of both theoretical and practical information on gamification theory and background. Due to this, the cluster results are broader and more heterogeneous than the others, reflecting the trend in the gamification design framework production to remain anchored to a few theoretical papers and struggle to build a comprehensive network. Cluster #1, albeit presenting a smaller number of documents, presents a collection of articles related to the self-determination theory and persuasion technique, providing insight into how important this part of the literature is in the design phase of gameful systems and the creation of design frameworks. Clusters #3 and #4 have a lower impact on the overall network; however, their analysis suggests that is important to use specific frameworks for contextual learning, that is a teaching and learning approach that emphasizes the importance of understanding the context in which new information is presented, by connecting it with their daily life, personal, social, and cultural circumstance (Mahendra, 2016).

Overall, the field of gamification design frameworks and models appears to have been shaped initially by documents that focus on persuasion techniques, and generically on gamification definition and background. Lately, there seems to have been a shift in reference documents, giving increasing importance to methodological and critical references.

2.4.3.2 Who are the most influential authors?

To address this research question, we rely on the results of the ACA. Table 2.9 and Table 2.10 give an overview of the most influential authors according to citation frequency and burst strength.

Considering the burst strength, Koivisto J. is the author who attracted the most research attention over a period of time (burst strength = 17.34). The author burst is probably due to the paper *"The rise of motivational information systems: A review of gamification research"* (Koivisto & Hamari, 2019b), which is considered one of the most influential documents according to the DCA analysis. Since the burst ended in

2022, it is legitimate to think that the burst might continue in the foreseeable future. According to the burst strength, the second influential author is Sailer M. The burst is probably related to the paper *"How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction"* (Sailer et al., 2017), which is one of the most influential documents according to the DCA. According to the burst timing, the authors with the earlier burst are Malone T. W. (2011 to 2017), Schell J. (2013 to 2016), McGonigal J. (2013 to 2017), and Gee J. P. (2014 to 2016). Their works mainly focus on teaching and learning mechanics used by video game developers and how these game design elements can be used in education. While, the authors with the later burstness are Koivisto J. (2020 to 2022), Sailer M. (2020 to 2022), Dichev C. (2020 to 2022), Mora A. (2018 to 2022), and Landers R. N. (2019 to 2022). Their works mainly focus on critical reviews about gamification design and application, what are the main issues, and how to face them to design and use properly gameful systems. Taking into account the frequency, Deterding S. (647), Hamari J. (399), Werbach K. (213) Zichermann G. (189), and Ryan R. M. (189) are considered the most influential authors of the network. It is not surprising since these authors produced some papers that are considered a cornerstone in the field of gamification (Deterding, Björk, Nacke, Dixon, & Lawley, 2013; Deterding, Dixon, et al., 2011; Deterding, Sicart, et al., 2011; Hamari et al., 2014; Huotari & Hamari, 2017; Werbach et al., 2012; Zichermann & Cunningham, 2011), and their articles have helped shape what is the current literature on gamification. Interestingly, these authors do not present a citation burst because a highly cited paper may not be identified with a citation burst if its citation counts over the years are relatively stable.

Overall, the ACA results suggest that initially the structure of the literature about gamification design frameworks and models has been based on the analysis of game design and how to implement game design elements in educational contexts. According to the previous section, there seems to have been a shift in reference documents, and the authors have placed more emphasis on critical reviews and research agendas (Koivisto & Hamari, 2019b; Landers et al., 2018).

2.4.3.3 How have research trends changed over time?

To answer this question, we rely on an overview of the keyword's change over time and on the DCA and ACA burst strength time shift. The research trends in the field of gamification design have been analyzed by examining the burst strength of keywords and the burst of the most influential documents. A timeline of these trends was extrapolated from this analysis. Results indicate that there has been a shift in the focus of researchers over time.

According to the analysis of the keywords, the first trends that appeared in the field of gamification design frameworks were "design" (beginning year = 2013), "education" (2015), and "software engineering" (2015). This timeline might reflect the initial interest in the design of gameful systems (for examples see Deterding (2012); Deterding et al. (2013); Zichermann and Cunningham (2011)) and their first application in education, especially in the field of software engineering and computer science (see Barata, Gama, Jorge, and Gonçalves (2013); Kapp (2012) for examples). Later, the keywords changed to "teaching", "application program", "information use" and "artificial intelligence", indicating that the trend has not totally changed, but there has been a growing interest in information use techniques, mainly automated, to provide adaptation of gamified systems. According to the DCA and ACA mentioned above, results indicate how the trend in the literature was initially based on articles and authors that focused on persuasion techniques and the usefulness of gamification (both in a generic way and specific way), and that later the main trend was to question the elements of game design hitherto considered, to create new design methodologies that are effective.

Overall, these results seem to suggest that trends have changed considerably over time, first describing broad motivational intervention designs, and generic design techniques, then leading more and more resources in the direction of a new critical approach toward gamification, suggesting that there are no models or frameworks for the design of gameful systems that are widely used, or that are a reference for the production of further design frameworks or models.

2.4.4 Study conclusion

Gamification, as the application of game elements in non-playful contexts, to motivate and engage users became more and more adopted in the last decade (Koivisto & Hamari, 2019b), along with the increased number of gamification frameworks born with the intent of guiding designers in the definition of gamified systems (Mora et al., 2017). In the current scientometric review, we aimed to explore the structure of the references related to gamification design frameworks and models. Through Scopus, and then CiteSpace, we analyzed a final number of 1,751 publications and 64,109 unique references to identify the most influential documents (**RQ B.1**), authors (**RQ B.2**), and the main trends (**RQ B.3**). Through a DCA we were able to select the most impactful publication (Koivisto & Hamari, 2019b; McGonigal, 2011; Sailer et al., 2017; Seaborn & Fels, 2015) (**RQ B.1**), by ordering the results by burst strength, sigma and citation frequency (Σ). We performed an ACA to answer to **RQ B.2** and identify the authors that contributed more to the research on gamification design frameworks, sorting the authors by the burst strength. Based on the years of interest, we found two distinct sets of authors. Through the analysis of keyword change over time, and the burst strength time shift in both DCA and ACA we noticed a change in the trend of research interests in the study of gamification design frameworks. As for the analysis of the keywords, the results indicate a moderately increased interest in the use of automated techniques to design adaptive gamified systems, which have lately been replaced by an increasing interest in surveys. Also, from the results of the DCA and ACA we observed a shift of emphasis from the implementation of game elements in education, to critical reviews and research agendas (**RQ B.3**).

In conclusion, methodological papers about gamification, game design, and persuasive techniques have influenced the production of gamification design frameworks for nearly a decade. Then, due to numerous inconsistent results with the gamification use, the research interest appears to have changed (Hamari et al., 2014; Koivisto & Hamari, 2019b; Seaborn & Fels, 2015). In line with other findings in the literature (Bassanelli et al., 2022), the current trend seems to be to question current methodologies to research new reliable solutions.

2.5 Limitations

As reported in Section 2.3.3.4, scientometric analyses have limitations that must be considered when interpreting the data. Moreover, the studies I reported in this chapter have some limitations that restrict their validity. First, there was no careful control in selecting and excluding more or less relevant articles. Consequently, no relevant articles were added manually (commonly referred to as snowballing), practice that when combined with database searching, provides a better result in selecting primary studies, greatly increasing the validity of a study (Wohlin, Kalinowski, Felizardo, & Mendes, 2022). Second, reviews were not excluded from the analysis, since what we were interested in the study was an accurate analysis of the networks created around the relevant literature, which was comprehensively composed of survey papers and reviews, which on the one hand may have led to a bias in the creation of co-citation networks, while on the other hand, it made explicit the importance of reviews in the field of gamification.

2.6 Conclusion

Overall, both studies show that there has been a linear change in document and author selection over time. Above all, they show how from an initial pull of references related to persuasive and theoretical techniques, researchers have moved to the need for systematic and standardized methodologies. Specifically, in line with the documents of Koivisto and Hamari (2019b), Seaborn and Fels (2015), and Mora et al. (2017), both studies show that it is considered increasingly important to find new ways to do gameful systems design, improving the reliability and effectiveness of gameful systems.

Starting from these conclusions, we decided to specifically analyze each component that should be considered during the design of gameful systems, relying on what some of the most influential documents and authors from the two scientometric studies reported in their works. The findings are reported in Chapter 3, and are the basis on which the conceptualization and development of the GamiDOC model was based (Chapter 4).

Chapter 3

The challenges in gameful systems design

More and more sources in the literature are reporting that gamification's application doesn't consistently meet the intended standards and sometimes it yields unfavorable or inconclusive outcomes (Bassanelli & Bucchiarone, 2022; Bassanelli et al., 2022; M. A. Hassan, Habiba, Khalid, Shoaib, & Arshad, 2019; B. Huang & Hew, 2018; Koivisto & Hamari, 2019b; Seaborn & Fels, 2015). In other words, the implementation of gamification exhibits a wide range of results, mainly due to interfering factors. In the following chapter, to answer **RQ 1**, we elaborate extensively on the factors that contribute to the challenges in designing gameful systems. The topics discussed are excerpted from the documents *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024), and *"The role of game modality in the outcomes of gamification: A research agenda"* (Gini et al., 2023), and stems from the analysis depicted in Chapter 2. The presentation of the challenges follows the order in which they were investigated, following the same flow as other design frameworks (e.g. Morschheuser, Hamari, Werder, and Abe (2017), Klock et al. (2016), Hunter and Werbach (2012)), initially analyzing user- and context-centred features, then technology-centered, game-centered, and finally concluding with the evaluation features.

3.1 Contextual differences

The term “context” generally pertains to the conditions or surroundings in which something exists or takes place. It often refers to an analytical unit larger than the phenomena being directly investigated (Nilsen, 2020). The role of context can be seen as the environment or setting in which a proposed change is to be implemented. Alternatively, it can be perceived as something more active and influential, significantly shaping the implementation process and its outcomes (Nilsen, 2020). Considering the impact of context is crucial when explaining how or why specific implementation outcomes are achieved. Neglecting context could limit the generalizability of study findings to various settings or circumstances. As such, context can be attributed to the variations in outcomes across different studies (Nilsen & Bernhardsson, 2019).

Within the realm of gamification, the term “context” is often intertwined with “user” and “goals”. There is an undeniable link between these factors. Therefore, how users perceive gamification is contingent on various factors, encompassing individual user characteristics, the context of gamification implementation, and the specific task or activity being gamified. All these elements contribute to users’ perception and engagement with gamification. Nevertheless, contextual factors and the significance of the application domain are frequently underestimated in gamification research and design. Consequently, according to Koivisto and Hamari (2019b), the lack of theoretical comprehension concerning the contextual influence on gamification effectiveness might yield results that cannot realistically be applied to other contexts.

While designing gameful systems, several principal contextual features warrant evaluation. First, the application domain plays a pivotal role in gamification perception. Different gamification design solutions are proposed based on specific usage domains such as education (Bucchiarone et al., 2019b; Saggah, Atkins, & Campion, 2020), cognitive training (Khaleghi, Aghaei, Mahdavi, et al., 2021), behavior change (Epstein, Zemski, Enticott, & Barton, 2021; Hammerschall, 2019), and business (Kappen & Nacke, 2013; Klevers, Sailer, & Günthner, 2016; Kumar, 2013). For instance, in a study by Hallifax, Serna, Marty, Lavoué, and Lavoué (2019), efforts were made to identify factors supporting design choices for tailoring gamification to user profiles (HEXAD

(Tondello et al., 2016), BrainHEX (Nacke, Bateman, & Mandryk, 2014), and Big Five inventory (John, Donahue, & Kentle, 1991)). The outcomes highlighted that the motivational impact of certain game elements fluctuates based on activities or gamified system domains. Second, the device employed by users to access the gameful system is noteworthy in gamification design. The device can influence users' interaction with game elements and participation in activities. For example, wellness and fitness apps with gamified elements are often designed for mobile devices due to their inherent nature and input mechanisms. Third, users' differences can significantly impact how well a system engages and motivates them. These differences include prior gaming experience, learning styles, goals, motivations, physical, and cognitive abilities. Addressing these differences allows designers to tailor the gameful system to better align with user needs and preferences, enhancing the effectiveness and enjoyment of the learning experience.

In conclusion, considering the context in gamification design is vital because the effectiveness of gamification elements, and gameful systems in general, can vary significantly based on the context of use. Aspects like user goals, motivations, device usage, and environmental conditions can all strongly influence how users perceive gamification elements. By incorporating context, designers can craft more customized and effective gamification experiences that are more likely to engage and motivate users. Furthermore, accounting for context ensures that gamification elements are used ethically, respectfully, and appropriately within specific contexts.

3.2 Interpersonal differences

When embarking on the design of gameful systems, it is imperative to take into account the intended audience and the specific attributes of the target group (Kim, 2015). Undoubtedly, tailored interactive systems prove to be more efficacious than adopting a *one size fits all* approach (Tondello et al., 2016). When we talk about individual differences, two key concepts come to the fore: variances in the broader population and differences given by belonging to a specific group. In the former scenario, an eminent motivational theory within the realm of gamification, the Self-Determination Theory

(Ryan & Deci, 2000), illustrates that people are inherently driven by the fulfillment of three fundamental needs: autonomy, competence, and relatedness. While these three fundamental needs are universally shared, each individual is more strongly drawn to one of these needs and less so to the others (Ryan & Deci, 2000). In simple words, this implies that diverse individuals can find intrinsic motivation in various facets of the same activity, a consideration that should be factored in when crafting gamified software (Ryan & Deci, 2000; Ryan, Rigby, & Przybylski, 2006). For instance, an excessive integration of competitive elements might profoundly inspire those who seek to validate their competence, yet fail to captivate individuals driven more by autonomy or relational aspects within the software. Building on existing literature (Morschheuser, Riar, Hamari, & Maedche, 2017; Reeves & Read, 2009; Yee, 2006), competition stands out as a pivotal success factor in video games due to its allure for those seeking self-challenge and performance evaluation (Jung, Schneider, & Valacich, 2010; Reeves & Read, 2009). However, it can lead to demotivation among users based on their attributes, such as goal orientation (mastery or performance) and skill level (Ames & Felker, 1979; Deci, Betley, Kahle, Abrams, & Porac, 1981; Preist, Massung, & Coyle, 2014). According to the research conducted by Klock, Gasparini, Pimenta, and Hamari (2020), a limited number of studies have explored the preference for game elements through Bartle's player types taxonomy (achievers, explorers, killers, and socializers) (Bartle, 1996). Notably, explorers showed the highest affinity for game elements (20), while socializers suggested the fewest game elements (13). The variance in the count of game elements could arise from either an imbalanced selection of game elements in research studies or a genuine aversion to these elements among socializers. Furthermore, although studies concerning the correlation between personal traits (e.g., Big Five taxonomy (Goldberg, 1993)) and game elements are sparse, there seems to be an association between personality traits and distinct preferences for game elements (Klock et al., 2020). For instance, individuals characterized by higher extroversion levels tend to favor badges, competition, points, levels, feedback, meaning, and social networking elements (Codish & Ravid, 2014). Conversely, those with elevated neuroticism levels are inclined towards badges, levels, and prizes (Klock et al., 2020). Furthermore, elements of customization find favor among individuals displaying higher levels of openness to experience (Jia, Xu, Karanam, & Voida, 2016). Nevertheless, a wealth of evidence indicates that user preferences can evolve over time (Santos, Oliveira, Hamari, & Isotani, 2021) and in response to the application context (Hallifax et al., 2019), which underscores the multifaceted challenges encountered in the application of customization techniques within gameful systems.

When examining disparities within distinct segments of the population, compelling evidence suggests that demographic attributes such as gender and age wield substantial influence over individual preferences in the realm of gamified software. In broad terms, technology garners greater favor from younger demographics for a multitude of reasons (Koivisto & Hamari, 2014). The younger population is introduced to technology at an earlier stage in life (Koivisto & Hamari, 2014; Morris & Venkatesh, 2000) and tends to be more susceptible to social pressures (Koivisto & Hamari, 2014; Morris & Venkatesh, 2000; Venkatesh, Morris, Davis, & Davis, 2003; Wang, Wu, & Wang, 2009). Additionally, the younger cohort tends to perceive technology as more utilitarian compared to their older counterparts (Koivisto & Hamari, 2014; Venkatesh et al., 2003). Drawing on usage data from a gamified application called Fitocracy, which encourages users to engage in physical activities, Koivisto and Hamari (2014) discerned that women exhibit heightened enthusiasm for the software's social dimension. This social component appears to fulfill the affiliation needs of older individuals, thereby potentially extending the software's appeal to the elderly for more extended durations in contrast to the younger population. Furthermore, numerous investigations have scrutinized the divergence in the effectiveness of game elements across genders (Busch et al., 2016; Codish & Ravid, 2017; Denden, Tlili, Essalmi, & Jemni, 2017; Klock et al., 2020; Orji, Vassileva, & Mandryk, 2014; Tondello, Mora, & Nacke, 2017). In a comprehensive review, Klock et al. (2020) aggregated available data on gender-related disparities in gamification. While consensus remains elusive concerning the preferences of women versus men regarding game elements (for instance, prizes being recommended for women in Tondello et al. (2017), whereas the same element is suggested for men in Oyibo, Orji, and Vassileva (2017a)), it is crucial to analyze these differences in tandem with contextual factors (Zahedi et al., 2021). Cultural affiliation also emerges as a factor influencing the efficacy of distinct game elements: initial investigations by Oyibo et al. (2017a), and Oyibo, Orji, and Vassileva (2017b) propose that individualist and collectivist cultures could mold preferences for specific game elements. Despite the intriguing insights, the works of Oyibo and colleagues have not yielded consistent findings, necessitating further research to delve into this matter. Overall, during the design phase, it becomes paramount to account for individualistic and demographic disparities. This practice ensures effective engagement across the broader populace and facilitates the adept selection of game elements when the target audience comprises a more specific demographic. This approach circumvents the limitations of a simplistic *one size fits all* strategy (Böckle et al., 2018). Moreover, it is imperative to sustain research endeavors in this trajectory to foster a more profound comprehension of the interplay between game elements, individual disparities, and distinct user categories.

3.3 Goal differences

According to some works of Locke and colleagues (E. Locke & Latham, 2006; E. A. Locke & Latham, 2002), there are four primary mechanisms through which goals can impact users' performances:

- **Goals can steer attention and efforts toward actions relevant to the goals**, diverting resources from non-relevant actions.
- **Goals can serve an invigorating role**, where challenging or hard goals prompt greater effort compared to attainable or easy goals.
- **Goals influence persistence**, with challenging goals often extending sustained effort.
- **Goals indirectly influence actions** by triggering the awakening, exploration, and application of task-relevant knowledge and strategies.

Broadly, goals can either motivate individuals to utilize their existing skills or drive them to seek out new knowledge. As initially highlighted by Latham and Baldes (1975), and by E. A. Locke and Latham (2002) later, distinct goals yield varying impacts on performance, a notion consistently supported in pertinent literature. This variation stems from differing user orientations—some emphasize goal outcomes, while others prioritize the process of attaining these goals (Koivisto & Hamari, 2019b). Furthermore, explicit goals and well-structured activities foster a state of *"flow"* (Csikszentmihalyi, 1991), characterized by complete absorption in an activity. During a state of flow, an individual is entirely engrossed and immersed in their actions, losing track of time and surroundings. This state of flow correlates with optimal engagement, enhanced productivity, and feelings of pleasure and contentment. Achieving flow necessitates an activity that aligns with an individual's skill level. Such immersive experiences are commonly observed in the realm of computer games (Sillaots, 2014).

Given that gamification, and games in a broader sense, are goal-oriented activities (Deterding, Dixon, et al., 2011), it's crucial to emphasize the impact of goals on players' performance in gameful systems and subsequently highlight the significance of establishing clear objectives during their design (Fortes Tondello, Premsukh, & Nacke, 2018). However, the goals of numerous gamification endeavors often appear to lack clarity at the project's inception (Kim, 2015). Furthermore, there's a conspicuous gap in the literature concerning goals within gamification, prompting a growing number of authors to underscore the necessity of goal analysis in the design of gameful systems (Bassanelli & Bucchiarone, 2022; Fortes Tondello et al., 2018; Koivisto & Hamari, 2019b). Diverse methods exist for integrating objectives into gameful systems—these goals can be explicitly defined as missions or quests, or they can manifest implicitly as pursued outcomes (Fortes Tondello et al., 2018). Additionally, gamification encompasses several inherent game elements that serve as mechanisms for goal-setting, such as badges, leaderboards, levels, progress bars, rules, challenges, conflicts, points, achievements, and rewards (Fortes Tondello et al., 2018). While not obligatory, goals frequently find a place in gamification due to their ubiquity in game-based applications. In gamification, these goals are usually well-defined in line with the underlying theoretical framework. Fortes Tondello et al. (2018) propose specific elements to foster clear goal-setting in gameful systems, encompassing boss battles, certificates, collections, exploratory tasks, learning objectives, quests, unlockable or rare content, and access to advanced features. Furthermore, in well-crafted systems, goal difficulty should be calibrated based on users' skill levels to maintain an engaging and appropriately challenging experience. This necessitates the ongoing ability to monitor users' skill progression. For instance, in

PolyGlot, a tool devised by Bucchiarone, Martorella, Colombo, Cicchetti, and Marconi (2021) employing the GamiDOC gamification design framework, goal-setting assumes a pivotal role during the design phase. The tool grants control over goals through gamified components and articulates activities in terms of their objectives. As such, goals can be adjusted according to users' proficiencies while sustaining engagement and challenge.

To conclude, the overarching purpose of a gameful system, the assortment of in-game goals, and their adaptable nature based on users' competencies and advancements constitute pivotal elements for the effectiveness of gamification, warranting careful consideration during the design phase. It's imperative to take into account feedback and appropriate gamification elements that align with the intended purpose. Moreover, the individual preferences of users and their specific goal orientations should not be underestimated. Consequently, the design principles applied in a given scenario will inevitably vary contingent on the objectives users strive to accomplish. Ultimately, anticipating a solitary design solution to be universally applicable across all contexts is unrealistic.

3.4 Game and player modality

The configuration of user interactions, coupled with the selection of objectives, shapes the inherent modality of a gameful system. Thus, comprehending how modality impacts the effectiveness of gamification and identifying factors that aid designers in selecting the most fitting user interaction approach holds paramount importance.

In the context of gamification, we primarily categorize modalities into four: individual, cooperative, competitive, and cooperative-competitive (Liu, Li, & Santhanam, 2013; Morschheuser, Maedche, & Walter, 2017). This classification, originally introduced by Morschheuser, Maedche, and Walter (2017), is rooted in the categorization of video games into four main modalities established by Liu et al. (2013). This distinction draws on the principles of the *Social Interdependence Theory* (D. W. Johnson, 2003), categorizing video games and gamified activities based on the attributes of user interaction.

3.4.1 Social Interdependence Theory

Social Interdependence Theory was initially formulated by Morton Deutsch in 1968 (Deutsch, 1949, 1968; D. W. Johnson, 2003). He re-explored the concepts of group and inter-group dynamics originally conceived by the Gestalt school (Lewin, 1948, 1951), delving deeper into how the interconnectedness of individuals' goals can impact group dynamics. The degree of interdependence between individuals' goals can be categorized as either positive or negative. Positive interdependence arises when the attainment of one individual's goal facilitates other group members in achieving their objectives, often fostering a spirit of cooperation (Deutsch, 1949, 1968; D. W. Johnson, 2003). Conversely, negative interdependence occurs when the accomplishment of an individual's goal impedes others from reaching their goals, fostering a competitive atmosphere. In such cases, individuals tend to vie against each other to achieve their desired outcomes. Finally, a lack of interdependence is observed when there is no discernible relationship between individuals' goals (Deutsch, 1949, 1968; D. W. Johnson, 2003).

3.4.2 Modality in traditional activities

The realm of social and motivational psychology delves into the examination of how various forms of interdependence impact individuals' performance and psychological outcomes. Investigations concerning positive interdependence (which we will call simply "cooperation" from now on), have demonstrated that collaborative activities tend to boost performance in contrast to other modalities (D. W. Johnson & Johnson, 1989, 2005), such as individual tasks (characterized by a lack of interdependence) and cooperative activities with negative interdependence. This enhancement in performance can be attributed to the correlation between cooperation and a higher occurrence of insights and heightened cognitive functioning among participants. Moreover, during cooperative activities, individuals dedicate more time to task engagement compared to individual or competitive endeavors (D. W. Johnson & Johnson, 1989, 2005). On the psychological front, cooperative activities foster increased motivation and elevated levels of well-being (Bagozzi & Dholakia, 2006; D. W. Johnson & Johnson, 1989). These cooperative effects underlie the success of collaborative learning (Al-Rahmi & Othman, 2013; D. W. Johnson, 2003; D. W. Johnson & Johnson, 2009).

Further insights from the literature propose that the amalgamation of positive and negative interdependence amplifies the benefits of social interaction. In scenarios where both cooperation and competition (such as team competitions) are intertwined, participants exhibit heightened motivation (Erev, Bornstein, & Galili, 1993; D. W. Johnson & Johnson, 1989; Tauer & Harackiewicz, 2004). When participants are divided into teams, positive influences (Julian & Perry, 1967) and a sense of social identification can arise among team members, aligning with the concept of "we-intentions" proposed by Hogg and Turner (1985) and Tuomela (2000, 2005).

In the context of traditional learning, the effects of cooperation and competition have been extensively explored. Collaborative learning exhibits a favorable influence on motivational, emotional, cognitive, meta-cognitive, and social aspects of learning (D. W. Johnson & Johnson, 2009). Students perceive collaborative learning activities as more gratifying than individual ones (Al-Rahmi & Othman, 2013). However, opinions regarding the utility of competition in education vary (Cantador & Conde, 2010). Some authors assert that competition can motivate students and foster learning (Cantador & Conde, 2010; Fulu, 2007; Verhoeff, 1997), while others contend that it might encourage performance orientation over a mastery-driven approach to learning (Lam, Yim, Law, & Cheung, 2001). Additionally, competition can lead to stress, as highlighted by Vockell (2004), as substantiated by students' preference for anonymous competitions (Yu, Chang, Liu, & Chan, 2002). Striking a balance between the merits and drawbacks of competition, and considering the evident effectiveness of cooperation, certain scholars propose adopting cooperative-competitive structures as a motivating approach for students (Thousand, Villa, & Nevin, 2002). The literature on traditional learning also outlines guidelines for implementing "safe" competition in education, suggesting brief competitions with well-defined goals and prizes of low or symbolic value to minimize the influence of rewards on student motivation (Cantador & Conde, 2010; Shindler, 2009).

3.4.3 Modality in Video Games

Using these definitions as a foundation, Liu et al. (2013) was the initial proponent followed by Morschheuser, Maedche, and Walter (2017), who subsequently categorized video games and gamified activities into four primary modalities:

- **Individual modality**: This signifies a lack of interdependence among users, where no connection exists between the goals of players.
- **Cooperative modality**: In this mode, positive interdependence between users' goals is fostered, encouraging interaction among players.
- **Competitive modality**: Negative interdependence characterizes this modality, impeding interaction among users' goals.
- **Cooperative-competitive modality**: This modality combines both positive and negative interdependence and is commonly observed in team competitions. Here, interaction is promoted within the team but limited with external teams.

The decisions made by game designers concerning the structure of objectives, introduced game elements, and the synergy of players' roles collectively shape the degree of interdependence among players, ultimately determining the modality of the video game.

Within the literature, **competition** is recognized as a contributing factor to the success of video games (Morschheuser, Riar, et al., 2017; Reeves & Read, 2009; Yee, 2006). It offers players immediate feedback on their performance and cultivates a sense of competence (Jung et al., 2010; Reeves & Read, 2009). Meeting the need for competence, as outlined in *Self Determination Theory* (Ryan et al., 2006), serves as a cornerstone for fostering intrinsic motivation. While competition attracts many players, it can also exert negative influences on game appreciation if it becomes overly controlling if users become overly fixated on goals instead of performance, or if there's an imbalance in players' abilities (Ames & Felker, 1979; Deci et al., 1981; Preist et al., 2014). It's evident that competition doesn't consistently enhance motivation, and whether competition is motivating or demotivating hinges on the nature of the activity and the characteristics of the users. Conversely, cooperative video games yield a distinct array of positive impacts on user behavior and well-being. Positive interdependence within video games promotes cooperation not only within the virtual realm but also extends to the real world (Seif El-Nasr et al., 2010). Moreover, beyond enhancing enjoyment and effort (Hamari, Huotari, & Tolvanen, 2015), engaging in video games characterized by social interaction, particularly cooperative ones, has been correlated with overall psychological well-being (Marker & Staiano, 2015; Przybylski, Rigby, & Ryan, 2010; Teng & Chen, 2014; Yee, 2006). Lastly, the **cooperative** modality serves as a means to fulfill players' requirement for relatedness, another of the three fundamental needs identified within the framework of *Self Determination Theory* (Ryan et al., 2006).

3.4.4 Modality in gamification

Within this section, we elaborate on how modality can influence the user experience of gamified systems within the broader population. Furthermore, we introduce primary evidence suggesting a potential mediating role of demographic and cultural characteristics

3.4.4.1 Differences in the general population

Given the observed effects resulting from adopting various modalities, as revealed by studies in social psychology and video games, it's reasonable to suspect that varying degrees of interdependence could have implications for the effectiveness of gamified activities. Initially, competition garnered popularity within gamification due to the positive impact of introducing competitive game elements like leaderboards on motivation, as compared to the individual modality (Codish & Ravid, 2014; Hamari et al., 2014; Jia, Liu, Yu, & Voida, 2017). However, just as with video games, it is evident that competition cannot be considered a universal solution to engaging users (Codish & Ravid, 2014; Hamari, 2013; Hamari et al., 2014; Jia et al., 2017; Massung, Coyle, Cater, Jay, & Preist, 2013; Preist et al., 2014). The literature underscores that competitive elements yield differing effects on individuals based on personality and other interpersonal disparities (Codish & Ravid, 2014; Hamari, 2013; Hamari et al., 2014; Jia et al., 2017). Leaderboards, in particular, offer a strong incentive for competition as they represent goals to achieve (Landers et al., 2017). Extroverts tend to enjoy competitive elements more than introverts, and situations like crowd-sourcing can lead to skilled participants discouraging less active contributors (Massung et al., 2013; Preist et al., 2014). For instance, Gabrielle (2018) highlighted how hotel employees were negatively impacted by competitive gamification, resulting in increased stress and negative emotions. Comparative studies of different modalities have uncovered similarities between effects seen in gamification and traditional activities, such as levels of enjoyment and engagement (Y. Chen & Pu, 2014; Morschheuser, Maedche, & Walter, 2017). In a study by Y. Chen and Pu (2014), three versions of the same gamified software were compared, each featuring a different modality. The results indicated that the cooperative and cooperative-competitive versions had a more positive impact on individuals compared to the competitive version. Similarly, Morschheuser, Maedche, and Walter (2017) found comparable outcomes when examining three modalities (individual, cooperative, and competitive) in a gamified crowd-sourcing software. Other researchers have noted that cooperative gamification enhances user participation and leads to sustained effects on motivation, enjoyment, and intent to share knowledge when contrasted with other modalities (Morschheuser, Maedche, & Walter, 2017). Furthermore, just as in traditional activities, the cooperative-competitive modality in gamification appears to be even more effective than introducing cooperation or competition alone (Y. Chen & Pu, 2014). In a recent study (Morschheuser, Hamari, & Maedche, 2019), participants reported similar levels of perceived usefulness and usage duration, yet those in the cooperative-competitive condition demonstrated higher levels of enjoyment and participation.

In the systematic review conducted by Klock et al. (2020), preferences for game elements among different player types and personality traits were discussed. Drawing from Bartle's player types taxonomy, individuals are categorized as achievers, explorers, killers, and socializers based on their game preferences (Bartle, 1996). Within the review's findings, competition was specifically suggested for killers in some studies (AL-Smadi, 2014; Fuß, Steuer, Noll, & Miede, 2014), while leaderboards were recommended for achievers (Akasaki et al., 2016; Taspinar, Schmidt, & Schuhbauer, 2016), explorers (Akasaki et al., 2016), and killers (Akasaki et al., 2016; Fernandes & Junior, 2016; Taspinar et al., 2016). Additionally, in terms of personality traits, the majority of studies included in the review referred to the Big Five or OCEAN model (Goldberg, 1993), associating competition with higher levels of extroversion (Roccas, Sagiv, Schwartz, & Knafo, 2002).

3.4.4.2 Demographic differences

We have thus far explored variations within the general population. Evidence suggests that the social aspect of gamification software is perceived differently based on individuals' demographic traits (Klock et al., 2020; Koivisto & Hamari, 2014). For instance, in Klock et al. (2020) study on tailored gamification, the authors examined the game elements and modalities recommended for different genders. Notably, two studies (Codish & Ravid, 2017; Tondello et al., 2017) suggested leaderboards for women, while another indicated leaderboards as suitable for all genders (Denden et al., 2017). Additionally, two other studies recommended competition for men and guilds for both men and women (Busch et al., 2016; Orji et al., 2014). Interestingly, Busch et al. (2016) delved into femininity and masculinity instead of simply focusing on gender, linking social status and competition to femininity. The realm of gender differences in gamification, and particularly the interplay between gender and modality, remains relatively nascent, necessitating further research for clarity. Moreover, Koivisto and Hamari (2014) explored variances in user experience with an exercise gamified software (Fitocracy), differentiating responses by age and gender, and identified a preference for social features among women. Lastly, Itoko, Arita, Kobayashi, and Takagi (2014) observed that competitive gamification may yield more favorable results in younger users compared to older individuals. Their study involved motivating senior workers in crowd-sourcing tasks, such as proofreading, revealing that younger participants embraced competition more than their older counterparts, who exhibited reluctance in comparing their contributions to others. These insights suggest that the impact of modality might be influenced by demographic characteristics in addition to interpersonal disparities.

3.4.4.3 Cultural differences

Culture encompasses various habits and preferences that commonly manifest within a group of individuals, encompassing aspects like social conduct, culinary choices, artistic inclinations, and emotional expression (Miyahara, 2004; Tylor Edward, 1871; Wright, Nancarrow, & Kwok, 2001). Cultural disparities can potentially result in the exclusion of certain users, particularly in educational settings. Indeed, numerous authors have highlighted how factors such as age, ethnic diversity, gender, sexual orientation, and disabilities might influence the effectiveness of gameful systems (L. Hassan & Hamari, 2020; Thiel, Reisinger, & Röderer, 2016; Zahedi et al., 2021). In light of this, many researchers advocate for adopting a cooperative or cooperative-competitive approach (Bucchiarone, Cicchetti, et al., 2021; Hanghøj, Lieberoth, & Misfeldt, 2018; Ortega Sánchez & Gómez Trigueros, 2019; Quintero et al., 2022). Oyibo et al. (2017a, 2017b) delved into the reception of game elements across various populations, particularly individualistic and collectivist cultures. Individualistic cultures prioritize the self, self-sufficiency, and individual objectives, whereas collectivist cultures emphasize community and prioritize society over individual interests (Rhee, Uleman, Lee, & Roman, 1995). In their work, the authors compared individualistic (North American) and collectivist (Asian, African) societies to ascertain which gamification features (competition, learning, and rewards) better suited the respective cultures. Notably, the first publication proposed competition for individualistic cultures (Oyibo et al., 2017a), while the second publication revealed similar findings for both individualistic and collectivist cultures (Oyibo et al., 2017b). Additionally, A. Toda et al. (2022) evaluated the perceived significance of game elements among Brazilian and US citizens, noting a notable distinction in the perceived importance of cooperation but not competition between the two populations. Specifically, Brazilian participants attributed higher importance to cooperation compared to the US sample.

Overall, as presented in this section, and as noted by Gini et al. (2023), although the investigation of various game modalities within gamification is at an early stage of development, it remains crucial to incorporate these insights into the formulation of gameful system design frameworks.

3.4.5 Modality research

As mentioned in the introduction, recent literature (Bassanelli et al., 2022; Kölln, 2022; Mora et al., 2017) underscores the frequent absence of a modality section in gamification frameworks. However, empirical evidence highlights the influence of varying levels of interdependence on psychological and behavioral factors in participants, spanning traditional activities, video games, and even gamified systems (Ames & Felker, 1979; Deci et al., 1981; Erev et al., 1993; Hamari et al., 2014; D. W. Johnson & Johnson, 1989, 2005; Morschheuser, Maedche, & Walter, 2017; Preist et al., 2014; Tauer & Harackiewicz, 2004). Future endeavors should traverse two distinct yet interconnected paths to address this gap and establish stronger ties between gamification research and the design of gamified systems.

As depicted in Figure 3.1, one avenue (**column A**) for researchers involves delving deeper into the nexus between different modalities and the impacts of gamification, taking into account the characteristics of the target audience. This entails: (i) examining how different modalities intersect with interpersonal variations in the population, such as personality traits, (ii) exploring the correlation between different modalities and demographic categories of users, (iii) investigating the correlation between different modalities and cultural disparities, and (iv) assessing the connection between different modalities and less explored factors like contextual application and targeted behavioral outcomes within gamification.

Aside from advancing our comprehension of this subject, discerning how modality moderates the effects of gamification can significantly enhance the formulation of more effective gamification frameworks. Conversely, gamification frameworks ought to incorporate a dedicated modality section. Integrating such a section would yield two implications (**column B**): (i) heightened awareness among designers about the potential impact of different modalities on user experiences, and (ii) a directed effort by designers to contemplate the interdependence among users, potentially leading to the explicit documentation of modality choices. The latter practice would further contribute to building a repository of information about the adopted modalities in gamified systems and the subsequent effects on individuals' psychological and behavioral facets. This database could be harnessed in future reviews, meta-analyses, and the continued refinement of gamification frameworks.

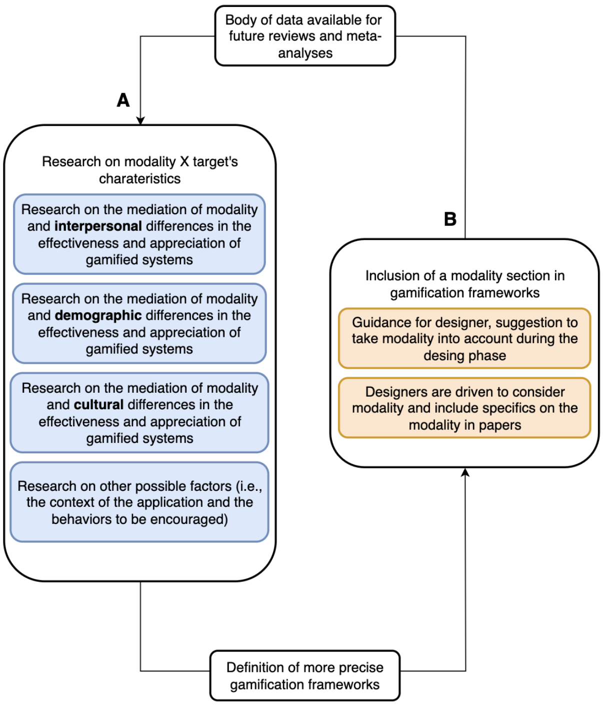


Figure 3.1: Reciprocal benefits between research in gamification and the development of a holistic framework, with the inclusion of a modality section.

3.5 Technology

The device and the different technological features are important in the design of gameful systems. It can affect the user's ability to interact with the game elements — and more generically with the gameful system itself — and participate in the activities. Different devices can rely on different technological features (such as augmented reality, virtual reality, wearable, portable, position tracking, and so forth) (Martinho, Carneiro, Corchado, & Marreiros, 2020), significantly determining a different interaction between users and the system. Some data in the literature (Koivisto & Hamari, 2014) report a different pattern in the approach to technology according to age and gender. This can result in different weights of game elements based on the relationship of technological features and the devices used, the context, and the end user. These elements make certain types of technologies and devices better suited for certain applications; indeed, it is possible to notice the limited use of mobile devices in the learning context (Chin, 2014), since by their inherent nature and with the possibility of integrating different input modalities, they would seem more suitable for health (Martinho et al., 2020) and behavioral change applications (Bucchiarone, Bassanelli, & Marconi, 2023). It is therefore necessary to assume that various game elements may be perceived differently depending on the type of device and the technology used. This implies that designers must be more careful in the selection (and subsequent interaction) of game elements. Despite these claims, further studies are needed.

3.6 Feedback

Feedback constitutes a process that furnishes insights, reactions, or evaluations of specific actions, behaviors, or performances, intending to facilitate growth, learning, or adjustments (Prue & Fairbank, 1981). It manifests in varied forms like spoken or written comments, constructive critiques, evaluations, ratings, or suggestions. Feedback's role is pivotal in personal and professional advancement, as well as in refining the quality of outcomes (Harro-Loit, 2019). It stands as a potent catalyst for motivating individuals to attain designated objectives. Yet, its influence can veer towards either positive or negative outcomes (DeNisi & Kluger, 2000). Thus, it becomes imperative to elucidate what encompasses "feedback" and delineate its integral components.

Broadly, feedback can be conceptualized as *"information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way"* (Ramaprasad, 1983). In essence, feedback addresses any system parameter, whether it's an input, output, or the underlying process. This necessitates specific prerequisites, including access to data concerning both the benchmark and current levels of the parameter in question. Moreover, a suitable mechanism is required to effectively compare and contrast these two datasets, ultimately generating pertinent information about the incongruity between the two levels.

The significance of feedback is primarily derived from its link to goal attainment. However, the interplay between feedback and goal-oriented challenges is intricate. For instance, when feedback falters in bridging the gap between existing comprehension and goals, learners tend to compensate by overestimating their current status or resorting to attributions that diminish their effort and engagement. If the goal itself is ambiguously defined, feedback struggles to effectively address this disparity. In such scenarios, the divide between ongoing learning and intended learning might not be distinct enough for learners to acknowledge the necessity for its reduction (Hattie & Timperley, 2007).

The functions of feedback can be broadly categorized into two main groups: *reinforcement* and *stimulus control* (Duncan & Bruwelheide, 1985). When feedback is dispensed after performances and leads to improved performance or desired behaviors, it fulfills a reinforcement role (e.g., Komaki, Barwick, and Scott (1978); Prue and Fairbank (1981)). However, feedback's realm extends beyond reinforcement alone; instances arise where feedback is considerably delayed after a performance, rendering its reinforcing influence less likely. An alternate function is that of stimulus control. This implies that feedback is often provided ahead of the next opportunity for a response (feedforward feedback). The aim here is to trigger and elicit the behavior that will subsequently be reinforced (e.g., Daniels (2000)). If feedback is given immediately before an opportunity to respond and effectively prompts the desired behavior, it can be considered as functioning as a discriminative stimulus. In addition to its discriminative function, feedback may also serve as an *establishing operation*. Duncan and Bruwelheide (1985) propose that feedback can alter the value of behaviors and outcomes by supplying information about their correlation to obtaining other reinforcers. Consequently, feedback can influence the motivation and significance attached to specific behaviors and outcomes.

Nevertheless, numerous factors moderate the efficacy of feedback, rendering the control of all these factors and the implementation of strategies and interventions equipped with appropriate feedback a challenging endeavor. An analysis by Kruger and DeNisi (1996) indicates that two-thirds of feedback interventions led to enhanced performance. However, the remaining third resulted in a counterproductive effect on performance.

Positive feedback that ensues accomplishments can function as an encouragement to persist in pursuing actions aligned with goals, signaling a heightened commitment to the objective. However, if positive feedback suggests that sufficient progress has already been attained, it might inadvertently diminish motivation. Conversely, negative feedback received after unsuccessful actions can inspire endeavors aligned with goals, conveying insufficient progress. Nonetheless, negative feedback may also dampen motivation if it implies a decline in commitment to the goal (Fishbach, Eyal, & Finkelstein, 2010).

3.6.1 Feedback theories

Several theories endorse the concept and utility of feedback. In this section, we introduce the most renowned ones.

3.6.1.1 Control theory

The core concept of Psychological Control Theory (CT) revolves around individuals' endeavors to regulate a specific aspect, typically their task performance, by adapting their behavior. This process of behavior regulation commences with an evaluation of a performance benchmark, often referred to as a "goal", and involves comparing it with information gathered from the individual's environment, such as feedback on their current performance (Gregory, Beck, & Carr, 2011). Substantial empirical support, as indicated by various studies (e.g., R. E. Johnson, Chang, and Lord (2006)), confirms that when individuals identify disparities between their goals and feedback, they respond by taking action, such as exerting greater effort, to minimize these discrepancies. Nevertheless, how individuals react to the detection of goal-performance discrepancies (GPD) can vary depending on the individual and the circumstances. For instance, addressing a GPD may involve intensifying efforts, but conversely, it could also lead to a reduction or complete abandonment of the goal (Gregory et al., 2011).

3.6.1.2 Goal-setting theory

The Goal Setting Theory (GST) aims to explain the connection between human motivation and the formulation of specific objectives, while also offering guidance on the proper process of goal setting (E. A. Locke & Latham, 2002). Locke emphasized that goals that are specific, clear, and challenging are more effective in prompting individuals to enact desired behaviors. While the definition of goals is crucial, consideration must also be given to individuals' capabilities, ensuring that the goals are realistically achievable (E. A. Locke & Latham, 2002; Lunenburg, 2011). Feedback assumes a vital role in the realm of GST, serving as a reference for the distance between individuals' performance and their goals. According to GST, there exists a relationship between feedback and performance, where the quality of feedback acts as a mediator between goal setting and performance (Ashford & De Stobbeleir, 2013). Notably, feedback enables individuals to establish more refined goals, ultimately facilitating goal achievement (Burns, Martin, & Evans, 2021). Moreover, personalized feedback steers students toward a greater focus on their goals rather than fostering competition (Burns et al., 2021).

3.6.1.3 Self-determination theory

Self-determination theory (SDT) is a macro-theory of human motivation, proposed by Deci and Ryan (Deci & Ryan, 2013) and aimed at describing how individuals voluntarily engage in activities. The authors point out how intrinsic motivation (i.e., internal drive to engage in an activity for its own sake) arises when individuals satisfy their needs for competence (i.e., feeling competent), autonomy (i.e., perceiving freedom of choice), and relatedness (having meaningful relationships) are satisfied through the activity performed. On the other hand, extrinsically motivated activities require rewards. While the satisfaction of the three basic needs plays a big role in human behaviors, sometimes individuals engage in non-motivating activities because they deem them as important for them or the society (i.e., autonomous regulation, see Organismic Integration mini-theory for a more exhaustive argumentation (Deci & Ryan, 2013)). The research highlighted how clear and well-defined feedback can help students' motivation, along with the satisfaction of the three basic needs (Cheon, Reeve, & Song, 2019; Mouratidis, Vansteenkiste, Michou, & Lens, 2013; Pat-El, Tillema, & Van Koppen, 2012). The results, though, are far from providing a clear picture of how feedback should be provided to satisfy the needs for competence, autonomy, and relationship. Krigksman et al. (Krijgsman, Mainhard, Borghouts, van Tartwijk, & Haerens, 2021) for example did not find any differences in students' motivation and satisfaction during PE classes based on different levels of clarification of goals and feedback during the activities. Rogers's (Rogers, 2017) research on video games and SDT demonstrated that feedback needs to be carefully balanced, to avoid players perceiving it as controlling and increasing their cognitive load. In gameful systems, feedback is often delivered through game elements. While game elements can easily function as a reward and increase extrinsic motivation, they can represent a threat in case users' are already intrinsically motivated, as users may shift their focus from the enjoyment of the activity itself to the desire to gain the rewards (Rogers, 2017). See the overjustification effect (Lepper, Greene, & Nisbett, 1973) for a comprehensive explanation of the phenomena.

3.6.1.4 Law of effect

Thorndike's Law of Effect (Thorndike, 1927) stands as the most influential theory in this domain. According to this principle, positive feedback is regarded as a type of reinforcement, while negative feedback is linked with punishment. Both reinforcement and punishment contribute to the enhancement of learning and, consequently, performance. It is anticipated that both positive and negative feedback will enhance performance since they serve to reinforce correct behavior and penalize incorrect behavior, respectively.

3.6.2 Feedback and behavior change

Habits are essential in our daily lives, helping us navigate the complex routines we face. They are often defined as behaviors that automatically respond to situational cues, formed through learned associations between these cues and our actions (Hermsen, Frost, Renes, & Kerkhof, 2016). Habits offer a way to manage the intricate nature of everyday life. Altering habits through interventions that require conscious thought, like persuasive messages, can be challenging due to the automatic nature of habitual behaviors, which operate without active consideration of the current context. An effective strategy for breaking habits is to modify the circumstances to prevent habitual cues from emerging or to change the external triggers that prompt these habitual actions. Because habitual behavior typically occurs unconsciously and without deliberate intent (Bargh, 1994), one successful approach to disrupt unwanted habits is to bring both the habitual behavior and its context into conscious awareness. Self-monitoring, a process in which individuals track and record instances of their target behaviors, enables us to perceive our actions and adapt to the current situation. Consequently, self-monitoring can lead to a reduction in undesirable behaviors (Hermsen et al., 2016). It's important to note that self-monitoring can be challenging, even for highly motivated individuals.

Feedback interventions are generally effective not only at disrupting undesirable habits but also at facilitating behavioral change. Digital technology, equipped with sensitive monitoring devices often worn on the body, can provide continuous, real-time updates on our progress. This capability makes gameful systems, with their feedback provision abilities, widely employed in the field of behavior change.

3.6.3 Feedback in education

According to Hattie and Timperley (2007), in the educational context, "feedback" is defined as information given by various sources (such as teachers, peers, textbooks, parents, self-reflection, and personal experience) regarding one's performance or comprehension of a subject. Consequently, feedback is viewed as a consequence of one's performance and is closely linked to monitoring progress. Its primary purposes are to highlight the learner's actions, provide feedback on their progress, motivate most learners to continue, and offer guidance on how to enhance future performance (Awais, Habiba, Khalid, Shoaib, & Arshad, 2019; Strmecki, Bernik, & Radosevic, 2015). Recent literature acknowledges the pivotal role of feedback in the learning process, considering it a fundamental aspect of the educational system (Panadero & Lipnevich, 2022; Winstone, Nash, Parker, & Rowntree, 2017). In general, students tend to prefer personalized feedback over grades, although an excessive amount of information can become overwhelming (Nicol & Macfarlane-Dick, 2006; Orsmond, Merry, & Reiling, 2005), especially when it doesn't clearly indicate where and how performance can be improved (Winstone et al., 2017). Furthermore, while students appreciate tailored feedback, it should focus on their performance rather than their attributes, as concentrating on a learner's characteristics can be demotivating (Schartel, 2012). Based on its functions, feedback can be broadly categorized into two main groups: *reinforcement* and *stimulus control* (Duncan & Bruwelheide, 1985). When feedback is provided after the performance and leads to improved performance or desired behaviors, it can be considered as serving a reinforcement function (e.g., see Komaki et al. (1978); Prue and Fairbank (1981)). However, rewards such as grades, stickers, and awards are less effective forms of feedback since they lack detailed information about performance and often act as extrinsic rewards, which can hinder student engagement and interest in the subject matter (Deci, Koestner, & Ryan, 1999). Fortunately, feedback is not limited to reinforcement alone. There are instances where feedback is delayed significantly after a performance, making it less likely to have a reinforcing effect. On the other hand, feedback can serve a discriminative function, often provided before the next opportunity for a response (known as "feedforward") to encourage the desired behavior that will be reinforced (e.g., Daniels (2000)). Hattie and Timperley (2007) emphasize that feedback is not merely about answering the question, *"How am I going?"* but should also help in understanding *"Where am I going?"* and *"Where to next?"*. If feedback is given immediately before an opportunity to respond and effectively prompts the desired behavior, it can be considered as functioning as a *discriminative stimulus*. Additionally, feedback may also serve as an *establishing operation*, modifying the value of behaviors and outcomes by providing information about their relation to obtaining other reinforcers, thereby influencing motivation and the significance attached to specific behaviors and outcomes (Duncan & Bruwelheide, 1985).

In education, feedback is conceptualized from three perspectives: that of the student, the teacher, and the teaching-learning process (Ovando, 1994). Carlson (1979) notes that feedback represents authoritative information that students receive to reinforce or modify their responses to instruction, guiding them more effectively in achieving course goals. According to Ovando (1994), feedback has an informative nature and follows a systematic process in which teaching and learning function cohesively to carry out the instructional process and gather specific information for teaching and learning. Therefore, feedback should be consistent, well-organized, and methodical. Importantly, learners actively engage in correcting their behavior based on the feedback received, making it essential to motivate and involve them in the feedback process to achieve the desired outcomes (Higgins, Hartley, & Skelton, 2001, 2002; Winstone et al., 2017). Integrating gamification into the feedback process could present an innovative way to engage students. However, badges alone, like other game elements, may not suffice to provide students with comprehensive feedback. As highlighted by Deci et al. (1999), stickers and awards often lack sufficient information about performance and the subsequent steps for improvement. Interestingly, students are not always motivated by praise, which can evoke positive emotions but may not necessarily lead to improved learning outcomes (Brummelman & Sedikides, 2020; Dev, 1997; Lipnevich & Smith, 2009). In particular, Lipnevich and Smith (2009) compared different forms of feedback and found that grades and computerized praise are not effective in guiding students toward better performance.

With the increasing role of technology in education, some studies have focused on the effectiveness of feedback provided by virtual agents (Winstone et al., 2017). Learners generally appreciate computerized feedback, as they perceive it to be more neutral and objective and less emotionally biased (Lipnevich & Smith, 2009). However, digital feedback may be challenging to personalize (Cutts, Carbone, & van Haaster, 2004; Lymn & Mostyn, 2010), and online courses often have a modular structure in which users are assessed only at the end of each section, reducing the likelihood of using feedback to enhance performance (Jonsson, 2013).

3.6.4 Feedback in gamification

Within the realm of gamification — where game elements are integrated into non-game contexts (Deterding, Dixon, et al., 2011; Huotari & Hamari, 2011) to yield motivational advantages, stimulate creativity, infuse playfulness, drive engagement, overall positive development, and well-being (Hamari, 2019) — feedback emerges as a pivotal factor in engaging and motivating participants, heightening their learning journey, and steering desired actions. Indeed, gamification can be perceived as an endeavor to offer users feedback regarding goal-oriented conduct or performance (Koivisto & Hamari, 2019b).

A cornerstone perspective in gamification literature, presented by Zichermann and Cunningham (2011), underscores feedback as among the simplest and most essential game mechanics. Their definition, describing feedback as "returning information to players and informing them of where they are at present, ideally against a continuum of progress" attributes equal significance to feedback in both gaming and e-learning systems. Since game elements serve as feedback to users, they can be viewed as direct feedback, while also falling into a distinct category of game elements (A. M. Toda, Oliveira, et al., 2019). Collectively, game elements can operate as feedback, or feedback can complement game elements (Awais et al., 2019), contributing to enhanced outcomes. Additionally, gameful systems, particularly in education, enable tailored, individualized game elements (Awais et al., 2019; Hallifax, Lavoué, & Serna, 2020) and automated feedback based on each user's distinct preferences (Serral, De Weerdt, Sedrakyan, & Snoeck, 2016). The extraction of insights from gameful systems is often facilitated through the integration of gamification elements within the tool.

In gamification, the fusion of feedback mechanisms and narrative components empowers users by fostering a sense of responsibility for their achievements. These elements also cultivate a feeling of being part of something greater, nurturing self-efficacy as users contribute to a larger purpose. The combination of feedback mechanisms and narrative can lead users to perceive direct accountability for their accomplishments (Fortes Tondello et al., 2018). The amalgamation of goals and feedback produces superior outcomes compared to either in isolation. Providing individuals with the means to track their progress towards goal attainment is vital, enabling adaptive strategies and effort allocation. Feedback consistently emerges as a crucial component in gameful design methodologies, alongside goals and challenges (Chou, 2019; Deterding, 2015; Marczewski, 2013). Progress bars, as highlighted by Landers, Bauer, Callan, and Armstrong (2015), represent a frequently employed form of feedback in gamification. Nevertheless, gamification presents a diverse spectrum of design elements that serve as feedback mechanisms (such as points, levels, achievements, badges, quest completion, leaderboards, avatars, narrative or storytelling elements, rewards, and more). Consequently, empirical evidence from goal-setting studies, revealing the moderating impact of feedback on the interplay between goals and performance, underscores the potential of feedback in gamification to elevate performance levels (Fortes Tondello et al., 2018).

Several elements make feedback important in gameful systems (Figure 3.2):

- **Feedback provides users with a sense of progress, achievement, and rewards**, which are essential for maintaining their motivation and engagement. It helps create a feedback loop where players receive information about their actions, performance, and progress, which can fuel their desire to continue playing and improving (Huotari & Hamari, 2012).

- **Feedback in gamification helps players learn and develop new skills**. Constructive feedback guides players on what they are doing well and what areas they need to improve. It can provide instructions, suggestions, or demonstrations to help players understand how to perform better and overcome challenges (Awais et al., 2019).

- **Feedback helps players understand their objectives and goals within a gameful system**. It provides clarity on what they need to accomplish and how they are progressing towards those goals. Clear and timely feedback enables players to make informed decisions, adjust their strategies, and stay focused on their desired outcomes (Fortes Tondello et al., 2018).

- **Feedback in gamification reinforces desired behaviors and actions**. It can provide positive reinforcement when players exhibit behaviors aligned with the intended objectives of the game. By providing immediate and tangible feedback, gamification can shape and encourage specific actions, helping players develop productive habits (Bassanelli et al., 2022).

- **Feedback enhances the overall user experience and enjoyment of a gameful system**. It creates a sense of interactivity and responsiveness, making players feel acknowledged and involved. Well-designed feedback systems in gamification can create a dynamic and immersive environment that keeps players engaged and entertained (Jahn et al., 2021).

Recognizing the importance of feedback within gaming contexts, it becomes imperative to delve into the diverse gameful methodologies employed and the potential avenues for integrating feedback within gameful education (Koivisto & Hamari, 2019b). Solely embedding game elements, like points and badges, as external rewards, falls short of cultivating the desired motivational advantages sought by certain gamification researchers. Thus, the critical nexus between formative assessment and effective gaming

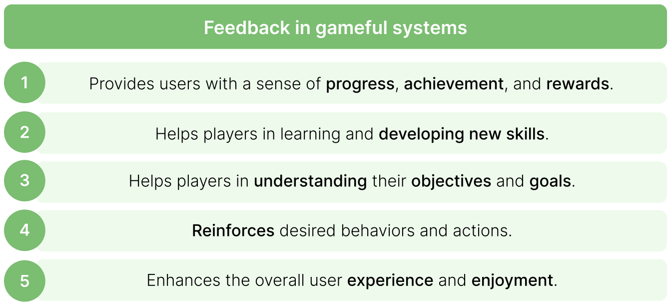


Figure 3.2: Five key points of feedback in gameful systems.

lies in their shared capacity to transparently showcase advancements and feasible objectives (Fuchs & Wolff, 2016). However, the gamification design can yield various implementations. As elucidated by Fuchs and Wolff (2016) and Morrison and DiSalvo (2014), badges and points can be employed in a meaningful manner to afford students a more profound comprehension of their progress. Nevertheless, a comprehensive exploration concerning specific feedback types, the most effective gamification features for their delivery, and the impact of feedback on system users remains an area that warrants more in-depth examination (Bassanelli & Bucchiarone, 2022; Koivisto & Hamari, 2019b).

3.7 Game elements selection

Game elements, which constitute the basic components of games, are utilized in non-playful applications to encourage positive changes in behavior. Game elements have been classified in various ways, such as through the Octalysis framework (Chou, 2019) and Toda et al.'s taxonomy (A. M. Toda, Klock, et al., 2019). While Toda et al.'s taxonomy focuses on the attributes of game elements, Chou's framework centers on the impact of these elements on human motivation. As mentioned in the introduction, gameful systems designers extensively relied on points, badges, and leaderboards (collectively known as the PBL triad) to gamify their digital services. The choice of points, badges, and leaderboards – also known as the PBL triad – in gamification as the default way to motivate users (i) did not allow the exploration of different, and possibly more appropriate, game elements, and (ii) promoted the adoption of a *one* *size fits all* approach.

Some authors pointed out how the PBL triad is not only potentially reducing the motivational effect of gamification to its short-term effects (Rapp, 2015; Trinidad, Calderón, & Ruiz, 2021), but it also limits users' experience to specific aspects of gamification, instead of engaging them in a vast possibility of meaningful experiences (Robertson, 2010; Trinidad, Calderón, & Ruiz, 2021; Werbach et al., 2012). The authors' reflections lead us to consider how gameful system designers should focus on other, less explored, game elements to broaden our knowledge and create better experiences for users. An interesting approach in this direction is the focus on the creation of meaningful experiences, rather than the choice of game elements as rewards for displaying specific behaviors. Nicholson (2015), for example, developed a framework based on aspects of game design such as *play, exposition, choice, information, engagement, and reflation*, based on Self Determination Theory and aimed to design intrinsically motivating experiences. Using the taxonomy reported by Toda et al. as a reference point, it is possible to identify other game elements that designers can use to create meaningful gameful experiences: *levels* represent a relationship between the user and their benefits as they progress within the gameful system; *stats* inform the user about their performance; while *progress bars* represent the advancement of the user, locating his position in respect to a goal (see A. M. Toda, Oliveira, et al. (2019) for a full list of game elements and their description). *Narrative*, as the order of the occurrences experienced by the user (A. M. Toda, Oliveira, et al., 2019), is recently gaining researchers' attention. P. T. Palomino, Toda, Oliveira, Cristea, and Isotani (2019) defined narrative as "the process in which the user builds his own experience through a given content, exercising their freedom of choice in a given space and period of time, bounded by the system's logic". In another paper, P. Palomino et al. (2019) explore users' preferences for game elements and their relationship with the narrative one. The authors underline the variety of rules that link narrative with the rest of the game elements, and the necessity to deepen our knowledge in that direction. In the Kids Go Green project, Bucchiarone, Bassanelli, and Marconi (2023) used virtual journeys to incentivize children to opt for sustainable means of transportation for going to and coming back to school. In Kids Go Green, children are rewarded virtual kilometers, which allows them to unlock new milestones in their class' educational journey. Similarly, Trinidad, Calderón, and Ruiz (2021)'s GoRace users participate in a legendary Olympic race and advance in a map based on their real-life performance.

From the literature emerges how the choice of game elements during the gamification design process is far from being an easy task, and designers need to take into account multiple factors to provide users with meaningful experiences and support their behavioral change.

3.8 Methodological rigor

In many gamification research papers, the reporting of the methods, data, analysis, and results presents inadequate methodological rigor, making interpretation, reliability, and generalization of data difficult (Koivisto & Hamari, 2019b). Gamification research and its theoretical conceptualizations should acknowledge the dynamic, cyclical nature of gamification; gamification research should increasingly employ controlled experimental research methods, to gain knowledge of the actual effects of gamification. Calculating the correct sample size (Francis et al., 2010), namely the number of people participating in the study, allows researchers to minimize false positive and negative results (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005; Francis et al., 2010; Smith & Bryant, 1975) (also referred to as type I error and type II error (Banerjee, Chitnis, Jadhav, Bhawalkar, & Chaudhury, 2009)), which is a first step in the generalization of findings. The minimum required sample size depends on the experimental design, along with the statistical tests chosen to analyze data (Francis et al., 2010; O'Keefe, 2007). Also, differences in parameters such as the critical alpha and the expected effect size influence the number of participants (Francis et al., 2010; O'Keefe, 2007). The procedure of calculating the sample size is called "a-priori power analysis" (O'Keefe, 2007) and it has to be done before the collection of data. Researchers can rely on different software for calculating the a-priori power analysis, or different sources from the literature (Boddy, 2016; Hwang & Salvendy, 2010; Louangrath, 2017). G*power (Erdfelder, Faul, & Buchner, 1996) and MorePower (Campbell & Thompson, 2012) are two of the most used software for the computation of the sample size with many different statistical tests; they allow automatic calculation of the necessary sample based on the distribution of the reference population or by selection of the statistical tests used. As reported by many authors (Hamari et al., 2014; Koivisto & Hamari, 2019b; Seaborn & Fels, 2015), the empirical research on gamification is limited in terms of sample sizes; as a result, the sample sizes of studies should be large enough to increase methodological rigor, as well as to amplify the transferability and explanatory power of the results. Finally, after the collection and analysis of data, it is important to calculate the effect size (Kelley & Preacher, 2012). The effect size is defined as the value that quantifies the strength of the relationship between two variables (Kelley & Preacher, 2012) and it can be calculated through many different methods (Rosenthal, Cooper, Hedges, et al., 1994). It represents the magnitude of the effect, and it is useful to compare different outcomes (Kelley & Preacher, 2012). For example, effect size in gamification can be used to evaluate differences in the effectiveness of two gamified software. Although it is not always easy, following these procedures is essential to increase the methodological rigor in gamification and shed light on some open issues.

In addition to the sample size and effect size calculation, also the time spans of studies should be properly considered. Several authors (Farzan et al., 2008; Koivisto & Hamari, 2014, 2019b) report that the time users spend interacting with a gameful system within a study should be long enough to enable novelty effects in the data to be minimized. Short timeframes pose an evident threat to the validity of study findings. In particular, novelty has been shown to affect users of gamification services (Farzan et al., 2008), and by gathering after a short period of time, the risk of findings being skewed by the novelty of the implementation is elevated. Unfortunately, there are no specific guidelines in the literature. Considering this aspect at the evaluation stage could significantly improve the reliability of the data obtained. Moreover, not many studies conducted follow-up analyses of whether gamification had been effective over a long period after the use; hence, it is possible that the positive changes observed may not be sustained over time (D. Johnson et al., 2017). It is therefore increasingly necessary to design long-term evaluations to make the true effectiveness of gameful systems more transparent.

Some authors (Chambers & Tzavella, 2022; Soderberg et al., 2021) suggest that registered reports and study preregistrations might provide an increased methodological rigor. According to Soderberg et al. (2021), the use of registered reports shows statistically robust and large improvements in attributes, such as methodological rigor and overall article quality, and at the same time, being statistically indistinguishable from comparison papers in terms of features such as novelty and creativity. Moreover, these techniques can also provide: (i) the possibility of successful replication (Camerer et al., 2016), (ii) the generalization, (iii) the data transparency (Reich, 2021), (iv) allow fast publications, and (v) provide control over the study design and the methodological rigor (Spiller & Olff, 2018). Furthermore, while it is true that research on gamification is still trying to understand how to motivate specific behaviors in the short term, the evaluation of the effectiveness of gameful systems should also move forward in the understanding of the long-term effect. To assess the long-term strengths and weaknesses of gamification, more longitudinal studies are needed.

Overall, in the gamification research domain, it is deemed necessary to improve the methodological rigor of the studies, to comprehensively evaluate the real weight of gamification components, determine the factors that contribute to its success, and then ensure the effectiveness and validity of gamification efforts. To pursue these results, several options that provide rigorous methods are available. Hence, we expect that in the foreseeable future, this factor will be taken into account to increase the effectiveness of the gameful systems.

3.9 Collected data

In empirical studies, two main outcomes can be collected: psychological and behavioral outcomes (Koivisto & Hamari, 2019b). For behavioral outcomes, several affordances, such as badges, points, and leaderboards, as well as course results or assignment grades, and other forms of measuring performance are used. Thus, it is not always possible to determine whether the behavioral change is due to the use of the gameful system. It is very difficult to have baseline behavioral data before the use of gameful systems, so any changes are inferred from a hypothetical baseline, or through subjective questionnaires. Moreover, few studies present a follow-up analysis, then it is difficult to check whether the behavior is maintained over time.

Research on gamified systems should also address ethical concerns. Behavioral data collection and analysis can introduce ethical considerations from both personal and societal perspectives. Individual players within gamified systems are not a homogeneous group, and researchers must consider the potential for overload, where excessive gamification can undermine its intended purpose. Designers and researchers need to carefully evaluate whether gamification provides enduring value within specific contexts or if it might backfire by causing overload and undesirable outcomes, both in the targeted environment and in real-life situations where the competition may spill over. Additionally, there are ethical concerns related to the hidden motivations behind gamification. In some cases, the use of player data for purposes other than those explicitly disclosed in the gamified system may arise, creating ethical dilemmas for both designers and users. These ethical issues highlight the importance of transparency, consent, and data protection in the design and implementation of gamified systems (Hyrynsalmi, Kimppa, & Smed, 2017a).

The extensive range of psychological results observed in empirical studies is vast. Nevertheless, because of the broad distribution of research frameworks, there is an absence of significant accumulation of knowledge from any particular standpoint. To illustrate, the assessment of fun, entertainment, enjoyment or usefulness, effectiveness, and efficiency outcomes frequently employs specific instruments tailored for each investigation. Additionally, although numerous studies have explored the perceptions regarding the utilization of a specific system, these findings remain confined to that particular system, offering limited potential for generalization. Overall, psychological elements are analyzed through questionnaires that are often constructed ad-hoc for a specific study. This leads to low reliability and, consequently, difficulty in generalization and comparison (Koivisto & Hamari, 2019b). Hence, it seems necessary to understand which elements are best suited to be collected and which questionnaires are more reliable for assessing psychological outcomes.

3.10 Gamification design

The creation of gameful systems is a task of intricate complexity, demanding meticulous forethought to ensure their effective functioning. The inclusion of chance-based gamification elements such as points and badges within software does not inherently guarantee triumphant outcomes. Consequently, a multitude of design frameworks have emerged to serve as guiding principles for both researchers and practitioners throughout the design process (Mora et al., 2017). As a result, a diverse array of gamification designs and methodologies has arisen to accommodate the unique objectives and requisites of varied contexts. However, it is important to note that only a limited fraction of gamification research employs these underlying theoretical frameworks (Bozkurt & Durak, 2018; Kölln, 2022), and only three out of ten frameworks for education have been validated so far (Rauschenberger, Willems, Ternieden, & Thomaschewski, 2019). The divergence in gamification design paradigms presents both advantageous prospects and challenges. On one hand, it fosters a broad spectrum of imaginative and pioneering approaches to gamification. Conversely, it complicates the comparative analysis and assessment of distinct gamification designs, making it arduous to ascertain which approaches are most efficacious within differing scenarios (Bassanelli et al., 2022). Furthermore, the prevailing absence of standardized and comprehensive design procedures for gameful systems introduces complications in implementing personalization techniques, accurately evaluating the significance of each gamified element and feedback mechanism, and appropriately gauging user experience.

The lack of standardized and controlled paradigms can also lead to ethical issues concerning the developed system (as also presented in Section 3.9). According to Hyrynsalmi, Kimppa, and Smed (2017b), gamification design can have implications for people and society. Cognitive limits apply to gameful systems as they do to every other aspect of life. It is therefore necessary to bear this in mind during the design phase so as not to create dangers for users. For an overly simplified example, a gamified mobility solution can seriously endanger someone's life whether the user can check it while driving. Furthermore, considering all users as a homogeneous group could lead those in weaker positions (elderly, children, etc.) into dangerous positions. Therefore, an analysis of target users and their dynamics is necessary to avoid certain problems. Many companies use systems with leaderboards. Employees could easily falsify information (cheating) to advance their position, damaging the working environment. For this reason, elements of ethics should also be considered during gameful system design. Despite being a fundamental element in system design, design frameworks rarely consider ethics. For example, Marczewski (2017) presents a short framework to prevent the potential dangers of personal morals, or lack thereof, overruling ethics: (i) Does the system offer a choice? (ii) What is the intention of the designer? (iii) What are the potential positive and negative outcomes of being in the system? (iv) Are the beneficial outcomes weighted toward the needs or desires of the user or the designer?

As reported in Chapter 2, the prevailing absence of standardized and comprehensive design procedures for gameful systems has induced a shift in research focus, prompting more scholars to scrutinize systematic methodologies for devising, constructing, and appraising gamified solutions, thus leading to the development of various gamification design models and methodologies. However, the mere establishment of novel methodologies does not intrinsically resolve design predicaments. The creation of fresh frameworks or approaches might not necessarily translate into their effective utilization. Indeed, a constructive approach could involve leveraging pre-existing models, while concurrently designing supportive services aimed at assisting the community in discerning and implementing the most pertinent system (Kölln, 2022).

3.10.1 Related works

Distinguishing the creation of a game from the development of a gameful system reveals a significant connection between the two processes, albeit with distinct nuances. The bedrock of gamification draws extensively from the tenets of game design theory (Mora, Riera, Gonzalez, & Arnedo-Moreno, 2015). However, the design and execution of gamification demand substantial investments of both time and funds, and the challenges encountered in this journey might outweigh the anticipated gains (O'Donovan, Gain, & Marais, 2013). Indeed, multiple experts posit that for more dependable and favorable outcomes, gamification design should be meticulously tailored to specific contexts, oriented toward predefined objectives, directed at target user groups, and should embrace innovative design methodologies that diverge from traditional approaches (Bassanelli & Bucchiarone, 2022; Koivisto & Hamari, 2019b; Morschheuser, Hamari, Werder, & Abe, 2017). In pursuit of enhanced efficiency and expeditious development of gameful systems, a series of targeted design frameworks have emerged. A majority of these frameworks have predominantly catered to business contexts, leaving a dearth of research dedicated to generic and other applications, such as learning and health-related frameworks (Mora et al., 2017). Furthermore, comprehensive analyses (Azouz & Lefdaoui, 2018; Mora, Riera, et al., 2015; Mora et al., 2017) underscore a diverse array of scrutinized frameworks, ranging from highly theoretical constructs to more conceptual viewpoints, encompassing perspectives on gamification from various domains of interest. Additionally, it's notable that much of the existing literature tends to concentrate on specific ad-hoc experiences rather than formalized design processes (Mora et al., 2017).

Numerous gamification design frameworks, exemplified by the works of Klock et al. (2016) and Mora, Zaharias, González, and Arnedo-Moreno (2015), trace their roots to the **MDA model** (Hunicke et al., 2004). This model strives to amalgamate game design, development, critique, and technical game research by presenting a structural approach to understanding games. It encompasses three primary components: (i) *mechanics*, which delineates the specific game elements at the level of data representation and algorithms; (i) *dynamics*, which elucidates the real-time behavior of mechanics as they interact with player inputs and each other's outputs over time; and (i) *aesthetics*, which encapsulates the desired emotional responses evoked in players during their interaction with the game system. Among the pioneering frameworks for gameful systems is the **6D model** by Hunter and Werbach (2012). This framework outlines a six-step roadmap for constructing a gamified system: (i) identifying the business objectives driving the gamification initiative, (ii) specifying the target behaviors to be encouraged or discouraged, (iii) defining the intended player demographic, (iv) devising short-term and long-term engagement loops, (v) ensuring the enjoyability of the gamified system, and (vi) implementing the requisite tools. Over time, this framework has inspired numerous authors to create additional design frameworks (see Brito, Vieira, and Duran (2015); Julius and Salo (2013); Schönen (2014); Tomé Klock et al. (2015))

or adapt it in different contexts (see Gomez-Jaramillo, Moreno-Cadavid, and Zapata-Jaramillo (2018); Yusof, Atan, Harun, Salleh, and Khamis (2022)). Among the most prominent gamification frameworks is the **Octalisys Framework** devised by Chou (2019) (Figure 3.3). This framework delineates eight "core drives" motivating players within gamified systems: epic meaning and calling, development and accomplishment, empowerment of creativity and feedback, ownership and possession, social influence and relatedness, scarcity and impatience, unpredictability and curiosity, and loss and avoidance. Notably, these "core drives" on the right are labeled "right brain core drives", aligning with creativity, self-expression, and social aspects, while those on the left are termed "left brain core drives", primarily associated with logic, calculations, and ownership.

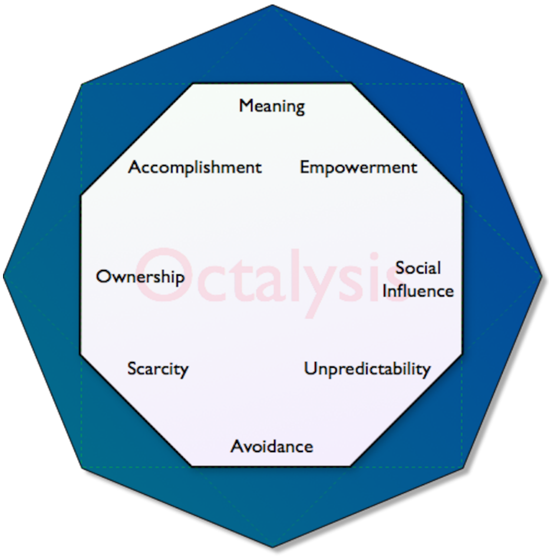


Figure 3.3: Octalysis framework, from Chou (2019).

Building upon literature reviews and expert insights, Morschheuser, Hamari, Werder, and Abe (2017) crafted a gamification design process comprising seven stages: (i) project preparation, (ii) analysis, (iii) ideation, (iv) design, (v) implementation, (vi) evaluation, and (vii) monitoring. This solution facilitates a thorough understanding of users and application areas before transitioning to prototype evaluation. On a different note, Klock et al. (2016) introduced a framework (**5W2H**) to guide the creation, development, and assessment of user-centric gamification through a series of steps addressing key questions (Figure 3.4): (i) who, (ii) what, (iii) why, (iv) when, (v) how,

(vi) where, and (vii) how much.

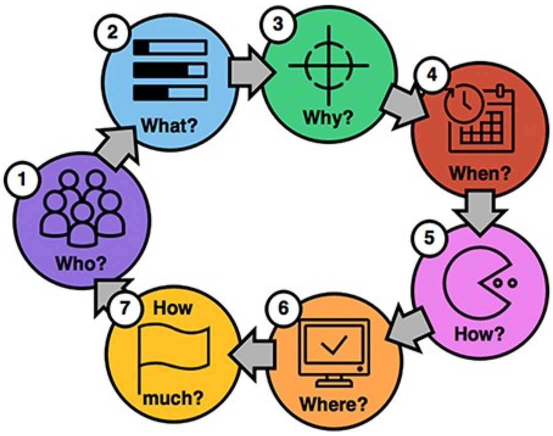


Figure 3.4: 5W2H framework, from Klock et al. (2016).

Conversely, Bucchiarone et al. (2019b) conceived the **GDF** (Gamification Design Framework), comprising five layers (Figure 3.5). These layers encompass three components in the game modeling process: the gamification model (GML), outlining game elements; the game model (GaML), focusing on game mechanics; and the game instance (GiML), addressing game dynamics. Additionally, two utility layers—the game simulation (GsML) and game adaptation (GadML)—can be applied to any of the game modeling layers.

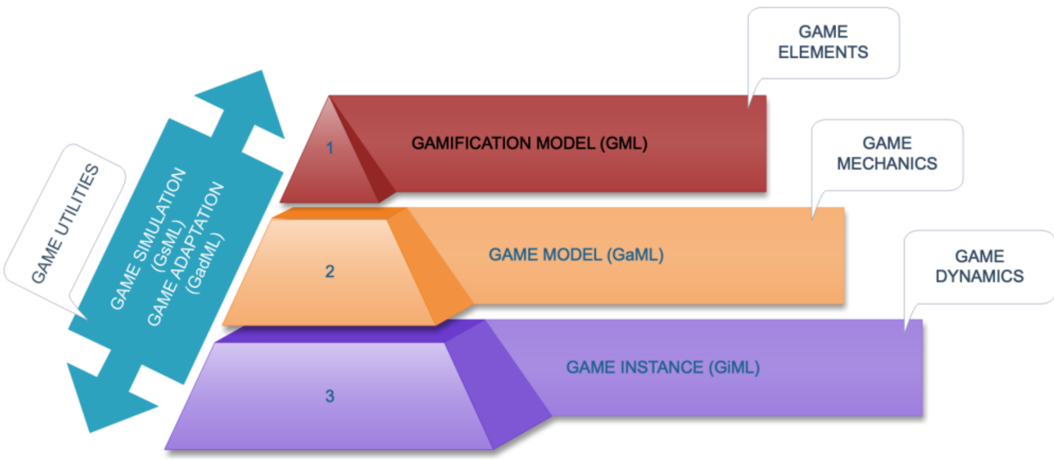


Figure 3.5: Gamification Design Framework (GDF), from Bucchiarone et al. (2019b).

Despite certain shared themes, the frameworks discussed exhibit a notable diversity in the realm of gameful systems. Therefore, the subsequent section outlines outstanding challenges that illuminate the ongoing endeavors required to enhance the design of gameful systems.

3.10.2 Open Issues in Gamification Design

Overall, it can be observed that the frameworks in the literature have a high degree of heterogeneity, with varying levels of theoretical depth and divergent conceptualizations, primarily stemming from different contexts and areas of interest (Mora, Riera, et al., 2015; Mora et al., 2017). Additionally, many existing references tend to emphasize ad-hoc experiences rather than formal design processes. While there are commonalities across these frameworks, the heterogeneity in gameful system design persists. This diversity leads to some challenges in comparing gamified systems and understanding the importance of specific factors in gamification outcomes. Additionally, most of these frameworks have their roots in game design, which operates under a different theoretical background and may not fully address onboarding and endgame processes (Mora et al., 2017). The great heterogeneity presented, along with the different areas of uses and concepts underlying the frameworks, is not necessarily a negative factor. However, gamification design frameworks, despite their potential, often exhibit low reliability due to the lack of standardization, a scarcity of validation, and difficulties in supporting the entire design process, including playtesting and evaluation (Mora et al., 2017). Addressing these issues — also reported in Section 3 — is crucial for advancing gamification design.

According to the review of Mora et al. (2017), gamification design frameworks can be divided into three main approaches to the design process:

- **User-centered**: users and their goals are the central focus during the design procedure.
- **Game-centered**: game design, elements, and artifacts are the central focus during the design procedure.
- **Technology-centered**: technological artifacts are the central focus during the design procedure. They are focused on the definition of architectures and systems as the basis of any gamification design and development.

These approaches allow some challenges, already presented in Chapter 3, to be addressed. However, none of the design frameworks attempt to consider all of them comprehensively. For instance, existing user-centered design frameworks can make designers pay the appropriate attention to the *interpersonal differences*, *goals*, and *game modality*; however, they may fail to address other *contextual* or *technological* challenges. Indeed, designers should not be based only on certain aspects of users, technology, or games, but should consider all the challenges depicted in Chapter 3, all the way from the initial design to the evaluation and monitoring (as presented in the model of Morschheuser, Hamari, Werder, and Abe (2017)). Moreover, researchers need to focus on the impact of these factors on the effectiveness of gamification. In fact, design frameworks should adequately guide designers during the process so that methodological rigor can be improved to facilitate the generalization and comparison of findings. Table 3.1 depicts what challenges have been addressed by the design frameworks described in Section 3.10.2).

Challenges	**Frameworks**
Contextual differences	Hunter and Werbach (2012); Klock et al. (2016); Morschheuser, Hamari, Werder, and Abe (2017).
Interpersonal differences	Hunter and Werbach (2012); Klock et al. (2016); Mora, Zaharias, et al. (2015); Morschheuser, Hamari, Werder, and Abe (2017).
Goal differences	Klock et al. (2016); Morschheuser, Hamari, Werder, and Abe (2017).
Game modality	Bucchiarone et al. (2019b); Chou (2019); Hunicke et al. (2004).
Technology	Klock et al. (2016).
Feedback	**Not addressed**.
Game elements	Chou (2019); Hunicke et al. (2004); Klock et al. (2016).
Methodological rigor	**Not addressed**.
Data collection	**Partially addressed by** Hunter and Werbach (2012); Klock et al. (2016); Mora, Zaharias, et al. (2015); Morschheuser, Hamari, Werder, and Abe (2017).

Table 3.1: List of challenges addressed by the most widely used design frameworks.

To address these challenges, the GamiDOC tool (Bassanelli & Bucchiarone, 2022) has been developed. The gamification design framework aims to guide designers in creating gameful systems while considering individual, demographic, and cultural variations, analyzing contextual and technological information, and guiding during the rules creation and game elements selections. Then, a gamification design document (GDD) helps the team with communication, speeding up timelines, and allows for monitoring of the technological components to help during the system requirements process. With the features under development, GamiDOC will provide a holistic approach to gameful system design, by including a design review phase — which promotes the use of preregistrations and registered reports to enhance methodological rigor and transparency —, an evaluation and monitor section to encourage the use of more rigorous methodologies, and an open access database to help users selecting the best fitting game elements. This approach aims to create a more systematic and reliable foundation for the design of gameful systems.

3.11 Conclusion

This chapter introduced the challenges that still exist in gamification design, including interpersonal and goal differences, game modality used, contextual factors, technological features, how to provide feedback, the selection of game elements, the use of good methodologies for data collection and analysis, and the collection of adequate data. Then, we introduced some gamification design frameworks that are used in the literature, and the open issues that still exist.

Overall, the elements to be considered during the design phase are many, and it is difficult to take into account every single variable to be analyzed. In the literature, many alternatives have been produced to guide during the design phase, however, there are currently no tools that take into account all the challenges we have identified. This prompted us to develop GamiDOC, a holistic solution that can guide throughout the entire journey, starting from design to final system evaluation. In the next chapter, GamiDOC will be described in detail.

Chapter 4

GamiDOC

To assist gamification designers throughout the design process, development, and user experience assessment, while addressing the challenges associated with gameful systems design and resolving the unresolved matters outlined in preceding sections, we have introduced a web-based tool known as GamiDOC[1], firstly described in *"GamiDOC: A Tool for Designing and Evaluating Gamified Solutions"* (Bassanelli & Bucchiarone, 2022)).

The creation of this tool followed a Design Science Research Methodology (DSRM) approach (which steps are depicted in Figure 4.1). DSRM (Peffers et al., 2007) serves as a robust approach enabling systematic generation and refinement of innovative solutions for challenges within the realm of Information Systems (IS) applications. Employing an iterative framework, DSRM empowers developers to craft artifacts for addressing specific problems. Our ongoing project, GamiDOC, aligns with the DSRM process to tackle the complexities of gamification.

4.1 Identify problem and motivate

The initial step in this journey involved a comprehensive review of pertinent literature to address **RQ 1**, and pinpoint challenges. We drew insights from the works of Koivisto and Hamari (2019b), Seaborn and Fels (2015), and Mora et al. (2017), who meticulously documented the hurdles encountered in gameful systems design, develop-

[1] `gamidoc.com`.

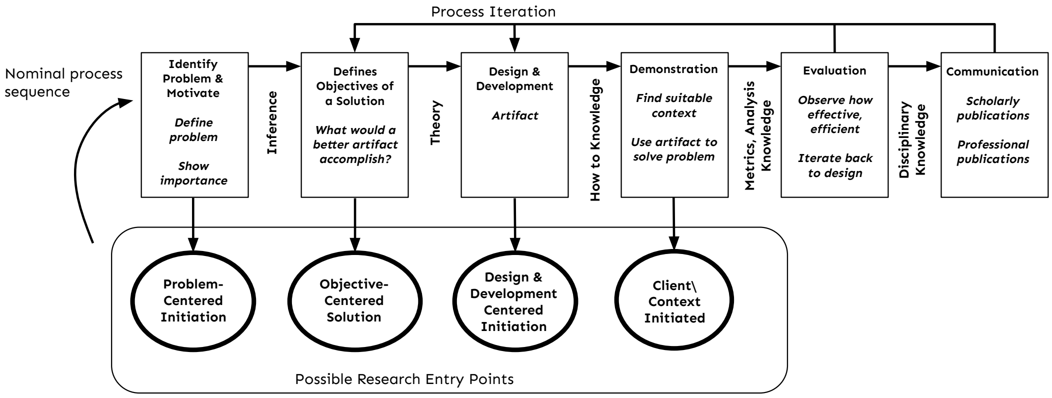


Figure 4.1: Design science research methodology (DSRM), rearranged from (Peffers et al., 2007).

ment, and implementation. Following their findings, we first analyzed co-citation and co-occurrence clusters of documents, authors, and keywords — as depicted in Chapter 2 —, identifying how the current literature has begun to move away from previously used design methods, looking for viable alternatives. Then, as presented in Chapter 3, we have comprehensively identified the elements that compromise the design of gameful systems, while also presenting the open issues that need to be addressed. Having these issues in mind, we went on to define an artifact that could address all the challenges presented.

4.2 Define objectives of a solution

Then, starting from the challenges identified, we engaged in a series of brainstorming sessions and interviews with prominent experts in the field to elucidate potential solutions. Our ideal final solution should guide designers and developers through the whole process, starting from the initial design, thus addressing the challenges related to the *target users*, *context*, *desired behaviors*, *technological constraints*, guiding during the *selection of game elements* and *feedback*, and allowing designers and developers to *com-*

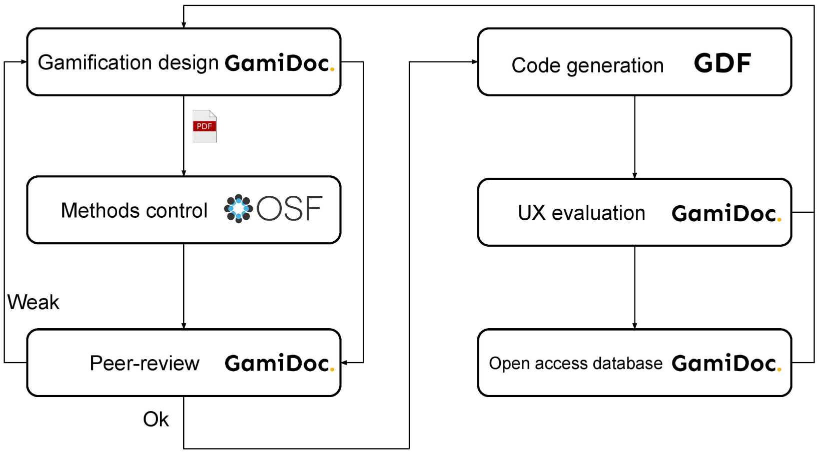


Figure 4.2: GamiDOC features in the overall process.

municate transparently through documentation that also took into consideration the *rules of the game* and the *aesthetic component.* Moreover, we identified other critical objectives to include, such as the facilitation of translation *from rules to programming language*, a possible *control over the final design* before the development, and the *guidance during the data collection and data analysis* phases. Overall, the final solution should pose as a viable solution for guiding entire teams in the design, development, and evaluation of gameful systems.

4.3 Design and development

As the optimal solution, a single platform containing several features to guide and help during the different stages of the development of gameful systems was identified. The processes supported by our solution are depicted in Figure 4.2. I present then each feature, describing the design, and the developmental progress state.

4.3.1 Design framework

We initially set out to create a design framework that would guide designers through the analysis of `contextual information`, such as *application domain*, *tool's aim*, *target users need and aim*, and *encouraged and discouraged behaviors.* Sequentially, through the analysis of `technological information`, such as software, and devices, and pos-

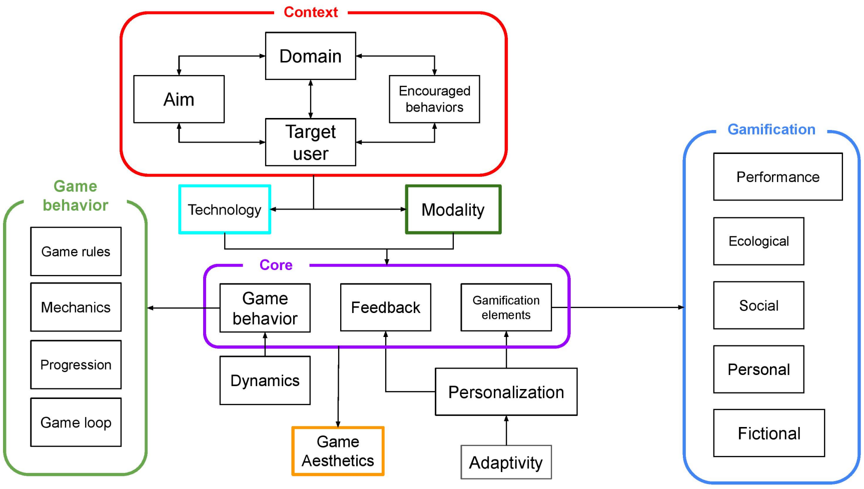


Figure 4.3: Gamification design model components.

sible `game modalities`. Then through the analysis of `core information`, such as *game elements selection* and their personalization, *feedback design*, and the *game rules, mechanics, dynamics, progression, and loop*. Lastly, through the analysis of the `game aesthetic`, including the graphic style. A graphical representation of the design framework is depicted in Figure 4.3.

The first implementation of the model was in Nuclino[2], an online tool for unified working space. However, the need to create an efficient GDD that would be integrated with the other features led to the creation of a stand-alone online website and the creation of some constraints within the different sections of the framework. In the following sections, we present a description of each component.

4.3.1.1 Context

Within the gamification design model, the context component holds immense significance. It permits the utilization of contextual details to tailor the design for each gameful system. This component gathers information about the *application domain*, *software aim*, *target user*, and the *behaviors to be encouraged*. Collecting contextual information is essential to customize recommendations for gamification elements, drawing from an accessible database.

As detailed in Section 3.1, gamification elements may operate differently based

[2] `https://www.nuclino.com/`.

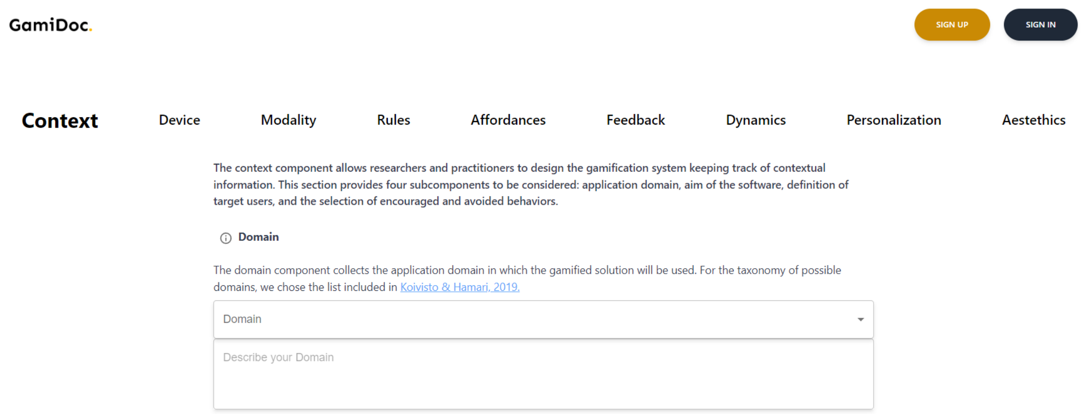


Figure 4.4: Domain feature in GamiDOC.

on the application domain (Kim, 2015; Koivisto & Hamari, 2019b). At present, the **domain** subcomponent employs the taxonomy established by Koivisto and Hamari (2019b) to classify the application domain where the gamified solution will be implemented. Thus, users can select one of the following categories: Education/learning, Health/exercise, Crowdsourcing, Social behavior/networking/sharing, Software development/design, Business/management, Ecological/environmental behavior, eCommerce/eServices, Software engineering, Marketing/consumer behavior, Citizen/public engagement/activity, Entertainment, Innovation, Transportation/mobility, Culture/tourism, Architecture, Communication, Emergency planning, Politics, Welfare/city/human/public services, Work, Theory. Then, after selecting the most suitable domain, users can describe in detail the application domain in the appropriate box (Figure 4.4).

The **aim** subcomponent gathers insights into the purpose of the designed software and the rationale behind its creation. In numerous gamification projects, the objectives were not clearly defined at the project's outset (Kim, 2015). The objective of a gamified solution could range from increasing student enrollment to fostering collaboration among users or even promoting a shift toward environmentally sustainable behaviors (see Section 3.3). In this part, GamiDOC gathers information about the purpose of the gameful system based on the taxonomy of Fortes Tondello et al. (2018): (i) **Outcome goal** — which refers to the accomplishment of a very specific result; (ii) **Performance goal** — which refers to doing well by one's performance standards; (iii) **Process or**

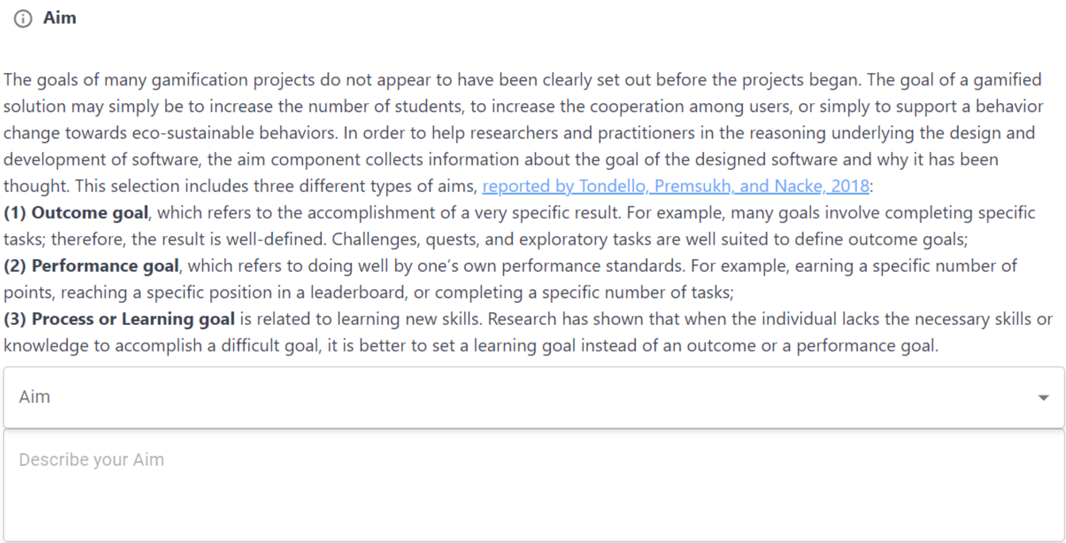


Figure 4.5: Aim feature in GamiDOC.

Learning goal — which is related to learning new skills. Users in this case have the opportunity to select one of the three elements and describe in more detail the ultimate purpose of the tool (Figure 4.5).

The **encouraged behaviors** subcomponent captures information about the behaviors designers intend to promote or discourage. This feature can be considered one of the fundamental factors of gamification (Bassanelli et al., 2022). Unlike the other components of GamiDOC, this one does not contain a categorized selection because there is no evidence in the literature reporting possible categories of behaviors to be encouraged and discouraged in gameful systems. For this reason, the feature allows designers who use GamiDOC to consider this element during the design phase by describing in detail the behaviors they wish to encourage through the use of the application (Figure 4.6).

The **target user** subcomponent compiles pertinent data concerning the users who will engage with the software, including categories, age ranges, and more. As outlined in Section 3.2, various sources within the gamification field emphasize the importance of individual differences and user preferences for the success of gamified solutions (Kim, 2015; Koivisto & Hamari, 2019b; Tondello et al., 2016; Zahedi et al., 2021). At the current state, designers can report the application's target user by selecting a range of age and describing the possible category of target user (Figure 4.7). As presented in

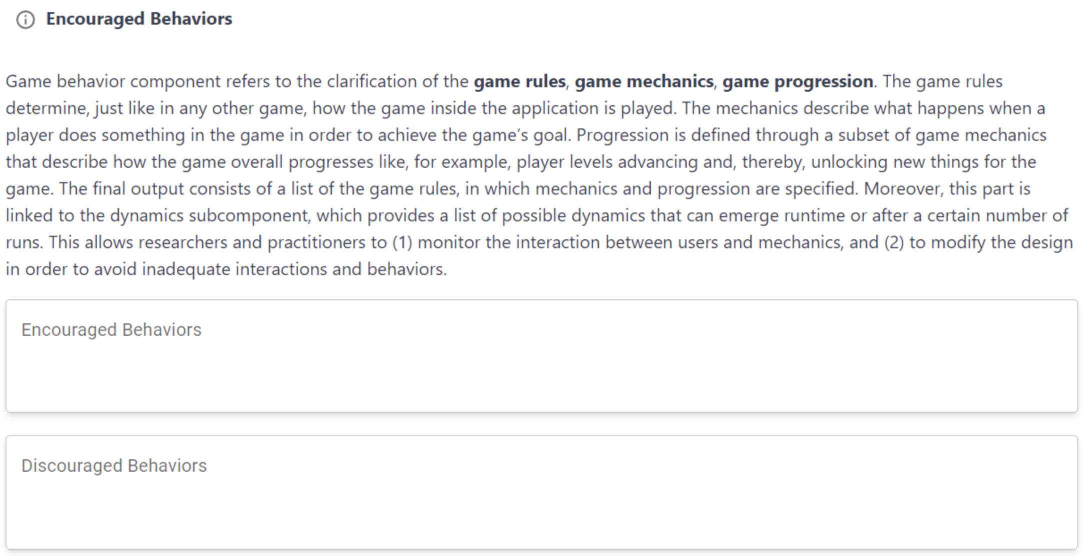


Figure 4.6: Encouraged and discouraged behaviors feature in GamiDOC.

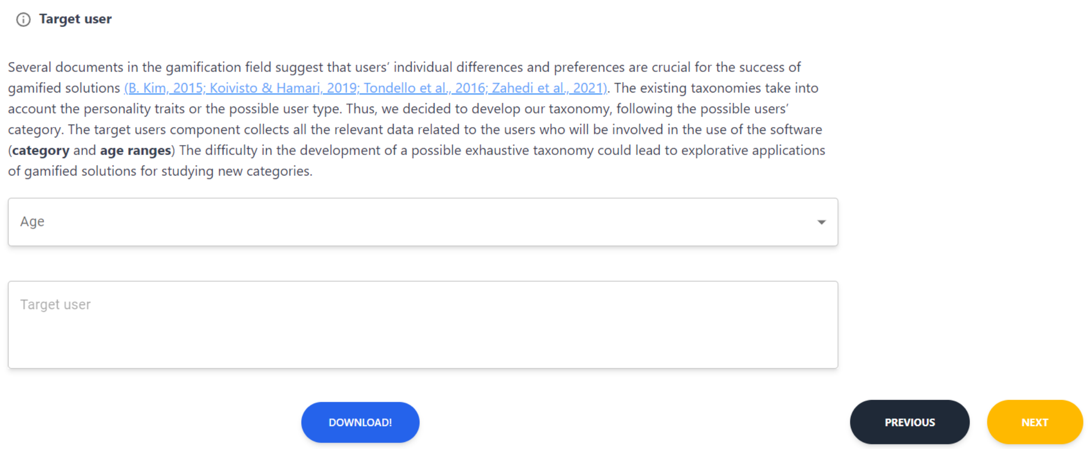


Figure 4.7: Target users feature in GamiDOC.

Chapter 7, the final aim of GamiDOC for this feature is to present reliable pre-existing categories of target users to allow personalization strategies based on these categories. This analysis is currently ongoing.

4.3.1.2 Techology

The technological aspect of gamification entails gathering details about the hardware, software, input/output devices, and other technological constituents utilized during software development. It is vital to take into account the specific device on which the software will be deployed, ensuring that the chosen gamification elements align opti-

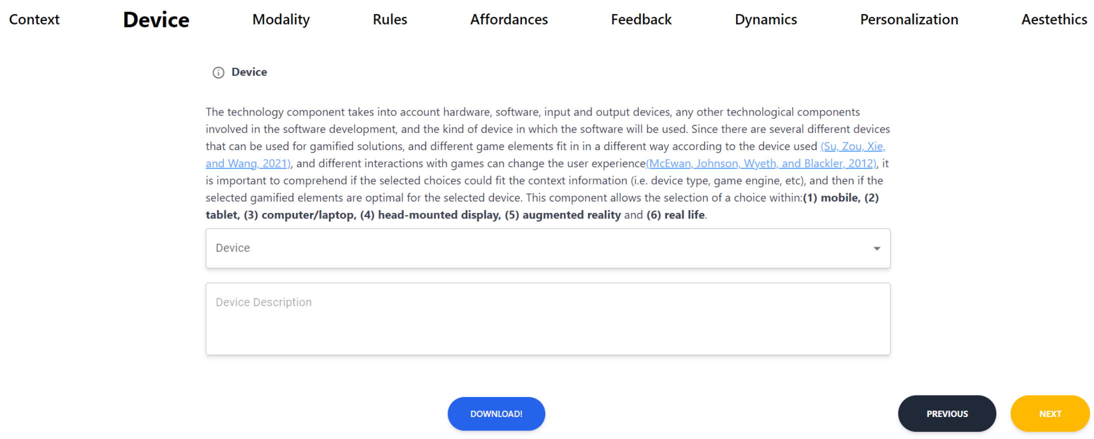


Figure 4.8: Device feature in GamiDOC.

mally with that device. This facet enables the selection of diverse device categories, encompassing mobile phones, tablets, computers, head-mounted displays, augmented reality devices, and real-world settings. A fundamental understanding of whether the technology selections align with the context of the gamified solution is crucial, including considerations of the game engine and device category. In this section, designers can report the different devices and technologies with which the gameful system is expected to be used, such as mobile, computer/laptop, tablet, head-mounted display, augmented reality, or non-digital. Moreover, an empty box allows users to give additional information about the device (Figure 4.8).

4.3.1.3 Modality

This section is devoted to establishing the primary mode integrated within the gamification software. Despite limited evidence on cooperative gamification, existing literature demonstrates distinctions between cooperative and competitive modalities of gamification, and how various game elements support these different modes (Y. Chen & Pu, 2014; Goh & Lee, 2011; Lee et al., 2013; Massung et al., 2013; Mekler, Brühlmann, Opwis, & Tuch, 2013; Morschheuser, Maedche, & Walter, 2017). To classify different gamification modes, we adopted the categorization proposed by Morschheuser, Maedche, and Walter (2017): (i) **individual**, (ii) **cooperative**, (iii) **competitive**, and (iv) **cooperative-competitive** gamification features (also elaborated in Liu et al. (2013)). The latter pertains to elements that facilitate users' collaboration within their group

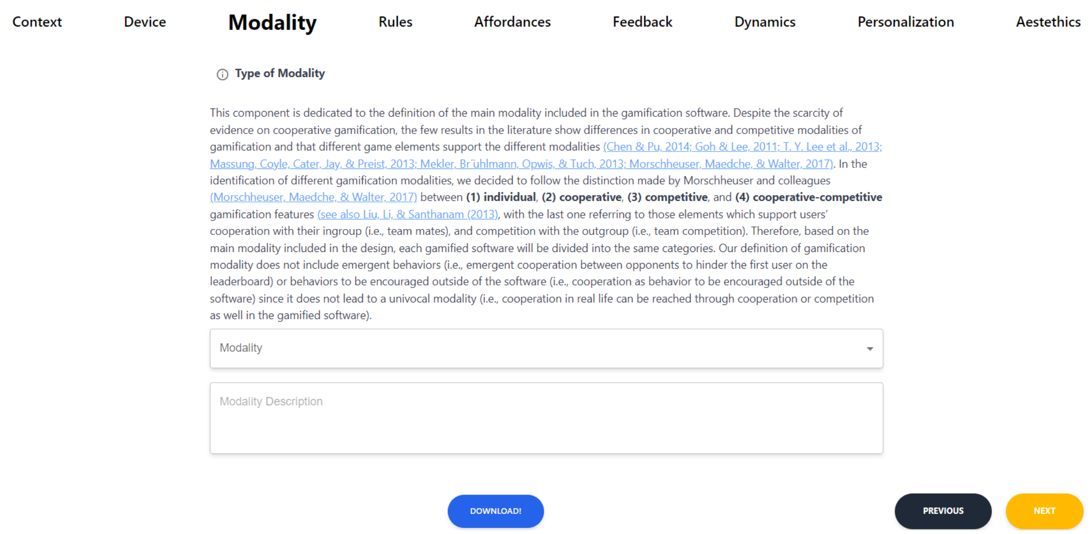


Figure 4.9: Modality feature in GamiDOC.

(e.g., team members) while fostering competition with external groups (e.g., team-based competition). Thus, the gamified software is categorized based on its primary modality. Our definition of gamification modality excludes emergent behaviors (e.g., opponents cooperating to hinder the leading user on the leaderboard) or behaviors to be promoted outside the software (e.g., encouraging cooperation beyond the software), as they don't lead to a distinct modality (e.g., real-life cooperation can be achieved through cooperation or competition within the gamified software). In this part of the tool, users report the gameful systems' modality and can describe it in detail (Figure 4.9).

4.3.1.4 Core

As presented in Figure 4.3, the **core** part is composed of three different parts: (i) **game behavior**, (ii) **game elements**, and (iii) **feedback**.

4.3.1.4.1 Game behavior

The **game behavior** subcomponent encompasses the game's rules and mechanics. Game rules, akin to any game, dictate in-game interactions within the application. Mechanics delineate outcomes when players take actions to accomplish the game's objectives. Progression is characterized by a subset of game mechanics that describe

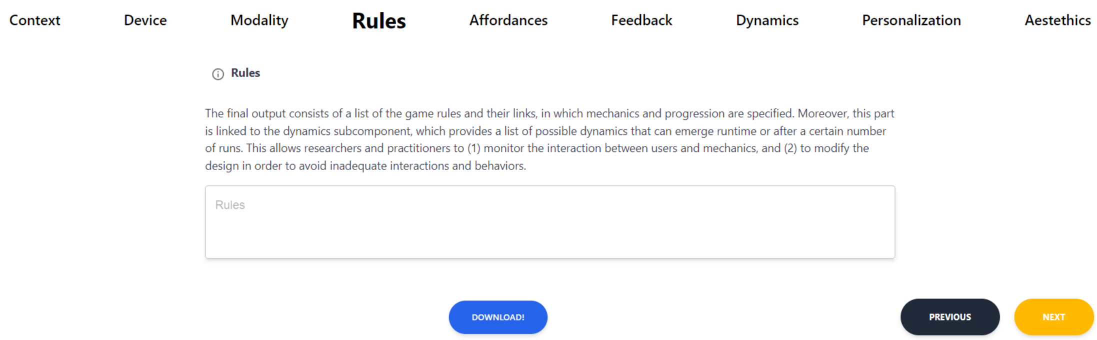


Figure 4.10: Rules feature in GamiDOC.

overall game advancement — like player leveling and unlocking new game elements. The outcome is a comprehensive list of game rules, specifying mechanics and progression. In this part, designers describe the whole game's rules. This allows them to monitor the interaction between users and mechanics, and to modify the design to avoid inadequate interactions and behaviors. The result is a list of primary mechanics, secondary mechanics, and their interactions (Figure 4.10).

Additionally, this section connects with the **dynamics** subcomponent, which presents a list of potential dynamics that can emerge during runtime or after a specific number of iterations. This empowers researchers and practitioners to (i) monitor user-mechanic interactions and (ii) refine the design to prevent undesirable interactions and behaviors.

4.3.1.4.2 Game elements

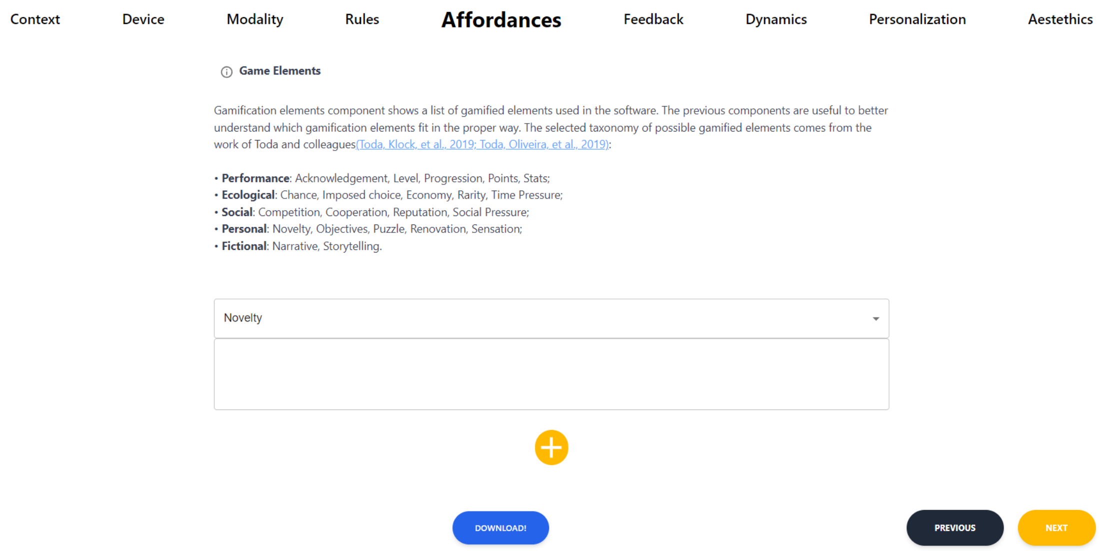
The **game elements** subcomponent records details about the utilized gamification elements, drawn from the taxonomy proposed in A. M. Toda, Klock, et al. (2019) and A. M. Toda, Oliveira, et al. (2019). Hence, designers can choose all the game elements used in the gameful system, describing them and their interactions. The taxonomy reports game elements divided into specific dimensions. The **Performance/measurement** dimension is composed of elements related to the environment response, which can be used to provide feedback to the learner, such as *Acknowledgement* (badges, medals, trophies, and achievements), *Level* (skill level, character level), *Progression* (progress bars, steps, maps), *Points* (scores, experience points, skill points), and *Stats* (information, Head Up Display (HUD) and data). The **Ecological** dimension is composed of elements that are linked to the environment in which gamification is implemented. These elements are mainly defined as properties, such as *Chance* (randomness, luck, fortune, or probability), *Imposed choice* (choice, judgment, and paths), *Economy* (transactions, market, exchange), *Rarity* (limited items, collection, exclusivity), and *Time Pressure* (countdown timers or clocks). The **Social** dimension is related to the interactions between the learners presented in the environment, such as *Competition* (conflict, leader boards, scoreboards), *Cooperation* (teamwork, co-op, groups), *Reputation* (classification, status), and *Social Pressure* (peer pressure or guild missions). The **Personal** dimension consists of elements that are related to the learner that is using the environment, such as *Novelty* (update, surprise, changes), *Objectives* (missions, side-quests, milestones), *Puzzle* (challenges, cognitive tasks, actual puzzles), *Renovation* (boosts, extra life, renewal), and *Sensation* (visual or sound stimulation). The last dimension, **Fictional** is composed of elements related to the user (through *Narrative*) and the environment (through *Storytelling*), tying their experience with the context. At the current state of development, the game elements (called **affordances** in the tool), are merely a list that provides a guideline for developers (Figure 4.11). As presented in Section 7, in the future, their selection will be reported in an additional feature that already contains a list of selected elements and will allow code to be produced from natural language (see Section 4.3.4).

The **personalization** section, which is linked to the game elements and the feedback, collects information on specific personalization techniques used in the gameful system to personalize feedback or game elements, such as the user type analysis with specific scales (Nacke et al., 2014; Roccas et al., 2002; Tondello et al., 2016).

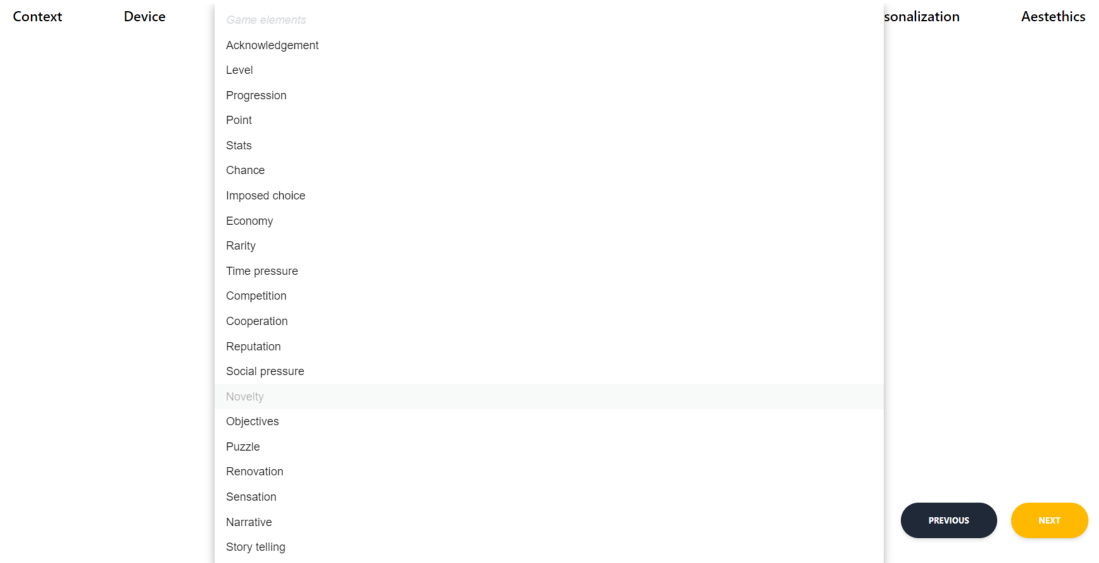
4.3.1.4.3 Feedback

The **feedback** subcomponent describes how the system provides feedback to users. The first version of this component, depicted in Figure 4.12, relied only on *timing* (e.g., immediate, delayed, personalized) and *content* (e.g., corrective, explanatory, reporting, personalized).

After the first use of the tool (Section 5.1.1), we noticed a lack of clarity and alternatives in feedback design; thus, we decided to develop a more structured feedback design



(a) Game elements section in GamiDOC.



(b) Game elements selection.

Figure 4.11: Game elements section and selection.

framework within GamiDOC. This framework, depicted in Figure 4.13, consists of a linear flow that allows users to specify different components of feedback and has been evaluated in Section 5.2. The first element is the selection of **the part of the process in which designers want to provide feedback**, which refers to the selection of a different part of the learning path: (i) learning, (ii) practice, or (iii) assessment. Later, with **"what do you prefer to communicate to the user?"**, it collects information on the information aim, such as (i) feedback, information on performances/paths, or

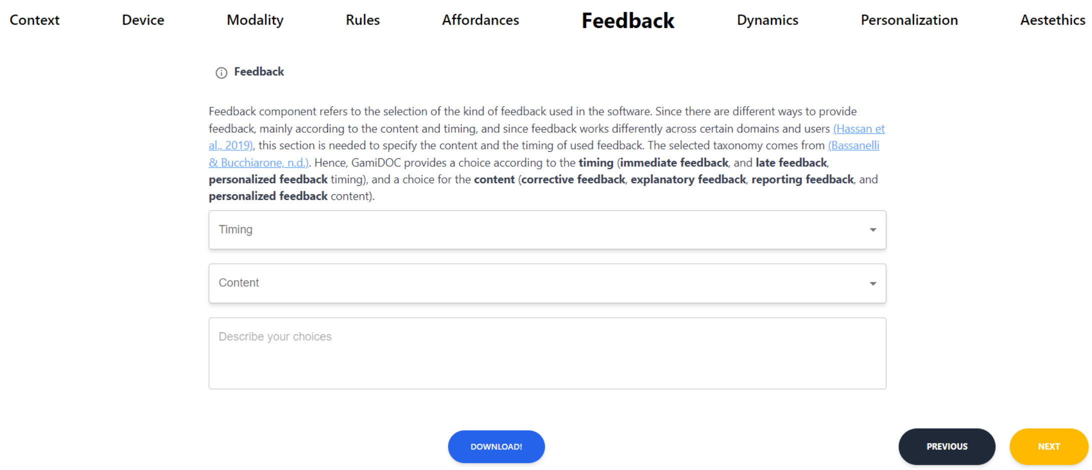


Figure 4.12: First feedback feature in GamiDOC.

(ii) feedforward, guidance for future performances/paths. After communicating the feedback aim, the framework guides through the feedback specification, with the description of the **feedback typology**, the specification of the **feedback content**, the **most suitable timing**, and the **best modality**. The feedback typology can be identified as positive, negative, and mixed. Positive refers to emphasizing the elements that went well during an activity, negative emphasizes the elements that could be improved upon, while the mixed one takes into consideration both elements that went well in an activity and, at the same time, identifies elements that need improvement. The feedback content refers to the ultimate meaning of the feedback. It comprehends *Evaluative* contents (evaluate the effectiveness of users' decisions), *Descriptive* contents (offer users detailed and specific information on how to improve; it is directly connected to the learning objectives, targeting any misconceptions or areas of insufficient understanding), *Constructive* contents (provide specific observations, suggestions, and recommendations to help an individual or a system enhance their performance, skills, or behaviors), *Corrective* contents (highlight and rectify errors, misconceptions, or performance deficiencies), *Appreciative* contents (provide insights with an emphasis on enhancing effectiveness and productivity in future endeavors rather than assigning blame for past actions), or *Tailored* contents (involve customization of the feedback content to suit the individual). The feedback timing comprehends three different timing: *Direct feedback* (provided immediately), *Delayed feedback* (provided later), and *Tailored* (per-

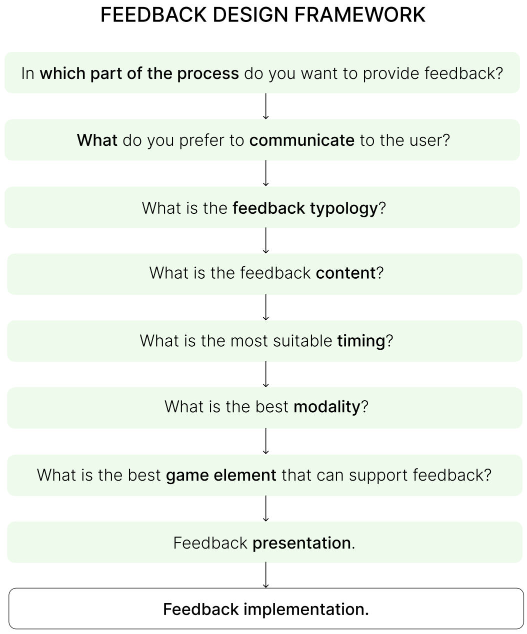


Figure 4.13: Feedback design framework.

sonalized for specific users or conditions. The feedback modality comprehends *written*, *visual*, *auditory*, and *multimodal*.

After selecting the feedback structure, the framework relates the feedback structure just described to the game elements taxonomy depicted in the game design framework (A. M. Toda, Oliveira, et al., 2019) to help the designer identify game elements that can support feedback in the game. The game elements can either replace feedback in its entirety or flank it. Indeed, in the later parts of the framework, the user is asked to identify how feedback is presented in detail.

4.3.1.5 Aesthetics

The **game aesthetics** segment encompasses all details of the visual aspect of the software and how these relate to the emotional aspects elicited. Aesthetics directly influence user experience: heightened visual appeal corresponds to increased user interest and captivation. Consequently, this elevation translates to enhanced motivation and engagement levels (Schell, 2008; Walk, Görlich, & Barrett, 2017). In this section, researchers and practitioners are required to document choices and provide examples of the visual facet and decisions regarding the user interface through an upload procedure.

4.3.2 Gamification design document

We subsequently decided to allow the automatic creation of a GDD containing all the information entered following the framework to help in the communication between design and development[3], thus helping the team in the selection of system requirements. This design document is a reference document that is helpful for members of the development team: they should constantly be using the document to find specific information for their specific needs (Baldwin, 2005). It's important that a GDD does not cause information overload and, for that reason, be shelved by the development team. Hence, the GDD must be organized and formatted to make it easy to use.

4.3.3 Methods control and Peer-review process

To optimize the effectiveness of the developed tool, and thus reduce the waste of resources and time, we came up with a feature containing control over the gameful systems design and methods design through a submission with a peer-review process. Although there is a gamification design framework within the tool that creates a GDD, it is possible to submit a gamification design document following any type of existing design framework. Then, users registered to GamiDOC may be selected for review based on their expertise in the field. Along with the document, preregistration of data collection and analysis may be attached. GDDs evaluated positively can then be posted in GamiDOC and published in the database, allowing a possible public search by keywords; if preregistration is present, a special mark is placed on the document, indicating that the modality of data collection and analysis has been validated. Finally, GamiDOC shows a board with the gamification design documents that have been approved on the homepage. It will be possible to scroll through them by keyword, authors, and other factors (such as domain, target users, and so forth). The board will show also the application results of the developed software (if possible).

This feature exists in the back end of the tool but has not yet been implemented because the peer-review condition carries with it certain concerns. Indeed, the simple final design and development of this feature can represent a Ph.D. project itself:

[3]A video of the use of GamiDOC and a GDD can be retrieved here: https://osf.io/zvayp/?view_only=5973f67e5f8743468f09809a2786b654.

- How can we assess properly the expertise in gameful systems design?
- Where can we allocate the database for the collection of submitted documentation?
- Which organization can take on the ethical constraints that the collection of personal data entails?
- What method of ascertainment and authentication can be used to determine the identity of members?

4.3.4 Code Generation

As gamification continues to gain traction across various application domains and reaches a broader user base, the complexity of gameful software is inevitably on the rise. Consequently, a significant challenge arises in bridging the abstraction gap between design and implementation (Bucchiarone, Martella, Muccini, & Fusco, 2023). The implementation phase becomes more arduous and error-prone due to the increasing rules and customization requirements. Additionally, managing maintenance and evolution activities becomes more challenging due to the disconnect between design and implementation.

To address this issue and narrow the gap between the design and implementation of gameful applications, we leverage the Gamification Design Framework (GDF) (Bucchiarone, Cicchetti, & Marconi, 2019a; Bucchiarone et al., 2019b). GDF comprises a set of domain-specific languages (DSLs) dedicated to specifying, implementing, and deploying gameful applications. This framework has been developed based on three fundamental principles:

- **Separation-of-concerns**: As the complexity of gamification grows, an effective way to manage it is by treating various aspects as separate viewpoints that can later be integrated into a comprehensive solution.
- **Correctness-by-construction**: With the increasing intricacy of gamification applications and their diverse user base, it becomes crucial to ensure that game rules align with the mechanisms and elements intended for the target application.

- **Automation**: To minimize the gap between design and implementation, manual code writing should be minimized, and the framework should provide a high degree of automation.

GDF embodies these principles through three DSLs that correspond to three abstraction layers through which any gamified application can be viewed: (i) The topmost layer defines general mechanics and elements that a solution may include, such as points, bonuses, challenges, and more; (ii) the second layer instantiates a subset of the abstract concepts defined in the layer above. It does this by specifying the particular gameful application under development, including details like the number of steps, "walker of the week", "hundred thousand steps week", and so on, in systems for green transportation. (iii) The third and bottom layer describes the implementation of the concepts above together with their deployment on a gamification engine. Configuration parameters can be set, like point thresholds, bonuses, challenge activation timing, and player or team assignments. Furthermore, these layers provide generators for automating the derivation of implementation code for the gamified application. In summary, GDF provides a structured and automated approach for designing and implementing gameful applications, helping to manage complexity and ensure alignment between design and realization.

Starting from GDF, Bucchiarone, Martella, et al. (2023) developed DSL4GaR — Domain Specific Language for Gamification Rules —, a domain-specific language for gamification rules definition, simulation, and deployment. DSL4GaR consists of three main modules:

- **Rule Generator**: It provides a collection of APIs that are valuable for defining a rule and can be broken down into two separate sub-modules: (i) **LHS generator**, a set of APIs to properly define the conditions for a rule to be executed; and (ii) **RHS generator**, a set of APIs to build the consequences of a rule when it is executed.
- **Rule Simulator**: It offers a range of APIs designed for simulating specified rules. Each simulation consists of an initial state, the applied rules, and all the state changes observed during the simulation process. These state changes are

illustrated through a graph, which helps assess the impact of each rule applied during the simulation.

- **Rule Deployer**: Using it enables interaction with the gamification engine and deployment of all tested rules once they have been simulated.

DSL4GaR functions as a tool designed to simplify the rule's lifecycle process for non-technical users. It offers user-friendly APIs within an integrated development environment (IDE), ensuring syntax highlights and auto-completion features while restricting user inputs to minimize errors. Additionally, DSL4GaR provides a structured method for testing rules and validating their outcomes. Designers can initiate simulations from a defined initial state of the game, monitoring the evolution of the state and rule interactions through a graph. Each node in the graph represents a game state, and each edge signifies the executed rule that transforms the system into a new state. During simulation, designers can define a list of assertions automatically verified during the process. If unexpected occurrences arise, the graph highlights encountered errors. Once rules are defined and tested, they can be deployed on the gamification engine (GE), the software responsible for their execution, allowing end-users to engage with the gameful system. DSL4GaR facilitates direct interaction with the GE, enabling users to deploy, update, or delete rules effortlessly.

We therefore decided to implement a direct link between the game rules and DSL4GaR to speed up code creation during development. However, this feature needs to specify the links between the various game elements and game rules before using DSL4GaR. Currently, we are designing a user interface that can easily be used directly within GamiDOC.

4.3.5 User experience

To guide designers in the selection of adequate evaluative methodologies and evaluation items, thus improving the quality of analysis (addressing the challenge depicted in Section 3.8), we designed a feature of GamiDOC which provides a list of different evaluation methodologies and standardized questionnaires (primarily for user experience) for both qualitative and quantitative data, along with the presentation of several important pieces of information for data collection and analysis (including sample size, effect size, and so forth). At the moment we have included two main sections: **qualitative methods**, including *Observation* (Lofland, Snow, Anderson, & Lofland, 2022), *Interviews* (Brooks, Horrocks, & King, 2018), *Focus groups* (Morgan, 1996), *Surveys* (Braun, Clarke, Boulton, Davey, & McEvoy, 2021), and *Secondary research* (Largan & Morris, 2019); **standardized questionnaires**, including *GAMEFULQUEST* (Högberg, Hamari, & Wästlund, 2019), *User Experience Questionnaire* (Schrepp, 2015), *Igroup Presence Questionnaire* (Vasconcelos-Raposo et al., 2016), *FunQ* (Tisza & Markopoulos, 2021a), *System Usability Scale* (Lewis, 2018), *Health-ITUES* (Yen, Wantland, & Bakken, 2010), *meCUE* (Minge, Thüring, Wagner, & Kuhr, 2017), *miniPXI* (Haider et al., 2022), *Player Experience of Need Satisfaction questionnaire (PENS)* (D. Johnson, Gardner, & Perry, 2018), *Game Experience Questionnaire (GEQ)* (IJsselsteijn, De Kort, & Poels, 2013), *Intrinsic Motivation Inventory (IMI)* (McAuley, Duncan, & Tammen, 1989), *User Motivation Inventory (UMI)* (Brühlmann, Vollenwyder, Opwis, & Mekler, 2018), *AttrakDIFF* (Ströckl, Oberrauner, Krainer, Wohofsky, & Oberzaucher, 2020), and two questionnaire developed by the team that are now under validation, *MEEGA360* (Gini & Bassanelli, 2022), and *AirBreak* (Bassanelli, Gini, Bucchiarone, & Marconi, 2024).

Currently, the feature design is complete, and we are developing a prototype. As a result of the development it will be possible to select the most suitable questionnaire according to needs, or compose a questionnaire based on the dimensions of the various questionnaires presented. Moreover, with a submission procedure, it will be possible to add new standardized questionnaires to the list, so that they can be used by users.

4.3.6 Open-access database

The selection of proper game elements can be a difficult part of design. So we thought we would tackle the challenge of properly reporting the effectiveness of game elements in several contexts. We thought the use of an open-access database that would allow users to observe the effectiveness of certain game elements in certain contexts could help (i) choose the best game elements for a gameful system, and (ii) produce a perpetual list of state of the art on gamification. Currently, we are exhaustively analyzing empirical documents in the field of gamification (with at least one control group in the study) to build the structure behind the database. After the final development, the database will interact with the contextual part of the model, suggesting the best game elements concerning the contextual info entered. Moreover, it will be possible to update the data in the database by uploading empirical data.

4.3.7 Design frameworks list

To make GamiDOC more holistic, we thought about the inclusion of an exhaustive list of gamification design frameworks and procedures to be included in a section of the tool that would serve as an analysis of the relevant literature and, when needed, provide users with an alternative to design procedures. Currently, we are still running a comprehensive systematic analysis. Before reaching the development stage, it will be necessary to implement a method to identify reliable design strategies and frameworks.

4.4 Ethical concern

As presented in Sections 3.9 and 3.10, while designing gameful systems it is crucial to face ethical concerns regarding the designed gameful system. The option in GamiDOC to specify in detail elements relating to domain, purpose, encouraged and hindered behaviors, additional information on technological components, game modality, and aesthetics, and the specification of the game dynamics, makes it possible to verify that there are no ethical problems, including dangers to users or the possibility of misuse of the tool. In addition, GamiDOC guidance and control features, namely methods control, peer-review process, and system evaluation allow further control on the ethical sustainability of systems produced with GamiDOC.

4.5 Conclusion

The solution identified through the DSRM approach considers all the challenges reported in Chapter 3 and attempts to fill the open issues (Section 3.10.2). However, the project can be considered extremely large, and it was not possible to complete the development of all artifact components during these 3 years. Table 4.1 provides an overview of the current state of the GamiDOC's features development.

Feature	Current state
Gamification design framework	***Done.*** The gamification design framework was fully developed and implemented in the online platform. ***To do.*** We just need to implement the new feedback design model, find a reliable categorization for target users to allow personalization strategies based on these categories, and address some usability problems listed in Chapter 5, including further examples of use.
Gamification design document	***Done.*** The creation of a GDD is possible from the compilation of the framework. ***To do.*** The page layout of the document still needs some improvement.
Gameful system design peer-review	***Done.*** The login component and the possibility to submit the final design of gameful systems as an attachment with evaluative methods exist in the back end. ***To do.*** Before the feature can be released into the tool, some ethical constraints must be resolved.
Code creation	***Done.*** Code creation from rule definition exists through an external tool called DSL4GaR. ***To do.*** DSL4GaR needs some refinements. Moreover, it is necessary to finalize the user interface that acts as a conduit between GamiGOC and DSL4GaR.
User experience	***Done.*** The design of this feature and the first prototype is ready. ***To do.*** We need to evaluate the prototype before proceeding to actual development.
Open-access database	***Done.*** We have finished the database design phase. ***To do.*** We need to complete an exhaustive and systematic analysis of empirical documents in the field of gamification (with at least one control group in the study) to build the structure behind the database; then, proceed to the development
Gamification design frameworks list	***Done.*** Feature design has been partially completed. ***To do.*** We need to complete an exhaustive and systematic analysis of gameful system frameworks, identifying reliable design strategies and frameworks. Then, proceed to the actual development.

Table 4.1: Overview of the current state of the GamiDOC's features development.

In the next Chapters, we present the final stages of the DSRM, namely **Demonstration**, which reports the use of the artifact to solve problems, and **Evaluation**, which reports observation on how effective and efficient the artifact is (and iterate back to the design to improve the tool. This document, along with Bassanelli and Bucchiarone (2022) and Bassanelli, Bucchiarone, and Gini (2024), represent the actual **Communication** stage.

Chapter 5

Evaluation

In this Chapter, I report studies that analyze the usability and usefulness of GamiDOC, related to both the gamification design framework and the online tool (**RQ 3**). The first two studies, which analyze usability, feature perception, and tool utility, respectively, are given in the paper *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024). As a result of the data obtained from these studies, the feedback component in GamiDOC was modified. The last study in this chapter presents an analysis of the new feedback design framework.

5.1 Design framework evaluation

5.1.1 First study

To conduct a comprehensive evaluation of the tool's usability, we employed the System Usability Scale[1] (SUS) (Brooke et al., 1996). We chose this tool due to its established reliability and widespread use in usability studies (Lewis, 2018), and because it presents an analysis methodology providing an overall value for the usability of a tool. Additionally, to gather insights into future developments for GamiDOC, we designed a custom questionnaire[2] comprising eight items that assessed the usefulness of GamiDOC features. Both questionnaires utilized a 5-point Likert-type scale, ranging from *"Strongly*

[1]The full scale and the related information can be retrieved here: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html.

[2]The full ad-hoc questionnaire can be retrieved here: https://osf.io/5cr9w.

disagree" to *"Strongly agree"*. We also included a final item to collect feedback and suggestions for enhancing the tool.

5.1.1.1 Participants and procedure

We recruited a total of 16 participants (10 males and 6 females) with experience in gamification design. Each participant utilized the tool online and completed the questionnaire on Qualtrics[3]. Existing literature (Hwang & Salvendy, 2010) recommends a sample size of 10 ± 2 for usability studies to detect at least 80% of usability issues. Among the participants, nine had limited experience in gamification design, while seven identified as experts in the field. Participants were recruited through online platforms such as Slack[4] and Discord[5] channels, as well as via email within gamification design courses at the University of Trento. We instructed participants to use GamiDOC's gamification design model to create a gamification design document. Subsequently, users were asked to respond to the online questionnaires. Participation in the study was voluntary.

5.1.1.2 Results

We analyzed questionnaire results using RStudio[6] (version 2022.07.2). The initial version of the tool achieved a SUS score of 65.67, slightly below the threshold of 68. This suggests the need for further improvements to enhance usability. The feedback from the questionnaire highlighted specific areas for improvement:

- **Hypertext connections**: Users encountered issues with hypertext connections, particularly when navigating to the bibliographic section, as it did not retain the information entered.
- **Information gaps**: Some elements in the model lacked sufficient information, leading to difficulties for users. Gamification elements and feedback selections were identified as needing more explanatory content.

[3]https://www.qualtrics.com/.
[4]https://slack.com/.
[5]https://discord.com/.
[6]https://posit.co/.

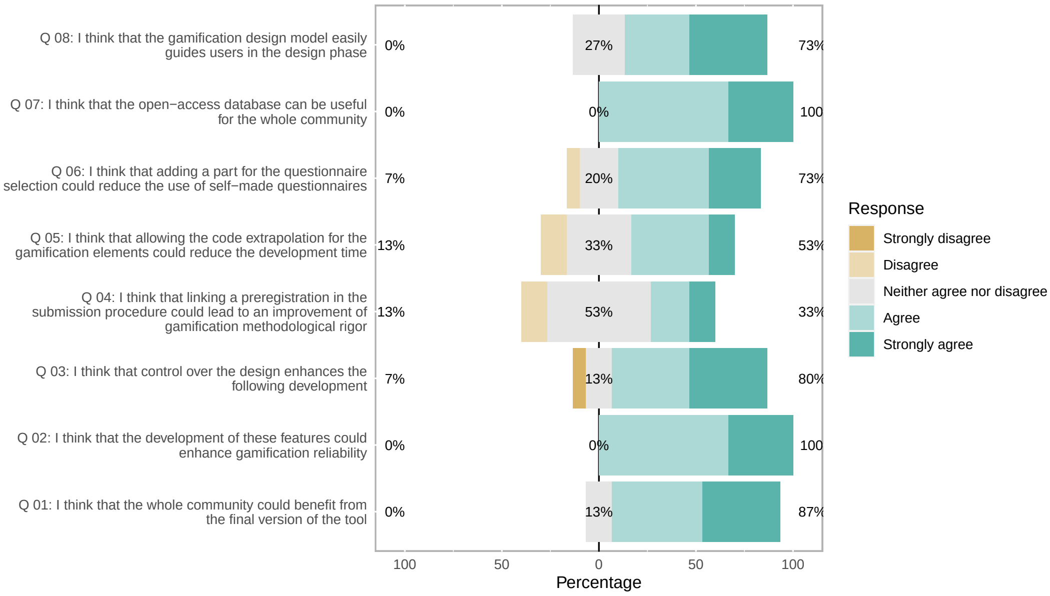


Figure 5.1: Usefulness of GamiDOC features expressed by users.

- **Lack of examples**: Users noted a general absence of examples within the model, which could have guided them in filling out the form effectively.
- **Hierarchy**: Users found it necessary to establish a well-defined hierarchy during the design phase to guide them through the top-down approach of the model.

Analyzing the SUS score concerning user experience in gamification design, less-experienced users reported a higher overall score (SUS score = 72.5) compared to expert users (SUS score = 60.71). Regarding the evaluation of the components under development, each item received high ratings (mean value = 3.91 ± 0.4). Notably, items such as *"I think that the development of these features could enhance gamification reliability"* (mean value = 4.3 ± 0.48), *"I think that the open-access database can be useful for the whole community"* (mean value = 4.3 ± 0.48), *"I think that the whole community could benefit from the final version of the tool"* (mean value = 4.23 ± 0.72), and *"I think that the gamification design model easily guides users in the design phase"* (mean value = 4.07 ± 0.86) received particularly positive responses. Interestingly, expert designers reported higher values (mean value = 4.23 ± 0.47) compared to users with less experience (mean value = 3.80 ± 0.34).

A visual summary of the results is provided in Figure 5.1.

5.1.2 Second study

We run a second experiment to further evaluate the usefulness of GamiDOC during the design phase of gameful systems.

5.1.2.1 Participants and procedure

During two different classes focused on gameful systems design at the University of Trento, a group of 13 students had the opportunity to utilize GamiDOC to design several gameful systems. In response, we devised a custom questionnaire to systematically assess the tool's strengths[7]. This questionnaire encompassed various aspects, including demographic information (such as age, gender, and educational background), the participants' self-perceived competence in gameful systems design, and their impressions regarding the GamiDOC gamification design framework's utility in the context of designing gameful systems.

5.1.2.2 Results

Among the participants (n = 13), 9 identified as males, while 4 identified as females, with an average age of 25.4 years. Their educational backgrounds varied, with 7 users having a background in *user interfaces*, and the remaining 6 coming from diverse specializations within *engineering* and *computer science*, including *biomedical engineering*, *telecommunication engineering*, *mechatronics engineering*, *computer vision*, and *computer science*.

Before the course, participants reported having little to no knowledge of gameful systems design techniques and skills (6 reported *"no experience"*, while 7 reported *"little experience in gameful systems design"*). However, they indicated that the course significantly improved their understanding of the subject and equipped them with the necessary tools to be competent in gameful systems design. Specifically, 10 students reported feeling *"somewhat competent"*, while only two students described themselves as *"neither competent nor incompetent"*, suggesting a perceived overall enhancement in

[7] The full questionnaire can be retrieved here: https://osf.io/t9uks/?view_only=7d1d26c2cb2f49d0a9c9d900d82bacfa.

their competence levels. During the course, detailed explanations of various gamification frameworks were provided, including the MDA model (Hunicke et al., 2004; Kim, 2015), the 6D model (Hunter & Werbach, 2012), Octalysis (Chou, 2019), Morschheuser and colleagues' framework (Morschheuser, Hamari, Werder, & Abe, 2017), the 5W2H model (Klock et al., 2016), and GamiDOC (Bassanelli & Bucchiarone, 2022).

Participants identified several key obstacles when designing gameful systems, with the most significant challenges including adapting systems to meet the needs and objectives of target users, understanding the psychological rationales behind the implementation of specific game elements, and defining rules clearly to optimize the feedback provided. For instance, one participant emphasized the importance of *"Get to know the target users and their problems in a precise way, so that it is possible to target them during the gamified system design"* to effectively target them during the gamified system design process.

Nearly all students (n = 10) reported that GamiDOC was valuable during the design phase and highlighted various reasons for its usefulness. Primarily, it was seen as beneficial because it offered a structured guide throughout the design process, facilitating the identification of weak and strong points and enabling the tracking of any changes. Moreover, it allowed students to *"understand all the phases and critical points in designing gameful systems"* and, as a result, *"structure the design workflow"*. One student mentioned that *"it was helpful to clear out the ideas in a structured way. Thus, it was easier to understand the weak and strong points"*. Additionally, students noted that GamiDOC prompted them to pay attention to components that might have otherwise been overlooked (n = 10), including the dynamics section, the pipeline, the interaction between target users and game elements, the aesthetic component, the description of rules, and various types of feedback.

Overall, one user described GamiDOC as "extremely useful", eight users found it "very useful", two users regarded it as "moderately useful", one user deemed it "slightly useful", and one user reported that it was "not at all useful" for guiding in the design of gameful systems.

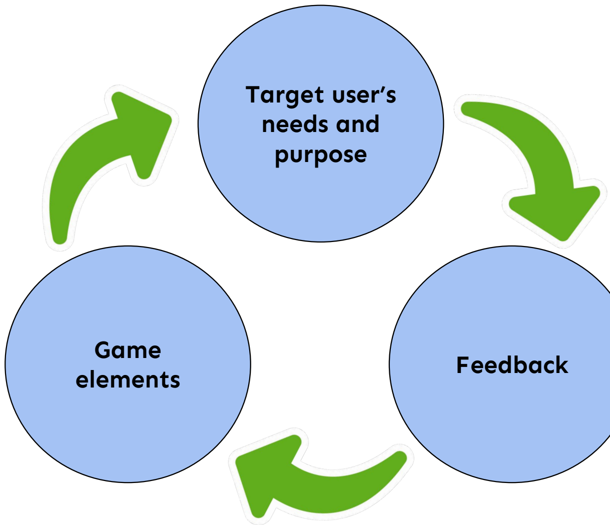


Figure 5.2: Critical elements in gameful systems design identified by users.

5.1.3 Discussion and limitations

Our findings indicate that users with limited design experience encounter the most significant challenges when designing gameful systems. These challenges revolve around the identification of end users' needs and objectives, the proper provision of feedback, and the selection of appropriate game elements for both real-time feedback delivery and user motivation (as illustrated in Figure 5.2). Given that GamiDOC's design framework places a strong emphasis on the interaction between users, game elements, and feedback, it has proven to be a valuable resource for users during the design phase of gameful systems, particularly for individuals new to game design.

Even though these data are useful in providing us with (i) feedback on the effectiveness of the tool and (ii) direction on which components to prioritize in development, the data collected are related to 16 users in the first study and 13 users in the second one. The sample size, although big enough for usability analysis (Hwang & Salvendy, 2010) and qualitative data analysis (Boddy, 2016), is not large enough to provide adequate consistency for generalizing the results of each data collection. Indeed, according to Boddy (2016), at least 20-30 participants are needed for each study to establish data saturation and thus generalize qualitative data. Likewise, according to Louangrath

(2017), for Likert-type data, an optimal sample size should be at least 30 participants. However, the small sample size reflects the non-numerosity of the target population, i.e., users with experience in gameful systems design. Moreover, in the analysis, we mainly focused on the utilities of the tool, partly leaving out the usefulness of analyzing the critical issues as well. Overall, the two sample sizes can be accepted as we collected sufficient data to have estimates with a desired level of accuracy for a preliminary study (Lakens, 2022).

The adopted analyses present some limitations. Indeed, the use of the SUS provides a quality index on usability. However, it does not provide the necessary elements to significantly improve the usability of a tool such as a qualitative analysis. Last, the selection criterion for experience in gamification design in the first study was dichotomous. In the future it would be more useful to use a more comprehensive analysis, specifically asking for users' past experiences.

5.2 Feedaback framework evaluation

To conduct a comprehensive evaluation of the feedback framework's effectiveness, we asked some faculty with progressive experience in education to try the framework (depicted in Section 4.3.1.4.3) during the design of a gameful system. To assess the effectiveness, we formulated the following research question:

> ***RQ C.1***: *To what extent the proposed feedback framework can be useful to users?*

5.2.1 Methods

5.2.1.1 Participants

Participants in the study (n = 35) were recruited through the mail from IEEE MODELS conference[8], and direct connections (mail and Slack messages) with other researchers with expertise in gameful systems design. The decision to send direct mail stems from

[8] https://conf.researchr.org/home/models-2023. Last accessed January 18th, 2024.

the fact that otherwise, it would have been difficult to identify people with experience in gameful systems design and teaching.

5.2.1.2 Procedure

The full procedure has been implemented in Google Forms[9]. First, some demographic data related to the number of years of teaching experience, and then the type of institution in which they taught were collected. We didn't collect any other demographic information, such as gender or age. The answer for years of teaching was dichotomous, *"More than 5 years"* and *"1-5 years"*, and for the institution was open-ended. Then, the survey collected information related to the difficulty of designing feedback and previous experiences with gamification. All these questions were closed-ended. Specifically, we used the three levels *"Yes"*, *"No"*, *"Sometimes"* to analyze difficulties in feedback design and gameful systems design, *"Yes, more than once"*, *"Yes, once"*, and *"No"* for the experience in gameful systems design, and a 5-point Likert-type scale (from 1 = *"Disagree"* to 5 = *"Agree"*) to report whether game elements could represent an effective medium to vehicle feedback to students.

Later, we presented to the participants an exhaustive description of the feedback framework and a contextualization for the use through the ENCORE project[10]. ENCORE is an Erasmus+ Project that seeks to produce a personalized and gamified solution to support teachers in educational settings. In ENCORE it is possible to design a learning path for different skills in the Digital, Green, and Entrepreneurial domains, providing the learning materials, and specifying which are the learning nodes — identifying whether they are textual (reading) or multimedia learning — which are the practice nodes, and which are the assessment nodes. Then, through personalization using AI, ENCORE provides practice and assessment exercises based on a specific learning objective identified.

After the presentation of ENCORE and the feedback framework, we asked participants to use the framework for each node of the learning path depicted in ENCORE (*Learning node*, *Practice node*, and *Assessment node*) to design feedback properly, and

[9] The survey can be retrieved here https://osf.io/sq4ek/?view_only=79207dc622ff4f049649db879d0ed69e.

[10] https://modis.fbk.eu/encore/. Last accessed January 18th, 2024.

to link the feedback to specific game elements. Specifically, this part of the survey allowed responses according to the categories specified in the framework, as described in Section 4.3.1.4.3. For each node, each user was asked why they had chosen certain game elements to go alongside the feedback with an open-ended question. Next, we asked how useful the proposed framework could be and whether they would use this framework in the future through two 5-point Likert-type items in the survey (from 1 = *"Disagree"* to 5 = *"Agree"*). Last, we collected suggestions, comments, and feedback on the proposed framework. All data were collected anonymously.

5.2.2 Results

Most of the participants reported to teach at the University (74,2%) and to have at least 5 years of teaching experience (54,2%). The totality of the participants (100%) reported having problems selecting appropriate feedback in educational contexts. 80% of the participants (28) reported having had experience with gameful systems, of whom 75% (21) also with the design phase. Among them, 95,24% (20) reported having problems in properly selecting feedback during the design phase of gameful systems. Game elements have been reported generically as an effective way to provide feedback to students ($M = 4,05; SD = 0.8; Mdn = 4$).

After the use of the framework, users reported it as extremely useful (**RQ C.1** $M = 4.03; SD = 0.78; Mdn = 4$), and some of them reported the will to use the tool in the future ($M = 3.66; SD = 0.90; Mdn = 4$). Then, only 7 users without experience in gameful systems answered the survey, hence we decided not to run inferential analyses to assess differences in the results according to gameful systems experience as the groups were inconsistent in size. The suggestions and comments analysis are reported in Table 5.1. Not all the users reported comments or suggestions.

5.2.3 Discussion

We assessed the effectiveness of the feedback design framework implemented in Gami-DOC in a structured survey. The purpose of this data collection was to check if the direction taken in the design of the new feedback framework was appropriate. All in all, the usefulness reported by users was positive, although there were some problems

Coded responses	Participant IDs
Expand the framework to other domains	P8
More control over the assessment part	P12
Add more contextual information in the framework	P19
Create a link to use the framework easily	P22, P23
Integrate storytelling in the framework	P25
Add integration within game elements	P25
Add the selection of more choices	P8, P27, P30
Add feedback from peers	P32
Difficult to understand without a demo	P9, P13, P20, P35

Table 5.1: Comments from the survey.

related to the study design. Users reported the framework useful since it addresses those difficulties that are often encountered during the selection of coherent feedback. Interestingly, the framework was also found to be useful in other contexts. However, as depicted in Table 5.1, some users reported difficulties as they needed more examples and contextual information in the experiments, while others suggested some more features to add to the feedback framework.

5.2.3.1 Limitations

This study presents some limitations. First, the application context was provided through the presentation of a project. To generalize its effectiveness, it would be more correct to include multimedia contextual information or use it in a real-world context. Second, the background of the users was not homogeneous, and the sample size was not sufficient to run reliable inferential analyses (the sample size required with a power of 0.80 is a total of 106 users). Thus, it was not possible to analyze differences in the utility expressed concerning prior experience. However, given the accuracy of the data obtained (Lakens, 2022), we can consider the sample size for this first analysis on the feedback framework efficacy as acceptable.

5.3 Studies conclusion

In this Chapter, we addressed the *Evaluation* of the usability of GamiDOC's gamification design framework (n = 16), the perceived usefulness of each feature (n = 13), and a first evaluation of the feedback design framework (n = 35). Usability data suggested that the tool was sufficiently usable but with ample room for improvement, specifically regarding hypertext connections, information, examples, and a more specified hierarchy. Subsequently, all GamiDOC features (exception made for the check on methodological rigor) have been indicated as useful elements in the design of gameful systems; specifically, the gamification design framework has been identified as useful to address several key obstacles reported by users (depicted in Figure 5.2). Last, the new design for the feedback framework was reported as effective by users. However, a real application is needed to assess its reliability and real effectiveness. In the next Chapter, we present an *Evaluation* of two different real case uses of GamiDOC: the first one is the evaluation of GaMoVR (which was presented above), followed by the evaluation of Untitled Bee Game, a serious game on eco-sustainability awareness.

Chapter 6

Real-case scenarios

In this Chapter, to assess **RQ 4**, I present the two real cases in which the GamiDOC design framework has been used along with a final evaluation of the tools. The first study depicts the design, development, and analysis of GaMoVR (Yigitbas, Schmidt, Bucchiarone, Gottschalk, & Engels, 2024), a tool to support students in learning Unified Modeling Language (henceforth UML). The evaluation is described in the paper *"Gamification- and Virtual Reality-Based Learning Environment for UML Class Diagram Modeling"* (Yigitbas, Schmidt, Bucchiarone, Bassanelli, & Engels, 2024), while a description of the design procedure (**RQ 3**) is depicted in the paper *"GamiDOC: The Importance of Designing Gamification in a Proper Way"* (Bassanelli, Bucchiarone, & Gini, 2024). Then, I present the design, development, and evaluation of Untitled Bee Game (henceforth UBG), a serious game to increase eco-sustainable awareness. The study is taken from the document *"Untitled Bee Game: Be(e)ing Mean to Learn More about Eco-sustainability"* (Bonetti et al., 2024). In this Chapter, the two systems are evaluated using subjective (e.g., user experience) and objective (e.g., behavioral outcomes) measures. Specifically, GaMoVR will be evaluated only through subjective measures, including usability, perceived learning, and player experience components(engagement, motivation, and fun). Instead, UBG will be evaluated relying on user experience components (motivation, playfulness), users' future behavior as subjective measures, and real learning improvement due to the game as an objective measure. Building on these results, an analysis of the effectiveness of GamiDOC in creating gameful systems will be presented in Section 7.4.

6.1 Design, development, and evaluation of GaMoVR

6.1.1 Rationale behind GaMoVR development

Over time, modeling has evolved into a crucial practice in both conceptual and system design. Hence, upcoming software engineers must acquire proficiency in modeling and modeling languages such as UML. To improve learning and learning motivation, research suggests an active learning style (Jensen, Kummer, & Godoy, 2015; Mitchell, Petter, & Harris, 2017). To implement active learning, different learning applications exist that are designed to help learners become comfortable with modeling languages like UML and actively practice modeling. With this regard, many interactive UML learning applications exist, and gamification-based alternatives have been proposed in recent years to promote the engagement of learners (Bucchiarone, Savary-Leblanc, et al., 2020; Dæhli, Kristoffersen, jr, & Sandnes, 2021; Jurgelaitis, Ceponiene, & Drungilas, 2018). According to current studies such as (Saleem, Noori, & Ozdamli, 2022) or (Shin, 2017), gamification and Virtual Reality (VR) are promising approaches for gaining knowledge and can improve necessary capabilities such as decision-making, cooperation, and communication.

To create a comprehensive gamified tool to support students in learning modeling, the GamiDOC game design framework has been used with the design of a tool called GaMoVR[1] (Gamification-based MOdeling learning in Virtual Reality), an immersive gamification-based UML learning environment in Virtual Reality (VR) which aim is to practice modeling class diagrams interactively (see Figure 6.1). The framework outlined in GamiDOC was used during the design phase to define two layers in the application and the subsequent system requirements for development. The first layer encompasses the entire application, where game elements are not confined to minigames but instead, guide the overall learning experience. These elements include a player-controlled avatar and the aforementioned progression system, which is driven by experience points earned upon completing individual tasks or minigames. The second

[1] A video of that game can be retrieved here: https://www.youtube.com/watch?v=NOQ4m_r61kA&.

layer consists of the minigames that envelop the modeling tasks. These are metaphorically modeled after the Hangman metaphor, creating a challenging environment that motivates players to perform at their best, avoid mistakes, and attempt to find solutions through strategic thinking. Within these minigames, additional actions like shooting or swinging a sword are integrated to enhance the fun factor and harness the unique capabilities of VR technology.

6.1.2 Design with GamiDOC

6.1.2.1 Procedure

To assess the usefulness of GamiDOC's framework, the main designer and developer reported in a structured interview[2] how GamiDOC — through the definition of the tool and the creation of a gamification design document — helped the team during both the design and development phases. The answers were collected using Qualtrics at the end of the development of the tool. The questionnaire consisted of two 5-point Likert-type items (from 1 = *"Disagree"* to 5 = *"Agree"*), *"Do you think GamiDOC helped you with the design of the gameful system?"*, and *"Do you think using GamiDOC for design has somehow optimized the definition of requirements and the selection of software components for the gameful system?"* with the option to openly explain the reason for the choice and a final item to collect information about other elements of the tool's utility (*"Please, report in detail in what other ways it has helped you (if any)"*.

6.1.2.2 Results

During the design phase, GamiDOC was evaluated positively. Specifically, the designer reported that the tool played a role in shaping the design of the game. Moreover, the designer specified: *"GamiDOC helped massively to understand all the different parts and influences I have to keep in mind if I am designing and developing a gamified application. When creating a gamified system for the first time, it is difficult to think about all aspects involved like the context of the application, the target user group, the desired outcomes via the game's mechanics and dynamics which are important for*

[2] The questions can be retrieved here: https://osf.io/7vqm2/?view_only=74f1e8e5158c47da83c570732383a674.

design decisions, including which game elements to use, where to include them. GamiDOC's process helped in thinking about all these steps and parts in a structured and well-documented way throughout the development of GaMoVR".

Moreover, GamiDOC played a pivotal role in defining the requirements and the selection of software components for the gameful system. Specifically, the designer reported *"By clearly defining aspects like the target user group, it was easier to select the specific game elements because I clearly knew which type of users (age, roughly their background, amongst other things) I want to target. This made it easier to decide on the game elements and game interactions I wanted to include in GaMoVR. Because GamiDOC provides a holistic view of the complete system including the technology, I was able to think about what might be possible and what might not be possible with the target platform and development frameworks right at the start and not notice complete design mistakes halfway through the development".*

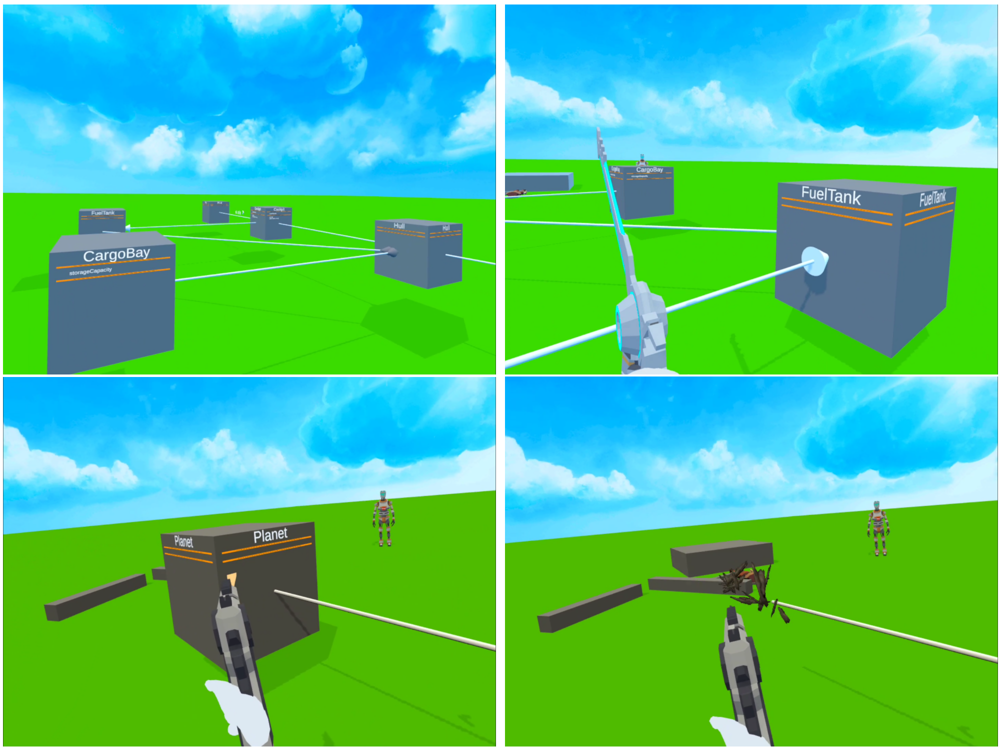


Figure 6.1: Screenshots from GaMoVR gameplay.

Overall, during this first real application, GamiDOC has been evaluated as useful for both the design phase and development phases, showing that, despite being not fully developed, it can be a useful tool in the design phase and to define the system requirements for the development.

6.1.3 Limitations

The reported data show us how GamiDOC can be useful in guiding designers and developers during the development of gameful systems. However, it should be mentioned that any critical issues encountered by users were not collected in this study. Moreover, what is reported is derived from the experience of an individual designer and developer, hence it should be kept in mind that the data reported are merely a subjective description and consequently have limited validity.

6.1.4 Evaluation of GaMoVR

After the design and development of GaMoVR, we run an evaluation of the tool. To assess its reliability, we stated the following research questions[3]:

> ***RQ D.1***: *Is GamiDOC helpful in designing a VR-based gameful system?*

> ***RQ D.2***: *Is using GaMoVR providing advantages over a different gameful system in terms of motivation, perceived learning, and/or user engagement?*

6.1.5 Setup and participants

To conduct a comparative analysis between the VR-based learning environment and a desktop application sharing similar educational objectives, we opted for a within-subject experimental design (Barnum, 2020). Under this approach, each participant was tasked with evaluating both applications. To mitigate potential bias, we employed a between-participants strategy wherein the order of application use was alternated; the first participant commenced with the VR application, the second with the desktop application, and so forth. For our comparison, we selected PapyGame (Bucchiarone, Savary-Leblanc, et al., 2023) as a representative desktop application. PapyGame is

[3]The research questions have been changed to better fit the content of the thesis.

a gamified learning tool designed as an Eclipse IDE plugin[4], offering students a platform to practice class diagram modeling, aligning with the objectives of the VR-based learning environment under scrutiny.

Participants allocated to the PapyGame group were provided with laptops equipped with a fresh PapyGame installation, along with a mouse. Conversely, participants assigned to the VR-based learning environment were equipped with Meta Quest 2 headsets[5], preloaded with the application. The Quest 2's play area (guardian) was configured to a 2.5m by 2.5m square, affording participants sufficient space for interactive engagement. To facilitate supervision and assistance, the VR headset's view was streamed to a laptop, allowing the experiment supervisor to offer guidance if participants encountered difficulties. Recruitment of study participants was conducted through multiple channels. Firstly, a request for participation was disseminated within a Slack channel established by one of the authors, frequented by students engaged in thesis work related to VR or AR topics. Additionally, the same request was extended to a project group, supervised by the same author, and to students attending a lecture on VR and AR development. While no stringent prerequisites were imposed on potential participants, prior familiarity with UML was encouraged, as neither PapyGame nor the VR application provided a comprehensive introduction to UML concepts. Interested students registered for specific time slots via a Doodle link embedded in the participation request, ultimately yielding a cohort of 16 participants for the evaluation.

6.1.6 Procedure

The user study transpired between May 30th, 2022, and June 7th, 2022, within the premises of Paderborn University. During this period, meticulous efforts were undertaken to maintain uniformity in the experimental setup across all participants. Upon the participants' arrival, a concise orientation session was provided, offering insights into the overarching study theme and outlining their prospective engagement in the user study. Subsequently, participants were introduced to the specific educational game they would commence with, bearing in mind that the order of application usage

[4] https://www.eclipse.org/ide/.

[5] https://www.oculus.com/quest-2/?locale=en_US.

was alternated for each participant, as previously elucidated. For those embarking on the PapyGame segment, a brief introduction to the Eclipse Integrated Development Environment (IDE) was provided, with a focus on the modeling plugin, Papyrus (Bucchiarone, Savary-Leblanc, et al., 2020), which furnishes both the modeling functionality and the user interface for PapyGame. Following this interface orientation, participants immersed themselves in the various levels of PapyGame. These levels were thoughtfully crafted to scaffold the participant's progression in constructing a class diagram for a fictitious food ordering system. Each successive level necessitated the incorporation of increasingly advanced elements of UML class diagrams, ultimately contributing to the participant's successful passage through the game's challenges.

Upon completion of the final level within PapyGame, participants were administered a questionnaire to solicit their impressions and evaluations of their gaming experience with PapyGame. In this user study, we opted to employ a modified version of the MEEGA360 questionnaire (Petri, von Wangenheim, & Borgatto, 2018) (modified by Gini and Bassanelli (2022)), a tool meticulously designed to assess educational games. Given its holistic approach, encompassing aspects of usability, player experience, and perceived learning outcomes, the MEEGA360 questionnaire enabled us to garner insights while accounting for the distinct attributes of both VR and desktop applications. Furthermore, the data gleaned from the MEEGA360 questionnaire empowered us to address the research queries posited in the introductory section of this paper. The questionnaire employed a 5-point Likert-type scale (Likert, 1932), spanning from 1 (*"Strongly Disagree"*) to 5 (*"Strongly Agree"*). In evaluating the VR application, participants received a brief orientation pertaining to the VR controls and were allowed to acquaint themselves with the VR device. Subsequently, they were instructed to engage with at least one level per Hangman game mode while unlocking a minimum of two UML objects, prerequisites for unlocking the subsequent levels. Upon the completion of this VR gaming experience, participants were once again presented with the MEEGA360 questionnaire, complemented by supplementary inquiries concerning their prior experiences. These additional questions probed into participants' familiarity with UML modeling, utilization of digital UML modeling tools, and previous interactions with VR devices, capturing nuanced insights into their backgrounds and proficiencies.

6.1.7 Results

The data analysis was conducted employing Rstudio, version 2022.07.2. Initially, we organized the data[6] into groups aligned with the dimensions delineated by the MEEGA360 scale, namely usability, perceived learning, and player experience. Subsequently, we placed particular emphasis on the assessment of engagement, fun, and motivation, integral facets within the player experience category. This deliberate focus on these three dimensions is grounded in their established significance as predictors of the efficacy of gameful systems (Koivisto & Hamari, 2019b; Seaborn & Fels, 2015) and educational processes (Tisza & Markopoulos, 2021b). The results derived from the MEEGA360 questionnaire indicated that GaMoVR outperformed PapyGame across all scrutinized components (Table 6.1). Specifically, GaMoVR exhibited a mean score of at least 4 (with 5 being the maximum) and a standard deviation below 1, indicative of a high degree of user satisfaction, usability, and minimal variability. In contrast, PapyGame displayed an array of average scores equal to or less than 3, accompanied by a standard deviation approaching 1, signaling lower levels of user satisfaction and usability, along with heightened variability. An examination of the overall scores obtained from the MEEGA360 scale (ranging from a minimum of 20 to a maximum of 80, where $\theta < 42.5$ indicates low quality, $42.5 \leq \theta < 65$ suggests good quality, and $65 \leq \theta$ implies excellent quality) revealed that GaMoVR achieved a score of 68.17, denoting excellent quality. Conversely, PapyGame garnered a score of 48.66, indicative of an overall good quality assessment. Turning attention to the additional questionnaire items, participants reported a moderate level of prior experience in UML modeling, digital UML modeling tools, and VR devices (Table 6.2).

Dimension	**Results**	
	GaMoVR	PapyGame
Usability	4.32 ± 0.85	3.05 ± 1.20
Perceived learning	4.00 ± 0.77	2.54 ± 0.87
Player experience	4.26 ± 0.79	2.40 ± 1.15

Table 6.1: Mean and standard deviation at the MEEGA360 scale.

[6] The raw data can be retrieved here: https://osf.io/uqmj5/?view_only=2697bf363c7a458ebfdf9a32f002cd61.

To discern significant distinctions in user perceptions of the two applications, we initially clustered the data into the MEEGA360 groupings: usability, perceived learning, and player experience. Subsequently, we conducted a Multivariate Kruskal-Wallis analysis (MKW) (H. Oja & Randles, 2004) to analyze significant differences between the two groups (GaMoVR and PapyGame) in the questionnaire answers. The findings illuminated an overarching significant variance in tool perception ($H_{(3)} = 25.29, p < 0.001$). The selection of a non-parametric test was motivated by two considerations: firstly, Likert-type data often exhibit skewness, and secondly, our study's sample size fell short of that recommended by power analysis standards (Gardner & Martin, 2007). Specifically, G*Power (version 3.1) suggested an overall sample size of 56 participants, with $\alpha = 0.05$, to achieve a power of 0.80.

Item	**Results**
Prior Experience with UML Modeling	3.40 ± 0.73
Prior Experience with digital UML Modeling Tools	2.80 ± 0.77
Prior Experience with VR Devices	2.73 ± 1.39

Table 6.2: Results about the prior experience with VR and UML (modeling tools).

Subsequently, we analyzed the disparities between the results of each application with regard to engagement, fun, and motivation. Due to ordinal data and small sample size, a Wilcoxon signed-rank test (Woolson, 2007) was employed for each dimension, yielding results indicative of significantly higher values for GaMoVR as compared to PapyGame in terms of *engagement* ($V = 120, p < 0.001$), *fun* ($V = 136, p < 0.001$), and *motivation* ($V = 136, p < 0.001$) perception. These outcomes are graphically represented in Figure 6.2. To ensures that any detected differences between groups are not simply due to chance or the result, pairwise comparisons, fortified by Dunn's test with Bonferroni adjustments were employed, exhibiting statistical significance across all three dimensions ($p < 0.001$ in all instances).

6.1.8 Discussion and limitations

The results from the interview reported that GamiDOC can be useful for both the design and development phases (**RQ D.1**), supporting the definition of the system requirements for further development. The results from the evaluation reported that GaMoVR reached adequate levels in the MEEGA360 scores, suggesting a good balance

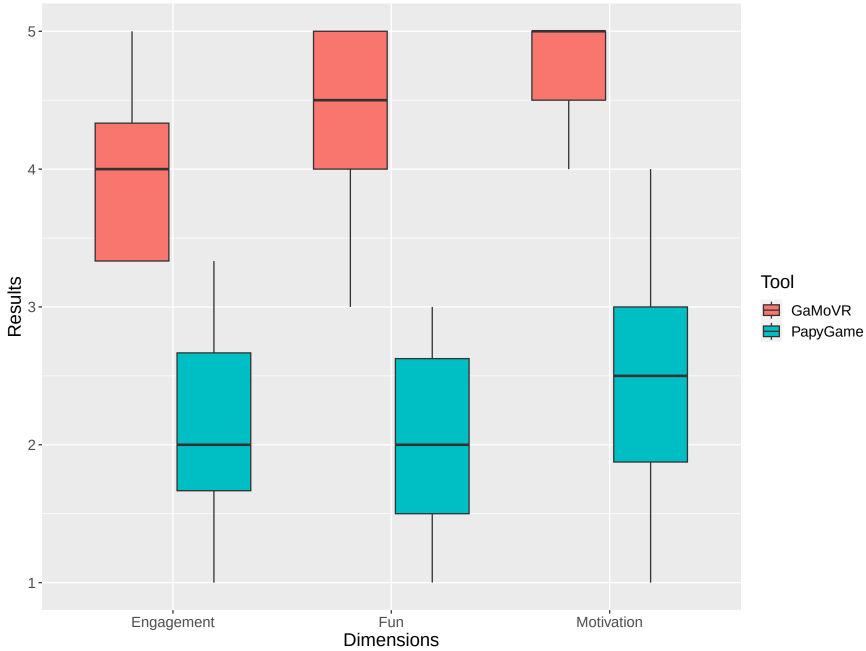


Figure 6.2: Results of the three analyzed dimensions - GaMoVR vs. PapyGame.

of the tool's usability, player experience, and learning perceived by users through the use of the tool. Moreover, the inferential results suggest that GaMOVR could provide advantages in motivation, perceived learning, and engagement over a different gameful system (**RQ D.2**). Overall, in this application, GamiDOC proved to be a useful approach for gameful system design, producing a reliable tool.

In light of the favorable outcomes delineated, it is imperative to acknowledge certain constraints that necessitate due consideration when interpreting the data. Firstly, the utilization of a diminutive sample size presents a constraint, precluding unequivocal extrapolation of the findings. A power analysis conducted via G*Power advocated for a minimum sample size of 28 participants to attain a requisite power level of 0.80. Nonetheless, it is worth noting that, as an initial pilot investigation, the selected sample size was deemed adequate for the inaugural analysis. Secondly, the study's inability to assess the ultimate user outputs when employing the two distinct tools represents a limitation. This limitation impedes a comprehensive evaluation of the authentic efficacy in the learning trajectory, restricting the assessment solely to the subjective perception of learning gleaned from the questionnaire items.

For prospective investigations, it is imperative to augment the sample size to facilitate the generalization of outcomes and facilitate a comprehensive assessment of GaMoVR's efficacy. Such assessment should extend beyond its influence on user motivation, enjoyment, and engagement, to encompass their tangible impact on the actual acquisition of UML modeling skills.

6.2 Design, Development, and evaluation of Untitled Bee Game

6.2.1 Rationale behind Untitled Bee Game development

The rationale behind the development of UBB stems from the great changes within the ecosystem and the climate due to human action. Human action is making great changes within the ecosystem and the climate through numerous unsustainable behaviors (Celik, 2020).

Every year, about 25% of the plastic we produce, which equates to a staggering 80.5 million tons, ends up in our environment, making plastic a pervasive environmental pollutant (Alimi, Farner Budarz, Hernandez, & Tufenkji, 2018; Unep & ASSESSMENT, 2016). Moreover, the use of non-eco-friendly elements, such as most detergents, causes enormous problems, including soil structure deterioration, with detrimental consequences for plant well-being (Mousavi & Khodadoost, 2019), and eutrophication, which results from the discharge of detergents into water bodies, leading to depletion of dissolved oxygen levels in the water, which can be harmful to aquatic organisms such as fish, seaweed, corals and other aquatic life. Human action — e.g. urbanization, light pollution, chemical pollution, plastic pollution, deforestation, noise pollution, and air pollution (Barker & Tingey, 2012; Hölker, Wolter, Perkin, & Tockner, 2010; Sordello et al., 2020) — is therefore producing a decrease in terms of biodiversity, contributing to the decline of some species, especially pollinators, such as wild bees (Potts et al., 2016). Bees play a pivotal role in delivering a spectrum of ecosystem services that not only enhance human well-being but also sustain the fundamental life support systems of our planet (Patel, Pauli, Biggs, Barbour, & Boruff, 2021). These ecosystem services are inherently aligned with the pursuit of global sustainable development goals (Wood, Kaplan, & Szendrei, 2018).

This underscores the necessity for a fundamental transformation in people's conduct (Oskamp, 2000). It is imperative to replace practices such as consumerism, wasteful resource utilization, environmental degradation, contamination, selfish and unjust behaviors with sustainable habits, using the world's resources in ways that will allow human beings to continue to exist on Earth with an adequate quality of life (Oskamp, 2000). With the continuous upgrading of educational technology, gameful systems, known for being highly motivating, have often been implemented to promote behavior change approaches or to support positive behaviors (Bassanelli et al., 2022; Seidler et al., 2020). Overall, gameful systems present a reliable approach to enhance and encourage users to engage in the desired behavior, potentially fostering environmental protection. Some investigations concentrating on the connection between video games and ecological knowledge have demonstrated noteworthy quantitative outcomes. In particular, there is a notable increase in players' motivation to grasp ecological concepts through video games (Coroller & Flinois, 2023).

We therefore developed a serious game called Untitled Bee Game[7], in which players have the goal of discouraging non-player characters (NPCs) from engaging in specific non-eco-sustainable behaviors while learning notions related to pollution through dialogue with flowers. The long-term goal of the game is to foster positive behavioral change in terms of eco-sustainability, by providing knowledge engagingly and enjoyably.

To assess the game's usefulness and qualities, we formulated the following research questions:

> ***RQ E.1****: To what extent was the game perceived by the players as playful and meaningful?*

> ***RQ E.2****: How would learning and perceived learning be affected by this game?*

> ***RQ E.3****: What design directions should be taken in developing and improving such a game?*

[7] The full game can be downloaded here: https://osf.io/eznjw/?view_only=9cf4572c4b1e41e3a39be622a9ab433d.

6.2.2 Design and Development

UBG borrows its title and mechanics from Untitled Goose Game[8], a 3D puzzle game for computers and consoles. Untitled Goose Game is part of a rather recent video game genre that could be defined as “animal mayhem games” (Caracciolo, Marco, 2021). These games let players control animals that wreak havoc and punish human beings, generally relying on humor and disruptive game mechanics. Other such games are, for example, Goat Simulator[9] and DEEEER Simulator [10]. This strand of games seems appropriate to focus on the theme of sustainability and environmental preservation: they shift the point of view from humans to animals and let players correct or condemn human behavior as an external observer. We also suspect that carrying out hostile in-game actions as a human instead of an animal could potentially have negative educational consequences. The game might be thought to suggest that it is advisable to be hostile to peers who do not follow certain virtuous behaviors. The core mechanic of Untitled Goose Game consists of playing tricks on and stealing stuff from human non-playing characters (NPCs) to disturb their daily activities. Albeit relying on such a simple mechanic, the game managed to sell one million copies in three months (Forbes, 2019). In addition, its simple, toon-shaded low-poly 3D graphics make for the ideal style for rapid prototyping and development.

As a “game with a purpose”, we carefully designed the game from the contextual information and purpose. We then relied on the GamiDOC design framework to write a game design document[11] that would allow the team to communicate and at the same time, keep track of the development concerning the purpose. The game had no specific target and was inspired by a graphic style that could be considered pleasurable across different genders and ages. The idea behind UBG is to explore the potential of immersing players in a 3D virtual world inspired by recent commercial games, with NPCs reacting to the avatar’s actions, to foster novel gameful interactions for learning and positive behavior change. UBG lets players control a bee in a small town (see Figure

[8] https://goose.game/.

[9] https://www.goatsimulator3.com/it.

[10] https://playism.com/en/game/deeeer-simulator/.

[11] The game design document can be retrieved here: https://osf.io/eznjw/?view_only=9cf4572c4b1e41e3a39be622a9ab433d.

6.3). We selected a bee because of the importance of bees in the environment (Patel et al., 2021). Indeed, facts about bees are part of the learning content of the game, which players are quizzed on. This choice also brought about advantages under the technical aspect of the implementation, such as having to deal with simpler animations. Imitating commercial games may imply adopting control schemes that are a bit difficult to master. Therefore, at the beginning of the game, players are asked to select the control style they prefer: overhead camera or free camera. In both schemes, players control the bee in the third person, but in the latter, they can also rotate the camera (and the bee) freely in all directions. The overall objective is to dissuade the NPCs from pursuing certain behaviors. If correct actions are taken by the player, the garden (a small area where flowers can appear) and the river are influenced: the garden produces flowers and more fish come to the river.

Figure 6.3: The scenery of Untitled Bee Game. The screenshot shows the small town that is explorable in the game.

Two NPCs were implemented. The first one wastes plastic bottles, while the second one uses polluting substances to wash her car. An icon above the head of the NPCs shows the player what behavior is being pursued. If it is a damaging behavior, players must sting the NPC or drop rubbish on their heads. Stinging can be performed at any time, even if the behavior shown by the NPC is already correct. Therefore we differentiate between two stinging actions the player can perform: correct and incorrect stinging. The game plays different sounds accordingly. If correct stinging is performed, new flowers appear in the garden and new fish appear in the river. Once a flower has appeared, players can interact with it and read facts about the three topics (see Figure 6.4), one paragraph at a time, shown in a classic video game message box. Once a knowledge paragraph has been unlocked, it becomes possible to read it again in the pause menu.

As a side mechanic, players may also pick up and recycle the rubbish they find. This grants them one additional fish. In addition, rubbish can be thrown at NPCs instead of stinging them to obtain the same effect. If players do not do anything for a certain amount of time, fish start to disappear. It is worth noting that there is no game over and the players' actions are only driven by the possibility to make the virtual environment better and more lively. The mechanics of UBG are still rather simplistic, allowing as many participants as possible to complete the experiment without too much effort. Apart from a couple of side mechanics, the task is rather straightforward.

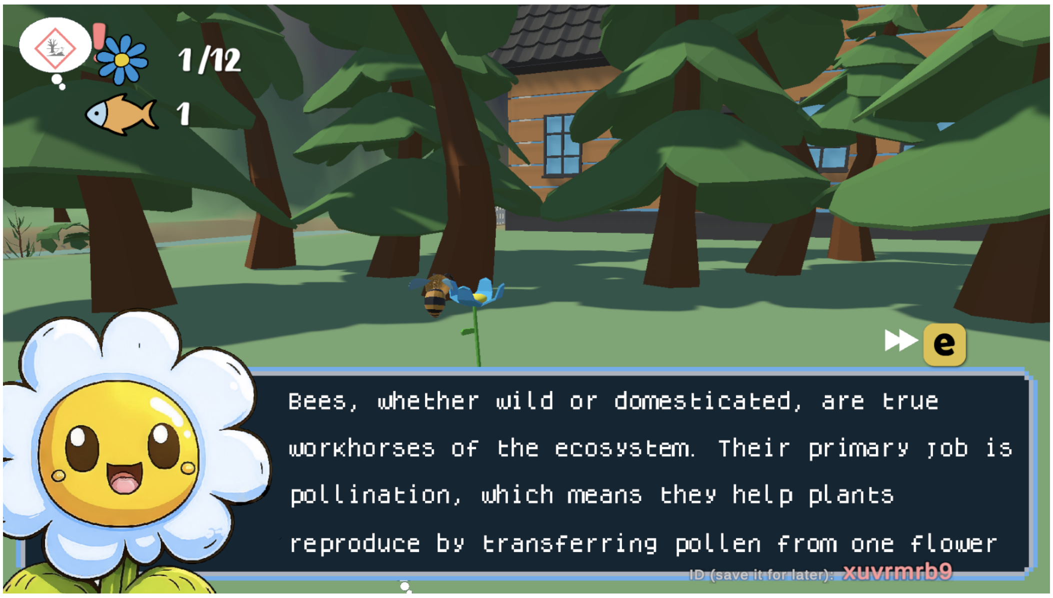


Figure 6.4: The messages displayed by pollinated flowers.

6.2.3 Tools and resources

UBG was developed using Unity3D[12], an industry-standard environment in the field of video game development. Some of the 2D resources and images were found on Adobe Stock[13], while others were created ad-hoc. The 3D models were either found in the Unity Asset Store[14] and Quaternius[15] or created using Blender[16], a popular 3D modeling tool. The music was either created ad hoc or found online for free[17].

All of the text within the game and the questions presented in the initial and final quizzes were created following an extensive literature search on the topic. Following an initial analysis, the final data were selected based on their association with the player and NPC behaviors within the game[18].

6.2.4 Methods

6.2.4.1 Data collection method

Before deploying the game and administering it to the participants, we created a REST API based on a Mongo[19] database that communicated with the game during the sessions. For this reason, players were asked to remain connected while playing. We collected the following information in JSON format: (i) a game ID; (ii) the number of times each NPC was stung; (iii) the overall number of correct actions; (iv) the overall number of incorrect actions; (v) the quiz scores; and (vi) the paragraph reading times.

6.2.4.2 Participants and procedure

Players were mainly recruited in Italian and Canadian university facilities (University of Trento, University of Verona, University of Turin, and University of Waterloo) and there were no exclusion criteria except for being underage. Recruitment requests were

[12] http://www.unity.com.
[13] https://stock.adobe.com/.
[14] https://assetstore.unity.com/.
[15] http://quaternius.com.
[16] https://www.blender.org/.
[17] https://audionautix.com/.
[18] The final documentation used in the game can be retrieved here: https://osf.io/eznjw/?view_only=9cf4572c4b1e41e3a39be622a9ab433d.
[19] https://www.mongodb.com/.

circulated by two lecturers among bachelor students primarily. They could choose between a browser-based, a Mac, and a Windows version. They were both Italian and English-speaking and could change the game language at the beginning of the game. The procedure consisted of playing for approximately 20 minutes and responding to an online questionnaire. The game assigned a random alphanumeric ID to each player to be entered in the final questionnaire, which granted complete anonymity. The procedure was employed in two phases: the usability test and the main experiment. The main differences were (i) the version of the prototype, and (ii) the content of the final questionnaire, administered after the post-session quiz, as described below.

Game session. Players could play a desktop or a web version of the game. After being shown their ID, they were taken to the initial quiz (pre-test) where they could express a choice among 4 to several questions related to eco-sustainability to establish a baseline of knowledge about sustainable behavior. This quiz established a baseline for each participant. After the initial quiz, players had to complete a brief tutorial about the game commands and mechanics. Then they could start exploring the environment and stinging the humans showing undesirable behavior. In so doing they earned a fish and a flower, and then had to pollinate a flower to reveal a new fact. After revealing a certain number of facts, players were taken to the second quiz (post-test) with an additional 4 questions on the same topics as the initial one.

The content and order of the questions in the two quizzes were the same for all participants, both in the usability test and the experimental phase. Accordingly, the eco-sustainability facts were shown to all players in the same order. The game informed players whether their answer was correct or incorrect, and in the latter case, it also displayed the correct one.

Usability test design. To collect usability data, we used the System Usability Scale (SUS), which had to be filled out at the end of the game. The choice of using this tool stems from its reliability and its widespread use for usability studies (Lewis, 2018). Moreover, we added the item *"Do you have any suggestions, feedback, comments that can help us in improving the game?"* to collect more detailed information about possible improvements in the game. We recruited 9 participants (female = 4; male = 5) and had them play the first version of the prototype, where they had to unlock

16 flowers and answer 16 quiz questions in total (8 pre-session and 8 post-session). Each participant tried the game through the web or desktop version and answered the online questionnaire on Google Form[20]. Data in literature (Hwang & Salvendy, 2010) suggest that for usability studies the sample size of 10 ± 2 is recommended to detect at least 80% of usability problems. During the usability data collection, we still collected data on playfulness, meaningfulness, and learning perception to check whether usability could in any way negatively affect these aspects. Hence, we selected some items from reliable questionnaires: **Q1-** *"This game gives me an overall playful experience"* from the playfulness dimension of the GAMEFULQUEST (Högberg et al., 2019), **Q2-** *"I think this was an important activity"* and **Q3-** *"I believe this activity has been beneficial to me"* from the value/usefulness dimension of IMI (McAuley et al., 1989), and **Q4-** *"The game contributed to teach me some of the most relevant environmental issues"* from the perceived learning dimension of MEEGA360 (Gini & Bassanelli, 2022). We then interviewed the participants in a think-aloud session.

Experiment design. We recruited 28 participants (age: $M = 26.62; SD = 8.80$; 13 identified as female, 15 identified as male), among which 18 completed the game, and 24 completed the final questionnaire, of which 14 completed both the game and the final questionnaire. In this phase, players were administered a refined version of the prototype, where many things were changed according to the previous round of feedback. In particular, the number of paragraphs to read and the number of quiz questions were both lowered to 12 in total. After playing the game, participants had to fill out a questionnaire where they were asked to provide: (i) demographic information and gaming habits; (ii) an informal assessment of the degree of sustainability of their behavior; (iii) engagement and value provided by the game by using the items from the GAMEFULQUEST, the IMI and the MEEGA360 described above, and two more items — **Q5-** *"Thanks to the game, I am likely to inquire about how to adopt other sustainable behaviors in the future"* and **Q6-** *"Thanks to the things learned in the game, I am likely to adopt more sustainable behaviors in the future"* — created ad-hoc to assess how much the game may have influenced users to become informed about environmental issues and adopt environmentally sustainable behaviors in the future. All of the materials

[20] https://www.google.it/intl/it/forms/about/.

were available both in English and Italian.

6.2.5 Results

We report here both the results from the usability test (expressed mainly in terms of SUS scores and feedback received) and the main experiment.

6.2.5.1 Usability test

The game's SUS overall score (62.2) was below the set threshold for a good level of usability (68), indicating marginal usability; hence, additional improvements were needed (Bangor, Kortum, & Miller, 2009).

6.2.5.1.1 Questionnaire results

The results of the additional items have identified that during this first application, not all users found the game completely entertaining (**Q1** $M = 3.33, SD = 1.11$). However, they recognized its educational power and its importance (**Q2** $M = 3.44, SD = 1.01$; **Q3** $M = 3.66, SD = 1.11$; **Q4** $M = 3.88, SD = 0.78$).

6.2.5.1.2 Feedback comments

At least one participant (U1) found that the game contained too much text. Therefore we lowered the number of flowers to 12. This meant less text to read (12 facts instead of 16) and less playtime. Three participants (U3, U4, U8) mentioned that the objective was not completely clear or that the icons were difficult to interpret. A participant wrote that they would have liked to know how important it was to read the flower facts. This would have allowed them to answer more correctly to the final quiz. As a solution, we added a summary of the objectives in the pause menu. Another participant (U9) mentioned the difficulty of understanding interactions with fish within the game. We therefore decided to specify this more clearly. The controls were a bit difficult for at least two participants (U3, U7). This led us to offer a choice to the players at the beginning, asking which control scheme they preferred. However, one participant (U1) liked how they worked. Finally, one participant (U7) stated that the background music was a bit annoying. This stresses the importance of giving players an option menu to personalize the experience.

Think-aloud session. After the usability testing, we interviewed the participants informally in a think-aloud session to gather additional feedback. Some points were raised in this session that had not emerged as part of the written comments. We summarize the main topics that emerged during the think-aloud session in Table 6.3. The table includes both suggestions that were later implemented into the game and those that were not, followed by an explanation.

The suggestion **"Add flowers around to encourage exploration"** was not followed because the game only allows to unlock flowers in a dedicated area to optimize the traveling time and the experiment duration, but will be considered in the future; **"Add a narrative"** was not followed because, as mentioned, the game was optimized to be completed in a reasonable amount of time, so we have not yet included a story behind the characters or the avatar, but it is a direction for future work.

Comment	Implemented
Add a progress bar for the quizzes	Yes
Make tutorial more visual	Yes
Include recycling mechanics	Yes
Make it easier to unlock the fish	Yes
Add flowers to encourage exploration	No
Add a deeper narrative	No

Table 6.3: Suggestions from the think-aloud session, with an indication as to whether they were followed.

6.2.5.2 Experiment

What follows is a summary of the results obtained with the main experiment, conducted after the usability test. We proceed to describe and analyze the data obtained from the questionnaire, the change in quiz performance in pre-test and post-test, and the feedback received at the end.

6.2.5.2.1 Questionnaire results

All of the analyses were conducted using RStudio (version 2023.03.0). Among the participants, 24 answered the questionnaire. The results (Figure 6.5) show an overall positive evaluation of the tool. Answers were given on a 5-point Likert-type scale ranging from *"strongly disagree"* to *"strongly agree"* (See Table 6.4). Most of the players found the game playful (63%); however, a few subjects thought the game was not playful, indicating the need to bring changes to the game in the future (12%). Interestingly, almost all of the players found the game meaningful (96% and 83% to **Q2** and **Q3** respectively), probably because of the importance of the themes it dealt with, and they felt they learned something from the experience (88%). Last, 58% of the players reported the will to inquire (**Q5**) and then adopt other sustainable behaviors in the future thanks to the game (**Q6**). We then ran a linear regression to analyze whether the expressed playful level could be explained by users' prior experience with video games. The results show no correlation between the two variables (Adjusted $R^2 = -0.07, F_{1,12} = 0.14, p > 0.05$), suggesting that the game is equally perceived despite users' experience with video games. To assess whether the meaningfulness and playfulness of the game, and the perceived learning were conveyed by the amount of information read, we ran some linear regression between the reading time and the playfulness value (**Q1**), the interaction of the items related to meaningfulness (**Q2** and **Q3**), and the perceived learning (**Q4**). The results show that no correlation occurred between the reading time and the expressed game meaningfulness ($R^2 = 0.58, F_{1,12} = 2.54, p > 0.05$), and playfulness ($R^2 = 1.82, F_{1,12} = 0.301, p > 0.05$). The analysis showed a significant relationship between reading time and perceived learning ($R^2 = 1, F_{1,12} = 5.79, p < 0.001$).

Item	Mean	SD	Mdn
Q1	3.66	1.05	4
Q2	4.37	0.71	4
Q3	4.16	0.82	4
Q4	4.37	0.82	5
Q5	3.58	1.10	4
Q6	3.75	0.94	4

Table 6.4: Questionnaire mean, standard deviation, and median.

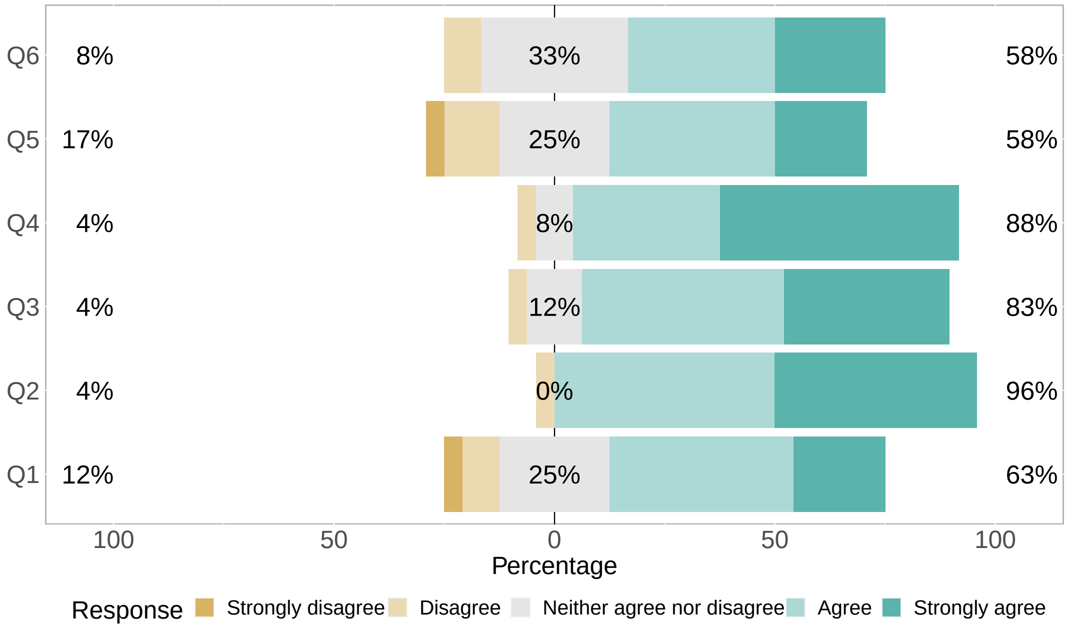


Figure 6.5: An overview of the users' answers to the 5-point Likert-type items (**Q1-Q6**) in the questionnaire.

6.2.5.2.2 Quiz score improvement

Among the participants, 18 completed an entire game session and did the post-test quizzes. To answer **RQ E.2**, we investigated the difference in scores between the pre-test and post-test quizzes. The post-test quiz obtained generally higher scores ($M = 4.55; SD = 1.46; Mdn = 5$) than pre-test one ($M = 3; SD = 1.23; Mdn = 2.5$) (Figure 6.6). At first glance, it was possible to see that only 5 players did not get an improvement, of which 4 maintained the same score, and one only had a lower score. We then analyzed whether these differences in users' answers before and after the use of the game were significant. To analyze the distribution of variables and guide us in the choice of tests to be used, a Shapiro-Wilk test was conducted, indicating that the two distributions of both variables differed from the normal distribution (Pre: $W = 0.79, p = 0.001$; Post: $W = 0.83, p = 0.005$). Considering this result, we opted for a non-parametric test. Expecting an improvement due to the use of the game, we then performed a one-tail Wilcoxon signed-rank test (Woolson, 2007) finding a statistically meaningful disparity in the results at the in-game eco-sustainability questionnaire before and after playing the game ($V = 11, p < 0.005$). To measure the magnitude of this data, effect size calculation ($r = 0.49$), which for the Wilcoxon signed-rank test is based on the formula $r = \frac{Z}{\sqrt{n}}$ (Fritz, Morris, & Richler, 2012), suggested a moderate magnitude of the effect (Tomczak & Tomczak, 2014). The post-game results are not explained by the subjects' prior experience with video games (Adjusted $R^2 = 0.02$, $F_{1,12} = 1.32, p > 0.05$) or the flower reading time (Adjusted $R^2 = 0.76, F_{1,16} = 4.49, p > 0.05$).

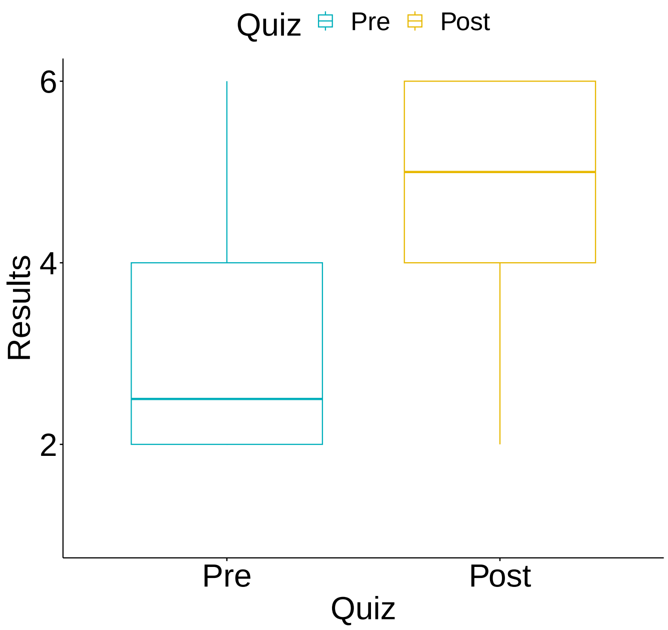


Figure 6.6: A boxplot showing the pre-test and post-test scores on the eco-sustainability quizzes, taken at the beginning and at the end of the game session. There were 6 questions in each quiz.

6.2.5.2.3 Feedback comments

We were able to extract a substantial amount of comments from the feedback collected with the final questionnaire, which helped us answer **RQ E.3**. We conducted a thematic analysis by coding all feedback manually under macro-themes. We summarize the comments here, grouped by theme. In Table 4, we provide a detailed breakdown of the main comments grouped by participants. In Table 6.5, we provide a detailed breakdown of the main comments grouped by participants.

Theme 1. Repetitive tasks: E1, E2, and E4 mentioned that they would have liked to have different areas to explore and to find flowers in. This mirrored what emerged during the usability testing. E7 also mentioned that the positioning of the

Coded responses	Participant IDs
There should be additional areas	E1, E2, E4
Exploration is not encouraged	E7
Text is too verbose	E6, E11, E14, E16
Text is too formal	E6, E10, E18
There should be a voiceover	E3, E13, E16
The font is not very readable	E18, E20, E24
The game was a nice experience/fun	E6, E11, E19
Being a bee is a nice concept	E16
Stinging humans is fun/satisfying	E9, E18
Input controls are easy	E6
Input controls are difficult	E10, E16, E17, E19

Table 6.5: Suggestions from the experimental session, with an indication as to who provided them.

garden and the humans may be exploited to take shortcuts. **Theme 2. Too much text**: in general, many users expressed that they would have liked to read less or simplify the text. E6, E11, E14, and E16 found that the text was too verbose. E6, E10, and E18 stated that they would have liked less formal language. **Theme 3. Voice over**: this theme is largely connected to Theme 2. In particular, players E3, E13, and E16 suggested employing text-to-speech techniques or voiceovers to decrease the amount of text to be read, especially for people with dyslexia. **Theme 4. Unclear goals**: player E8 stated that they found the stinging context button to be somewhat difficult to activate or interpret. E9, E14, and E19 asked for more succinct indications as to what to do with the flower and the fish. E19 suggested to include more visual cues. **Theme 5. Font readability**: flowers taught players their facts with a pixelated font, which was chosen to be reminiscent of old-school video games. In this particular context, it produced mixed results. Three participants (E18, E20, and E24) had difficulties reading the font and this made them want to skip the messages (see Figure 6.4). **Theme 6. Playing as a bee and stinging is fun**: overall, it was considered a nice experience by E6 and E19. E6 and E11 stated that the game was fun; E16 stated that they liked the concept of being a bee; E9 and E18 explicitly stated that being mean to humans made them happy or satisfied. **Theme 7. Mastery of controls**: although someone explicitly stated that controls were easy to master (E6) some people expressed their difficulty in mastering them (E10, E16, E17, and E19).

Finally, it is worth noting that only 4 players did not report any comments.

6.2.6 Discussion and limitations

We assessed the enjoyability and usefulness of UBG in two phases: a usability testing phase and the main experiment. All in all, the experience reported by users was positive, although there were some problems in the usability phase. Comments were particularly useful in the usability testing phase since usability appeared to be suboptimal according to the SUS. Regarding the main experiment, we observed that there was indeed an increase in quiz performance, thanks to the knowledge presented in the game, which is an encouraging finding that contributes to answer **RQ E.2**. Interestingly, quiz performance is not explained by self-reported experience with video games according to our analysis. This suggests that potential problems are not likely to have influenced learning. The level of reported playfulness does not seem to be influenced by experience with video games. In this respect, in addressing **RQ E.1**, we can say that in the end the prototype was quite enjoyable for every participant and the difficulties with the controls did not invalidate the experience.

Several observations can be made by looking at the feedback collected to answer **RQ E.3**.

1. Consider implementing a voiceover: The proposal of removing some text and/or adding a voiceover was shared by 29% (7) of participants. In general, reading was fatiguing for players, and adding recorded speech or using text-to-speech techniques seems like the sweet spot between teaching notions and not overwhelming players, although we suspect that learning could be influenced by the two different methods. Speaking of text more in general, our findings mirror those of Fernández Galeote, Legaki, and Hamari (2023), where the authors suggest avoiding too much text and using instead more visual and impactful feedback.

2. Take into account individual experiences: Four players thought that either the controls were difficult to master or that they could be improved, while one player found them easy to master. This underlines the importance of taking into account not only all the possible peripheral configurations but also all the possible levels of experience with video games. Choosing the right game genre is key: some games, like first-person shooters, for example, require very specific skills such as aiming with the mouse.

3. Too much freedom may lead to unclear goals: We obtained 24 complete questionnaires in the face of 18 completed game sessions. This fact is likely to be related in some way to T4 (Unclear goals), but it could also be influenced by the fact that the game, albeit giving instructions, leaves players rather free to explore. We therefore recommend future endeavors in serious games to consider the aspect of directing player to the goals in the most straightforward and clear way possible. Finally, 5 players explicitly stated that either the game was a good experience or that controlling a bee and stinging the humans was fun. This contributes to addressing **RQ E.1** and suggests that being mean to humans can be perceived as satisfactory.

There are several limitations to the present study. The serious game employed is still in an early version and lacks a number of features that could make the experience more enjoyable (e.g., a more complex narrative and high-level goals that go beyond the stinging task alone). Since one of the aims of the study was to assess the educational potential of the game, the instructions and the mechanics guided players rather strictly towards reading and responding to quizzes, while it could be turned into a more relaxed activity in future versions. The second limitation is the limited number of participants involved in the study. Overall, we collected data from 37 participants, which can be considered adequate for a preliminary version of the tool. A greater sample could be recruited or a follow-up knowledge retention test could be introduced to further support the learning trend observed. It could be argued that telling players about the final quiz may have encouraged them to read more carefully, thus influencing the final results. Future directions may include a between-subjects experiment, where the treatment group does not know about the final quiz.

6.3 Studies conclusion

In this Chapter, I addressed the evaluation of two systems created using GamiDOC in the design phase following subjective and objective measures. Specifically, to analyze GaMoVR, we based the analysis on usability, perceived learning, and player experience, placing more emphasis on engagement, fun, and motivation, and comparing the results obtained with another gameful system to support UML learning, namely PapyGame. Instead, to evaluate UBG, we relied on motivation, playfulness, users' future behavior, and perceived and real learning due to the use of the game. In both cases, the games proved to be effective, producing optimal levels of subjective measures (player experience, usability, and perceived learning) in GaMoVR, with significantly higher levels than its counterpart, and obtaining excellent values in the subjective measure in UBG (motivation, playfulness, users' future behavior, perceived learning) and in the objective outcomes, showing a significant post-use learning effect. Overall, the two tools have proven effective in capturing attention and engaging users, helping users perceive improved learning compared to traditional solutions, and supporting real learning in UBG's case, producing environmental awareness and willingness to enact more sustainable behaviors. Nevertheless, the reported analyses have limitations. In the future, it is crucial to rely more on behavioral data, recruit an adequate sample size following prior power analysis, use the same standardized questionnaire for all the evaluations, and compare the developed system with more similar systems or alternatives recognized as valid in the literature. Additionally, in the future, it will be necessary to develop a detailed framework to verify the effectiveness by subjective and objective metrics of the various systems produced with GamiDOC and, consequently, of the effectiveness of GamiDOC.

In conclusion, the two systems developed using GamiDOC in the design phase proved effective in keeping users engaged and motivated and providing effective learning. Thereby, GamiDOC appears to effectively support development teams by allowing them to keep track of elements crucial to the effectiveness of gameful systems; as a result, the systems produced have proven effective in promoting learning and behavioral change. A more exhaustive discussion on GamiDOC effectiveness is presented in Section 7.4.

Chapter 7

Conclusion and future works

This chapter will conclude the thesis by summarizing the key research findings in relation to the research aims and questions. It will also report the elements for future research to complete the development of GamiDOC.

This dissertation aims to investigate the challenges in gameful systems design and present an effective solution in guiding designers throughout the process of creating a gameful system, namely GamiDOC, from the design to the final evaluation. Results on the usability and usefulness of GamiDOC indicate that the tool still needs some improvements; however, the design framework is considered useful during the design phase and in the selection of system requirements, particularly for individuals new to gamification design. Moreover, according to users, the whole community could benefit from the final version of the tool, which will allow for improved gamification reliability. Real application and systems evaluation indicate that the systems developed using GamiDOC in the design phase are effective and enjoyable and that the game elements used are appropriate to the context and purpose. In the next parts, I answer the research questions I initially defined. In each section, a single research question is discussed based on the findings described in the previous chapters.

7.1 RQ 1: What are the elements that interfere with the design of gameful systems and, subsequently, their effectiveness?

To answer this question, I first focused on the data reported in Chapter 2. The data obtained from the two scientometric analyses suggest that documents and authors are reporting a change in themes and trends in the literature related to the use of gamification and its design. This is due to an inconsistency in its effectiveness and the lack of comprehensive and reliable design methods. As a result, more and more studies are questioning the drawing methodologies used up to that point and emphasizing the importance of increasing methodological rigor. Specifically, the reference literature in gamification has shifted from papers related to persuasive systems to critical reviews and research agendas, suggesting that while there are no comprehensive solutions, there is agreement on the issues that undermine the effectiveness of these systems.

Building on this data, we analyzed the elements that are reported to interfere with the design of gameful systems in Chapter 3. According to the literature, the main challenges in gamification design are composed of (i) users' interpersonal differences and preferences, (ii) users and systems goal differences, (iii) game modality, (iv) contextual differences, (v) technological components, (vi) feedback design, (vii) game elements selection, (viii) methodological rigor for the tool evaluation, and (ix) which data collect. Therefore, these elements must be monitored during the design process, but also during development and evaluation, to obtain a system that is useful and effective, and reliable data on its use.

7.2 RQ 2: How can we address the challenges in gameful systems design, and create an effective way to support designers and developers?

As reported in Chapter 3, numerous strategies and frameworks for design have been developed in recent years. However, numerous challenges are not adequately addressed and, more importantly, no single solution can address them all. To address this research question, we designed GamiDOC, a tool to support the design, development, and evaluation of gameful systems, which is depicted in Chapter 4. To design it, we based on the DSRM approach, which allowed us to have a systematic approach during the creation and maintain a focus on design challenges throughout the journey.

To address the challenges i-vii from the previous section, we first created a gamification design framework that focuses on contextual information, users' preferences and differences, encouraged behaviors, application domain, technological features, game modality, feedback design, game rules, game elements, and aesthetics to guide in the design of gameful systems. Then, to facilitate communication between the various stages of development, and the various professionals, we implemented the framework in a tool that would allow a GDD to be produced automatically. To fully support the team in the whole process, we started working on several supporting features: a peer-review procedure to assess the design goodness, data collection, and methodological rigor (challenges viii, and ix); an evaluation section to guide in the selection of reliable solution for the evaluation (challenge ix); a code generation feature to facilitate the creation of the gamification code and to be able to make an initial check on the interaction between the game elements (challenge vii); an open-access database to suggest the best game elements according to the background information (challenge vii); a design framework list to allow users to use other GamiDOC features without necessarily use the design framework.

7.3 RQ 3: Does the identified procedure adequately support the design phase?

To answer this question, I presented three studies and a use-case scenario depicted in Chapter 5, and Chapter 6. Users reported that the gamification design framework helps during the design phase; specifically, it helps to take into account certain critical elements, such as individual differences, purposes, game elements, and type of feedback. Specifically, the framework's structure allowed the designers to focus on one element at a time and understand all the phases and critical points, structuring the workflow.

As a result, it was easier to understand the weak and strong points of each component. Interestingly, thanks to the analysis of technological components and the creation of the GDD, it proved useful in supporting the decision of systems requirements, allowing the selection of tools, packages, and components suitable for the purpose and users. Users, both experienced and non-expert, also reported that all the other features can be considered extremely helpful for the whole process and that the whole community would benefit from the fully developed tool, suggesting that the direction taken in the development of our artifact follows a correct path.

In conclusion, the identified design framework provides adequate support during the design phase. Moreover, the whole procedure implemented in the online tool will provide support for the rest of the development and evaluation process.

7.4 RQ 4: Does the use of GamiDOC enable the creation of effective systems?

Assessing the actual effectiveness of GamiDOC (**RQ 4**) is a relatively complex part. Utopically, an optimal evaluation should consider differences between several systems developed using GamiDOC and systems developed using different methodologies. However, it is hardly possible to consider these options during a 3-year doctorate, mainly because of the development time and costs of gameful systems. Hence, to address **RQ 4**, I must rely on the consideration reported in Chapter 5, and the results of the two systems developed, specifically analyzing subjective (user experience) and objective (behavioral) measures depicted in Chapter 6.

All in all, results in Chapter 5 show that GamiDOC is useful during the design phase of gameful systems and serious games. As depicted in the previous section, through the design framework, developers and designers can face the various concerns discussed in Chapter 3, producing optimal control of interfering variables during design and, at the same time, developing a greater awareness of gamification design challenges. In both systems presented in Chapter 6, GamiDOC was used for the design phase, enabling communication between the various team members during design and development. Its use helped, as presented in the previous section, in the selection of system requirements, also producing an optimal and rapid transition from the design phase to the development phase. As a result, both tools were appreciated by users and achieved the expected results. Specifically, GaMoVR produced excellent results in terms of subjective outcomes, including usability, player experience (especially motivation, engagement, and fun), and perceived learning. Moreover, it proved to be a more popular system than an alternative considered valid in previous analyses. Similarly, UBG proved motivating and playful, but more importantly, it produced learning related to its use, helping to improve users' environmental awareness and willingness to maintain it in the long term. It is therefore possible to hypothesize that, the guidance that GamiDOC provided allowed for optimal consideration of challenges in the design phase, enabling more attention to be paid to critical elements, such as contextual and technological factors, and fostering the successful outcome of systems, both in subjective and behavioral data.

In conclusion, in all the studies, GamiDOC proved useful and effective in the design and system requirements phase. The reported data suggest that GamiDOC can guide designers through the design and system requirements phases, allowing control over the various elements of the design phase, that the systems created are effective and enjoyable, and that the game elements chosen are appropriate to the context and purpose. Nevertheless, to answer this question more thoroughly, it is imperative to undertake the design and development of additional systems utilizing GamiDOC in the future and make further evaluations, comparing the outcomes (in terms of player experience and behavioral data) of systems created with it against those designed with different methods.

7.5 Future work

The work presented in this thesis has limitations that can be addressed with specific strategies in the future.

Scientometric analyses. The scientometric analyses, depicted in Chapter 2, present a detailed structure of the literature on gamification applied to behavior change and the design of gameful systems. However, although the results obtained are indicative of the literature and in line with other sources, the results shown are affected by a lack of manual selection of articles through the screening and snowballing procedure. In the future, to have more specific supporting data, it is essential to use these procedures.

Challenges. The challenges of gameful systems design, presented in Chapter 3, are the result of the analysis of specific challenges identified in the literature by several authors (Koivisto & Hamari, 2019a; Mora et al., 2017; Seaborn & Fels, 2015). Although the analysis presented can be considered relatively exhaustive, the data is not derived from a systematic procedure. In the future, a systematic analysis of the challenges and a more exhaustive presentation of design frameworks and open challenges could provide a more detailed overview and be more useful for experts in the field.

GamiDOC evaluation. GamiDOC evaluation, depicted in Chapter 5, reports an evaluation of the possible features of GamiDOC and the usefulness of the design framework. However, the studies only analyze the concept behind the tool and not a practical test of its use. Furthermore, the sample sizes used for the evaluations are inadequate to allow high reliability. In the future, it is essential to evaluate the features at the end of the development phase, so that users can actually use the complete tool. Furthermore, it would be optimal to evaluate the use of the framework during the design of real gameful systems, using standardized procedures that allow for greater reliability and generalization.

Systems evaluation. To evaluate the effectiveness of GamiDOC, two systems are presented in Chapter 6. As also presented in the previous sections, the methodology for evaluating systems created with GamiDOC has limitations. In the future, during the analysis phase, it would be necessary to collect information on the usefulness of GamiDOC for the design and development of these systems by analyzing the responses of several team members and using a more structured methodology, including standardized questionnaires, structured interviews, and so forth. Moreover, it is necessary to develop a standardized analysis methodology for each system produced with GamiDOC, based on subjective and objective data, and to evaluate the same participants even after a follow-up. In addition, it will be necessary to have further systems developed using GamiDOC and to use more comparisons with similar systems developed with other methodologies, along with a more reliable sample size.

Our journey to develop GamiDOC has yielded valuable insights. Nevertheless, specific components remain incompletely developed and require further attention to ensure the ongoing enhancement of the tool and its applicability within the gamification design domain (Table 4.1).

Gamification Design Framework. The design framework was developed and implemented in the online tool, and its applications have reported good results. However, the section related to feedback design needs to be updated and we need to report a reliable categorization for the different target users since it is currently a field related to age and employment. Moreover, some usability issues (listed in Chapter 5) need to be resolved, including the addition of further examples. Last, more explicit guidance for designing ethical gamification systems must be included within the framework.

Gamification Design Document. The creation of a GDD is possible from the compilation of the framework. However, the final layout of the document still needs some improvement.

Gameful System Design Peer-review. The login component and the possibility to submit the final design of gameful systems as an attachment with evaluative methods exists in the back end. However, we still need to resolve some ethical constraints before the feature can be released into the tool.

Introducing a Tool Board. The implementation of a tool board on the GamiDOC homepage that provides users with an interactive platform for viewing approved gamification design documents exists in the back end. However, we still need to fix the login feature and submission procedure before proceeding with this.

Code Creation. Incorporating a code generation feature is essential for the success of GamiDOC. While users can define new game elements and mechanics, the common usage involves the instantiation of pre-existing elements and mechanics within game definitions. Currently, code creation from rule definition exists through an external tool called DSL4GaR. However, DSL4GaR needs some refinements, and it is necessary to finalize the user interface that acts as a conduit between GamiGOC and DSL4GaR.

User Experience Analysis Component. Incorporating a guideline for the tool's evaluation is crucial to provide reliable analysis methodologies and for the assessment of the final systems. Currently, we finished the design phase of the component. However, we need to evaluate the prototype, and then continue with the development.

Open-Access Database. The open-access database plays a pivotal role in the gamification design model, helping identify key factors for enhancing the effectiveness of gameful systems. We are working on a semi-systematic literature review to gather data for the database, which will be implemented in GamiDOC. It will be dynamic, allowing for the addition of new data from ongoing analyses and experiments. The database will offer a separate page within GamiDOC, serving as a valuable resource for users.

Gamification Design Frameworks List. The list of gamification design frameworks and strategies has a dual function: first, it functions as a perpetual literature review, and second, it allows users to select a design framework appropriate to their needs. We are working on a systematic literature review to gather data for the list, which will be implemented in GamiDOC.

In conclusion, GamiDOC aspires to establish itself as a prominent reference in the field of gamification design. Through the provision of a diverse array of features and a standardized framework, GamiDOC holds the potential to cultivate a more cohesive and knowledgeable community, providing shared reference points for gamification design. To reach this aim, each feature needs to be assessed in real applications, and more gameful systems and serious games should be developed using GamiDOC in the design phase, and then properly evaluated.

References

Adams, E. (2014). *Fundamentals of game design.* Pearson Education.

Adrián, S.-R., & Elena, P.-G. M. (2019). Active methodologies in health. scientific production on gamification in health sciences. *Science for Education Today*, *9*(3).

Ahola, R., Pyky, R., Jämsä, T., Mäntysaari, M., Koskimäki, H., Ikäheimo, T. M., ... Korpelainen, R. (2013). Gamified physical activation of young men–a multidisciplinary population-based randomized controlled trial (mopo study). *BMC public health*, *13*(1), 1–8.

Akasaki, H., Suzuki, S., Nakajima, K., Yamabe, K., Sakamoto, M., Alexandrova, T., & Nakajima, T. (2016). One size does not fit all: Applying the right game concepts for the right persons to encourage non-game activities. In *International conference on human interface and the management of information* (pp. 103–114).

Alimi, O. S., Farner Budarz, J., Hernandez, L. M., & Tufenkji, N. (2018). Microplastics and nanoplastics in aquatic environments: aggregation, deposition, and enhanced contaminant transport. *Environmental science & technology*, *52*(4), 1704–1724.

Al-Rahmi, W. M., & Othman, M. S. (2013). Evaluating student's satisfaction of using social media through collaborative learning in higher education. *International Journal of advances in engineering & technology*, *6*(4), 1541.

AL-Smadi, M. (2014). Gameducation: using gamification techniques to engage learners in online learning. *European Summit on Immersive Education*, 85–97.

Ames, C., & Felker, D. W. (1979). An examination of children's attributions and achievement-related evaluations in competitive, cooperative, and individualistic reward structures. *Journal of Educational Psychology*, *71*(4), 413.

Armstrong, M. B., & Landers, R. N. (2018). Gamification of employee training and development. *International Journal of Training and Development*, *22*(2), 162–169.

Aryadoust, V. (2020). A review of comprehension subskills: a scientometrics perspective. *System*, *88*, 102180.

Aryadoust, V., Tan, H. A. H., & Ng, L. Y. (2019). A scientometric review of rasch measurement: The rise and progress of a specialty. *Frontiers in psychology*, *10*, 2197.

Aryadoust, V., Zakaria, A., Lim, M. H., & Chen, C. (2021). An extensive knowledge mapping review of measurement and validity in language assessment and sla research. *Frontiers in psychology*, 1941.

Ashford, S. J., & De Stobbeleir, K. E. (2013). Feedback, goal setting, and task performance revisited. In *New developments in goal setting and task performance* (pp. 51–64). Routledge.

Awais, M., Habiba, U., Khalid, H., Shoaib, M., & Arshad, S. (2019). An Adaptive Feedback System to Improve Student Performance Based on Collaborative Behavior. *IEEE Access*, *7*, 107171–107178. doi: 10.1109/ACCESS.2019.2931565

Azouz, O., & Lefdaoui, Y. (2018). Gamification design frameworks: a systematic mapping study. In *2018 6th international conference on multimedia computing and systems (icmcs)* (pp. 1–9).

Bagozzi, R. P., & Dholakia, U. M. (2006). Open source software user communities: A study of participation in linux user groups. *Management science*, *52*(7), 1099–1115.

Baldwin, M. (2005). Game design document outline. *Baldwin Consultant*, *24*.

Banerjee, A., Chitnis, U., Jadhav, S., Bhawalkar, J., & Chaudhury, S. (2009). Hypothesis testing, type i and type ii errors. *Industrial psychiatry journal*, *18*(2), 127.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual sus scores mean: Adding an adjective rating scale. *Journal of usability studies*, *4*(3), 114–123.

Baranowski, T., Buday, R., Thompson, D. I., & Baranowski, J. (2008). Playing for real: video games and stories for health-related behavior change. *American journal of* *preventive medicine*, *34*(1), 74–82.

Barata, G., Gama, S., Jorge, J., & Gonçalves, D. (2013). Engaging engineering students with gamification. In *2013 5th international conference on games and virtual worlds for serious applications (vs-games)* (pp. 1–8).

Bargh, J. (1994). The four horsemen of automaticity: Awareness, intention, efficiency. *Handbook of social cognition: Basic processes*, 1040.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century—a review. *Journal of informetrics*, *2*(1), 1–52.

Barker, J. R., & Tingey, D. T. (2012). *Air pollution effects on biodiversity.* Springer Science & Business Media.

Barnum, C. M. (2020). *Usability testing essentials: Ready, set... test!* Morgan Kaufmann.

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

Bassanelli, S., & Bucchiarone, A. (2022). Gamidoc: A tool for designing and evaluating gamified solutions. In *Extended abstracts of the 2022 annual symposium on computer-human interaction in play* (pp. 203–208).

Bassanelli, S., Bucchiarone, A., & Gini, F. (2024). Gamidoc: The importance of designing gamification in a proper way.
(Accepter for publication to *IEEE Transactions on Games*)

Bassanelli, S., Gini, F., Bonetti, F., Bucchiarone, A., & Marconi, A. (2024). Lost in gamification design: A scientometric analysis. *26th International Conference on Human-Computer Interaction (HCII 2024).* (accepted for publication)

Bassanelli, S., Gini, F., Bucchiarone, A., & Marconi, A. (2024). Promoting green mobility through gamified transportation campaigns. *8th International GamiFIN Conference 2024 (GamiFIN 2024), April 2-5, 2024, Ruka, Finland*. (in publication)

Bassanelli, S., Vasta, N., Bucchiarone, A., & Marconi, A. (2022). Gamification for behavior change: A scientometric review. *Acta Psychologica*, *228*, 103657.

Bidarra, J., Figueiredo, M., & Natálio, C. (2015). Interactive design and gamification of ebooks for mobile and contextual learning. *International Journal of Interactive*

Mobile Technologies (iJIM), 24–32.

Böckle, M., Micheel, I., Bick, M., & Novak, J. (2018). A design framework for adaptive gamification applications. In *Proceedings of the 51st hawaii international conference on system sciences.*

Boddy, C. R. (2016). Sample size for qualitative research. *Qualitative Market Research: An International Journal*, *19*(4), 426–432.

Bonetti, F., Bassanelli, S., Bucchiarone, A., Gini, F., & Marconi, A. (2024). Untitled bee game: Be(e)ing mean to learn more about eco-sustainability. *8th International GamiFIN Conference 2024 (GamiFIN 2024), April 2-5, 2024, Ruka, Finland*.

Bozkurt, A., & Durak, G. (2018). A systematic review of gamification research: In pursuit of homo ludens. *International Journal of Game-Based Learning (IJGBL)*, *8*(3), 15–33.

Braun, V., Clarke, V., Boulton, E., Davey, L., & McEvoy, C. (2021). The online survey as a qualitative research tool. *International journal of social research methodology*, *24*(6), 641–654.

Brito, J., Vieira, V., & Duran, A. (2015). Towards a framework for gamification design on crowdsourcing systems: The game approach. In *2015 12th international conference on information technology-new generations* (pp. 445–450).

Brooke, J., et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, *189*(194), 4–7.

Brooks, J., Horrocks, C., & King, N. (2018). Interviews in qualitative research. *Interviews in qualitative research*, 1–360.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational researcher*, *18*(1), 32–42.

Brühlmann, F., Vollenwyder, B., Opwis, K., & Mekler, E. D. (2018). Measuring the "why" of interaction: Development and validation of the user motivation inventory (umi). In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–13).

Brummelman, E., & Sedikides, C. (2020). Raising children with high self-esteem (but not narcissism). *Child development perspectives*, *14*(2), 83–89.

Bucchiarone, A., Bassanelli, S., Luca, M., Centellegher, S., Cipriano, P., Giovannini, L., . . . Marconi, A. (2023). Play&go corporate: An end-to-end solution for facilitating urban cyclability. *IEEE Transactions on Intelligent Transportation Systems*.

Bucchiarone, A., Bassanelli, S., & Marconi, A. (2023). How to foster sustainable behaviors through multi-campaigns rewarding mechanisms: The air-break experience. *Sustainability*, *15*(6), 5198.

Bucchiarone, A., Cicchetti, A., Bassanelli, S., & Marconi, A. (2021). How to merge gamification efforts for programming and modelling: a tool implementation perspective. In *2021 acm/ieee international conference on model driven engineering languages and systems companion (models-c)* (pp. 721–726).

Bucchiarone, A., Cicchetti, A., & Marconi, A. (2019a). Exploiting multi-level modelling for designing and deploying gameful systems. In *22nd ACM/IEEE international conference on model driven engineering languages and systems, MODELS 2019, munich, germany, september 15-20, 2019* (pp. 34–44). IEEE.

Bucchiarone, A., Cicchetti, A., & Marconi, A. (2019b). Gdf: A gamification design framework powered by model-driven engineering. In *2019 acm/ieee 22nd international conference on model driven engineering languages and systems companion (models-c)* (pp. 753–758).

Bucchiarone, A., Martella, S., Muccini, H., & Fusco, M. (2023). Dsl4gar: A domain specific language for gamification rules definition, simulation and deployment. *Simulation and Deployment*.

Bucchiarone, A., Martorella, T., Colombo, D., Cicchetti, A., & Marconi, A. (2021). Polyglot for gamified education: Mixing modelling and programming exercises. In *2021 acm/ieee international conference on model driven engineering languages and systems companion (models-c)* (pp. 605–609).

Bucchiarone, A., Savary-Leblanc, M., Pallec, X. L., Bruel, J., Cicchetti, A., Cabot, J., . . . Perillo, M. (2020). Papyrus for gamers, let's play modeling. In E. Guerra & L. Iovino (Eds.), *MODELS '20: ACM/IEEE 23rd international conference on model driven engineering languages and systems, virtual event, canada, 18-23 october, 2020, companion proceedings* (pp. 5:1–5:5). ACM. Retrieved from https://doi.org/10.1145/3417990.3422002 doi: 10.1145/3417990.3422002

Bucchiarone, A., Savary-Leblanc, M., Le Pallec, X., Cicchetti, A., Gérard, S., Bassanelli, S., ... Marconi, A. (2023). Gamifying model-based engineering: the papygame experience. *Software and Systems Modeling*, 1–21.

Bucchiarone, A., Savary-Leblanc, M., Pallec, X. L., Bruel, J.-M., Cicchetti, A., Cabot, J., ... Perillo, M. (2020). Papyrus for gamers, let's play modeling. In *Proceedings of the 23rd acm/ieee international conference on model driven engineering languages and systems: Companion proceedings* (pp. 1–5).

Burns, E. C., Martin, A. J., & Evans, P. A. (2021). The role of teacher feedback–feedforward and personal best goal setting in students' mathematics achievement: A goal setting theory perspective. *Educational Psychology*, *41*(7), 825–843.

Busch, M., Mattheiss, E., Reisinger, M., Orji, R., Fröhlich, P., & Tscheligi, M. (2016). More than sex: The role of femininity and masculinity in the design of personalized persuasive games. In *International conference on persuasive technology* (pp. 219–229).

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., ... others (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436.

Campbell, J. I., & Thompson, V. A. (2012). Morepower 6.0 for anova with relational confidence intervals and bayesian analysis. *Behavior research methods*, *44*(4), 1255–1265.

Cantador, I., & Conde, J. M. (2010). Effects of competition in education: A case study in an e-learningenvironment.

Caracciolo, Marco. (2021). Animal mayhem games and nonhuman-oriented thinking. *GAME STUDIES (COPENHAGEN)*, *21*(1), 16.

Carlson, C. (1979). Feedback for learning. *On College Training. San Francisco CA: Jossey-Bass*.

Carollo, A., Balagtas, J. P. M., Neoh, M. J.-Y., & Esposito, G. (2021). A scientometric approach to review the role of the medial preoptic area (mpoa) in parental behavior. *Brain Sciences*, *11*(3), 393.

Carollo, A., Bonassi, A., Lim, M., Gabrieli, G., Setoh, P., Dimitriou, D., ... Esposito, G. (2021). Developmental disabilities across the world: A scientometric review from 1936 to 2020. *Research in Developmental Disabilities*, *117*, 104031.

Celik, S. (2020). The effects of climate change on human behaviors. *Environment, climate, plant and vegetation growth*, 577–589.

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature human behaviour*, *6*(1), 29–42.

Chen, C. (2014). The citespace manual. *College of Computing and Informatics*, *1*, 1–84.

Chen, C. (2016). *Citespace: a practical guide for mapping scientific literature*. Nova Science Publishers Hauppauge, NY.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, *3*(3), 191–209.

Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for information Science and Technology*, *61*(7), 1386–1409.

Chen, C., & Morris, S. (2003). Visualizing evolving networks: Minimum spanning trees versus pathfinder networks. In *Ieee symposium on information visualization 2003 (ieee cat. no. 03th8714)* (pp. 67–74).

Chen, C., Song, I.-Y., Yuan, X., & Zhang, J. (2008). The thematic and citation landscape of data and knowledge engineering (1985–2007). *Data & Knowledge Engineering*, *67*(2), 234–259.

Chen, X., & Liu, Y. (2020). Visualization analysis of high-speed railway research based on citespace. *Transport Policy*, *85*, 1–17.

Chen, Y., & Pu, P. (2014). Healthytogether: exploring social incentives for mobile fitness applications. In *Proceedings of the second international symposium of chinese chi* (pp. 25–34).

Cheng, R., & Vassileva, J. (2005). User motivation and persuasion strategy for peer-to-peer communities. In *Proceedings of the 38th annual hawaii international conference on system sciences* (pp. 193a–193a).

Cheon, S. H., Reeve, J., & Song, Y.-G. (2019). Recommending goals and supporting needs: An intervention to help physical education teachers communicate their expectations while supporting students' psychological needs. *Psychology of Sport and Exercise*, *41*, 107–118.

Chin, S. (2014). Mobile technology and gamification: The future is now! In *2014 fourth international conference on digital information and communication technology and its applications (dictap)* (pp. 138–143).

Chou, Y.-k. (2019). *Actionable gamification: Beyond points, badges, and leaderboards.* Packt Publishing Ltd.

Codish, D., & Ravid, G. (2014). Personality based gamification-educational gamification for extroverts and introverts. In *Proceedings of the 9th chais conference for the study of innovation and learning technologies: Learning in the technological era* (Vol. 1, pp. 36–44).

Codish, D., & Ravid, G. (2017). Gender moderation in gamification: does one size fit all?

Coroller, S., & Flinois, C. (2023). Video games as a tool for ecological learning: the case of Animal Crossing. *Ecosphere*, *14*(3), e4463.

Csikszentmihalyi, M. (1991). *Flow.* Munksgaard København.

Csikszentmihalyi, M., & Csikzentmihaly, M. (1990). *Flow: The psychology of optimal experience* (Vol. 1990). Harper & Row New York.

Cutts, Q., Carbone, A., & van Haaster, K. (2004). Using an electronic voting system to promote active reflection on coursework feedback. In *Proceedings of intl. conf. on computers in education, melbourne, australia.*

Dæhli, O., Kristoffersen, B., jr, P. L., & Sandnes, T. (2021). Exploring feedback and gamification in a data modeling learning tool. *Electronic Journal of e-Learning*.

Daniels, A. C. (2000). *Bringing out the best in people.* McGraw-Hill Education.

Deci, E. L., Betley, G., Kahle, J., Abrams, L., & Porac, J. (1981). When trying to win: Competition and intrinsic motivation. *Personality and social psychology bulletin*, *7*(1), 79–83.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, *125*(6), 627.

Deci, E. L., et al. (1985). i ryan, rm (1985). *Intrinsic motivation and self-determination* *in human behavior*, *1*.

Deci, E. L., & Ryan, R. M. (2013). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how* (Vol. 488). Springer.

Denden, M., Tlili, A., Essalmi, F., & Jemni, M. (2017). An investigation of the factors affecting the perception of gamification and game elements. In *2017 6th international conference on information and communication technology and accessibility (icta)* (pp. 1–6).

DeNisi, A. S., & Kluger, A. N. (2000). Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Perspectives*, *14*(1), 129–139.

Deterding, S. (2012). Gamification: designing for motivation. *interactions*, *19*(4), 14–17.

Deterding, S. (2015). The lens of intrinsic skill atoms: A method for gameful design. *Human–Computer Interaction*, *30*(3-4), 294–335.

Deterding, S., Björk, S. L., Nacke, L. E., Dixon, D., & Lawley, E. (2013). Designing gamification: creating gameful and playful experiences. In *Chi'13 extended abstracts on human factors in computing systems* (pp. 3263–3266).

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic mindtrek conference: Envisioning future media environments* (pp. 9–15).

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification. using game-design elements in non-gaming contexts. In *Chi'11 extended abstracts on human factors in computing systems* (pp. 2425–2428).

Deutsch, M. (1949). A theory of co-operation and competition. *Human relations*, *2*(2), 129–152.

Deutsch, M. (1968). Field theory in social psychology. *The handbook of social psychology*, *1*, 412–487.

Dev, P. C. (1997). Intrinsic motivation and academic achievement: What does their relationship imply for the classroom teacher? *Remedial and special education*, *18*(1), 12–19.

Dicheva, D., Dichev, C., Irwin, K., Jones, E. J., Cassel, L. B., & Clarke, P. J. (2019). Can game elements make computer science courses more attractive? In *Proceedings of the 50th ACM technical symposium on computer science education, SIGCSE 2019* (p. 1245).

Domínguez, A., Saenz-de Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & education*, *63*, 380–392.

Duncan, P. K., & Bruwelheide, L. R. (1985). Feedback: Use and possible behavioral functions. *Journal of Organizational Behavior Management*, *7*(3-4), 91–114.

Dye, M. W., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current directions in psychological science*, *18*(6), 321–326.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.

Epstein, D. S., Zemski, A., Enticott, J., & Barton, C. (2021). Tabletop board game elements and gamification interventions for health behavior change: realist review and proposal of a game design framework. *JMIR serious games*, *9*(1), e23302.

Erdfelder, E., Faul, F., & Buchner, A. (1996). Gpower: A general power analysis program. *Behavior research methods, instruments, & computers*, *28*(1), 1–11.

Erev, I., Bornstein, G., & Galili, R. (1993). Constructive intergroup competition as a solution to the free rider problem: A field experiment. *Journal of Experimental Social Psychology*, *29*(6), 463–478.

Ernkvist, M. (2008). Down many times, but still playing the game: Creative destruction and industry crashes in the early video game industry 1971-1986.

Farzan, R., DiMicco, J. M., Millen, D. R., Dugan, C., Geyer, W., & Brownholtz, E. A. (2008). Results from deploying a participation incentive mechanism within the enterprise. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 563–572).

Fernandes, F. T., & Junior, P. T. A. (2016). Gamification aspects in the context of electronic government and education: A case study. In *International conference on hci in business, government, and organizations* (pp. 140–150).

Fernández Galeote, D., Legaki, N.-Z., & Hamari, J. (2023, October). Climate Connected: An immersive VR and PC game for climate change engagement. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (pp. 266–273). Stratford ON Canada: ACM. Retrieved 2024-01-09, from https://dl.acm.org/doi/10.1145/3573382.3616053 doi: 10.1145/3573382.3616053

Fishbach, A., Eyal, T., & Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. *Social and Personality Psychology Compass*, *4*(8), 517–530.

Fogg, B. J. (2002). Persuasive technology: using computers to change what we think and do. *Ubiquity*, *2002*(December), 2.

Forbes. (2019). *Indie darling 'Untitled Goose Game' clears one million copies sold in three months.* Retrieved 2023-11-13, from https://www.forbes.com/sites/lizlanier/2020/12/31/indie-darling-untitled-goose-game-clears-one-million-copies-sold-in-three-months/?sh=2f1695b84884

Fortes Tondello, G., Premsukh, H., & Nacke, L. (2018). A theory of gamification principles through goal-setting theory..

Foster, E. D., & Deardorff, A. (2017). Open science framework (osf). *Journal of the Medical Library Association: JMLA*, *105*(2), 203.

Franceschini, S., Gori, S., Ruffino, M., Viola, S., Molteni, M., & Facoetti, A. (2013). Action video games make dyslexic children read better. *Current biology*, *23*(6), 462–466.

Francis, J. J., Johnston, M., Robertson, C., Glidewell, L., Entwistle, V., Eccles, M. P., & Grimshaw, J. M. (2010). What is an adequate sample size? operationalising data saturation for theory-based interview studies. *Psychology and health*, *25*(10), 1229–1245.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35–41.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. *Journal of experimental psychology: General*, *141*(1), 2.

Fuchs, M., & Wolff, C. (2016). Improving programming education through gameful, formative feedback. In *2016 ieee global engineering education conference (educon)* (pp. 860–867).

Fulu, I. (2007). Enhancing learning through competitions. school of infocomm technology. *Ngee Ann Polytechnic*.

Fuß, C., Steuer, T., Noll, K., & Miede, A. (2014). Teaching the achiever, explorer, socializer, and killer–gamification in university education. In *International conference on serious games* (pp. 92–99).

Gabrielle, V. (2018). The dark side of gamifying work. *Fast Company*, *1*, 2018.

Gaggero, G., Bonassi, A., Dellantonio, S., Pastore, L., Aryadoust, V., & Esposito, G. (2020). A scientometric review of alexithymia: Mapping thematic and disciplinary shifts in half a century of research. *Frontiers in psychiatry*, *11*, 1405.

Gardner, H. J., & Martin, M. A. (2007). Analyzing ordinal scales in studies of virtual environments: Likert or lump it! *Presence: Teleoperators and Virtual Environments*, *16*(4), 439–446.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in entertainment (CIE)*, *1*(1), 20–20.

Gini, F., & Bassanelli, S. (2022). *The relationship between fun and the overall score of the meega360 scale.* Retrieved from https://assets.researchsquare.com/files/rs-1898746/v1/0ec676b0-eadc-450a-9eff-f02600e384f3.pdf?c=1659553858 (Preprint available at ResearchSquare)

Gini, F., Bassanelli, S., & Bucchiarone, A. (2023). The role of game modality in the outcomes of gamification: A research agenda. *7th International GamiFIN Conference 2023 (GamiFIN 2023), April 18-21, 2023, Lapland, Finland*.

Goh, D. H., & Lee, C. S. (2011). Perceptions, quality and motivational needs in image tagging human computation games. *Journal of Information Science*, *37*(5), 515–531.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, *48*(1), 26.

Gomez-Jaramillo, S., Moreno-Cadavid, J., & Zapata-Jaramillo, C. M. (2018). Adaptation of the 6d gamification model in a software development course. In *2018 xiii* *latin american conference on learning technologies (laclo)* (pp. 85–88).

Gregory, J. B., Beck, J. W., & Carr, A. E. (2011). Goals, feedback, and self-regulation: Control theory as a natural framework for executive coaching. *Consulting Psychology Journal: Practice and Research*, *63*(1), 26.

Griffiths, M. (2003). The therapeutic use of videogames in childhood and adolescence. *Clinical child psychology and psychiatry*, *8*(4), 547–554.

Haider, A., Harteveld, C., Johnson, D., Birk, M. V., Mandryk, R. L., Seif El-Nasr, M., ... Vanden Abeele, V. (2022). minipxi: Development and validation of an eleven-item measure of the player experience inventory. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CHI PLAY), 1–26.

Hallifax, S., Lavoué, E., & Serna, A. (2020). To tailor or not to tailor gamification? an analysis of the impact of tailored game elements on learners' behaviours and motivation. In *International conference on artificial intelligence in education* (pp. 216–227).

Hallifax, S., Serna, A., Marty, J.-C., Lavoué, G., & Lavoué, E. (2019). Factors to consider for tailored gamification. In *Proceedings of the annual symposium on computer-human interaction in play* (pp. 559–572).

Hamari, J. (2013). Transforming homo economicus into homo ludens: A field experiment on gamification in a utilitarian peer-to-peer trading service. *Electronic commerce research and applications*, *12*(4), 236–245.

Hamari, J. (2017). Do badges increase user activity? a field experiment on the effects of gamification. *Computers in human behavior*, *71*, 469–478.

Hamari, J. (2019). Gamification. *The Blackwell encyclopedia of sociology*, 1–3.

Hamari, J., Huotari, K., & Tolvanen, J. (2015). Gamification and economics. *The gameful world: Approaches, issues, applications*, *139*, 15.

Hamari, J., & Koivisto, J. (2013). Social motivations to use gamification: an empirical study of gamifying exercise.

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th hawaii international conference on system sciences* (pp. 3025–3034).

Hamari, J., & Tuunanen, J. (2014). Player types: A meta-synthesis.

Hammerschall, U. (2019). A gamification framework for long-term engagement in education based on self determination theory and the transtheoretical model of change. In *2019 ieee global engineering education conference (educon)* (pp. 95–101).

Hanghøj, T., Lieberoth, A., & Misfeldt, M. (2018). Can cooperative video games encourage social and motivational inclusion of at-risk students? *British Journal of Educational Technology*, *49*(4), 775–799.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & education*, *80*, 152–161.

Harro-Loit, H. (2019). Feedback and feedforward as a dialogic communication in the learning environment. *The Essence of Academic Performance*, *18*, 2019.

Hassan, L., & Hamari, J. (2020). Gameful civic engagement: A review of the literature on gamification of e-participation. *Government Information Quarterly*, *37*(3), 101461.

Hassan, M. A., Habiba, U., Khalid, H., Shoaib, M., & Arshad, S. (2019). An adaptive feedback system to improve student performance based on collaborative behavior. *Ieee Access*, *7*, 107171–107178.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81–112.

Hermsen, S., Frost, J., Renes, R. J., & Kerkhof, P. (2016). Using feedback through digital technology to disrupt and change habitual behavior: A critical review of current literature. *Computers in Human Behavior*, *57*, 61–74.

Herzig, P., Jugel, K., Momm, C., Ameling, M., & Schill, A. (2013). Gaml-a modeling language for gamification. In *2013 ieee/acm 6th international conference on utility and cloud computing* (pp. 494–499).

Higgins, R., Hartley, P., & Skelton, A. (2001). Getting the message across: the problem of communicating assessment feedback. *Teaching in higher education*, *6*(2), 269–274.

Higgins, R., Hartley, P., & Skelton, A. (2002). The conscientious consumer: Reconsidering the role of assessment feedback in student learning. *Studies in higher* *education*, *27*(1), 53–64.

Högberg, J., Hamari, J., & Wästlund, E. (2019). Gameful experience questionnaire (gamefulquest): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction*, *29*(3), 619–660.

Hogg, M. A., & Turner, J. C. (1985). Interpersonal attraction, social identification and psychological group formation. *European journal of social psychology*, *15*(1), 51–66.

Hölker, F., Wolter, C., Perkin, E. K., & Tockner, K. (2010). Light pollution as a biodiversity threat. *Trends in ecology & evolution*, *25*(12), 681–682.

Holman, C., Aguilar, S., & Fishman, B. (2013). Gradecraft: What can we learn from a game-inspired learning management system? In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 260–264).

Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education*, *125*, 254–272.

Huang, C.-K., Neylon, C., Brookes-Kenworthy, C., Hosking, R., Montgomery, L., Wilson, K., & Ozaygen, A. (2020). Comparison of bibliographic data sources: Implications for the robustness of university rankings. *Quantitative Science Studies*, *1*(2), 445–478.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). Mda: A formal approach to game design and game research. In *Proceedings of the aaai workshop on challenges in game ai* (Vol. 4, p. 1722).

Hunter, D., & Werbach, K. (2012). *For the win.* Wharton digital press.

Huotari, K., & Hamari, J. (2011). Gamification" from the perspective of service marketing. In *Proc. chi 2011 workshop gamification.*

Huotari, K., & Hamari, J. (2012). Defining gamification: a service marketing perspective. In *Proceeding of the 16th international academic mindtrek conference* (pp. 17–22).

Huotari, K., & Hamari, J. (2017). A definition for gamification: anchoring gamification in the service marketing literature. *Electronic Markets*, *27*(1), 21–31.

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation:

the 10±2 rule. *Communications of the ACM*, *53*(5), 130–133.

Hyrynsalmi, S., Kimppa, K. K., & Smed, J. (2017a). Gamification ethics. In *Encyclopedia of computer graphics and games* (pp. 1–6). Springer International Publishing AG.

Hyrynsalmi, S., Kimppa, K. K., & Smed, J. (2017b). Gamification ethics. *Encyclopedia of Computer Graphics and Games, Springer International Publishing, Cham*, 1–6.

IJsselsteijn, W. A., De Kort, Y. A., & Poels, K. (2013). The game experience questionnaire.

Itoko, T., Arita, S., Kobayashi, M., & Takagi, H. (2014). Involving senior workers in crowdsourced proofreading. In *International conference on universal access in human-computer interaction* (pp. 106–117).

Jahn, K., Kordyaka, B., Machulska, A., Eiler, T. J., Gruenewald, A., Klucken, T., ... Niehaves, B. (2021). Individualized gamification elements: The impact of avatar and feedback design on reuse intention. *Computers in Human Behavior*, *119*, 106702.

Jensen, J. L., Kummer, T. A., & Godoy, P. D. d. M. (2015). Improvements from a flipped classroom may simply be the fruits of active learning. *CBE—Life Sciences Education*, *14*(1), ar5.

Jia, Y., Liu, Y., Yu, X., & Voida, S. (2017). Designing leaderboards for gamification: Perceived differences based on user ranking, application domain, and personality traits. In *Proceedings of the 2017 chi conference on human factors in computing systems* (pp. 1949–1960).

Jia, Y., Xu, B., Karanam, Y., & Voida, S. (2016). Personality-targeted gamification: a survey study on personality traits and motivational affordances. In *Proceedings of the 2016 chi conference on human factors in computing systems* (pp. 2001–2013).

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). Big five inventory. *Journal of Personality and Social Psychology*.

Johnson, C., McGill, M., Bouchard, D., Bradshaw, M. K., Bucheli, V. A., Merkle, L. D., ... others (2016). Game development for computer science education. In *Proceedings of the 2016 iticse working group reports* (pp. 23–44).

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet interventions*, *6*, 89–106.

Johnson, D., Gardner, M. J., & Perry, R. (2018). Validation of two game experience scales: the player experience of need satisfaction (pens) and game experience questionnaire (geq). *International Journal of Human-Computer Studies*, *118*, 38–46.

Johnson, D., Horton, E., Mulcahy, R., & Foth, M. (2017). Gamification and serious games within the domain of domestic energy consumption: A systematic review. *Renewable and Sustainable Energy Reviews*, *73*, 249–264.

Johnson, D. W. (2003). Social interdependence: interrelationships among theory, research, and practice. *American psychologist*, *58*(11), 934.

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research.* Interaction Book Company.

Johnson, D. W., & Johnson, R. T. (2005). Training for cooperative group work. *MA, West, D., Tjosvold, KG Smith,(Eds.), The essentials of teamworking: International perspectives*, 131–147.

Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational researcher*, *38*(5), 365–379.

Johnson, R. E., Chang, C.-H., & Lord, R. G. (2006). Moving from cognition to behavior: What the research says. *Psychological bulletin*, *132*(3), 381.

Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active learning in higher education*, *14*(1), 63–76.

Julian, J. W., & Perry, F. A. (1967). Cooperation contrasted with intra-group and inter-group competition. *Sociometry*, 79–90.

Julius, K., & Salo, J. (2013). Designing gamification. *Marketing*.

Jung, J., Schneider, C., & Valacich, J. (2010). Enhancing the motivational affordance of information systems: The effects of real-time performance feedback and goal setting in group collaboration environments. *Management science*, *56*(4), 724–742.

Jurgelaitis, M., Ceponiene, L., & Drungilas, V. (2018). Using gamification for teaching uml in information system design course. *CEUR-WS*, *2145*, 88–94.

Kapp, K. M. (2012). *The gamification of learning and instruction: game-based methods and strategies for training and education.* John Wiley & Sons.

Kappen, D. L., & Nacke, L. E. (2013). The kaleidoscope of effective gamification: deconstructing gamification in business applications. In *Proceedings of the first international conference on gameful design, research, and applications* (pp. 119–122).

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological methods*, *17*(2), 137.

Khaleghi, A., Aghaei, Z., Mahdavi, M. A., et al. (2021). A gamification framework for cognitive assessment and cognitive training: qualitative study. *JMIR serious games*, *9*(2), e21900.

Kim, B. (2015). Designing gamification in the right way. *Library technology reports*, *51*(2), 29–35.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data mining and knowledge discovery*, *7*(4), 373–397.

Klevers, M., Sailer, M., & Günthner, W. A. (2016). Implementation model for the gamification of business processes: a study from the field of material handling. In *Simulation and gaming in the network society* (pp. 173–184). Springer.

Klock, A. C. T., Gasparini, I., & Pimenta, M. S. (2016). 5w2h framework: a guide to design, develop and evaluate the user-centered gamification. In *Proceedings of the 15th brazilian symposium on human factors in computing systems* (pp. 1–10).

Klock, A. C. T., Gasparini, I., Pimenta, M. S., & Hamari, J. (2020). Tailored gamification: A review of literature. *International Journal of Human-Computer Studies*, *144*, 102495.

Koivisto, J., & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, *35*, 179–188.

Koivisto, J., & Hamari, J. (2019a). Gamification of physical activity: A systematic literature review of comparison studies. In *3rd international gamifin conference, gamifin 2019.*

Koivisto, J., & Hamari, J. (2019b). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191–210.

Kölln, K. (2022). Maybe we don't need a new gamification framework after all. In *Extended abstracts of the 2022 annual symposium on computer-human interaction in play* (pp. 384–387).

Komaki, J., Barwick, K. D., & Scott, L. R. (1978). A behavioral approach to occupational safety: pinpointing and reinforcing safe performance in a food manufacturing plant. *Journal of applied Psychology*, *63*(4), 434.

Krijgsman, C., Mainhard, T., Borghouts, L., van Tartwijk, J., & Haerens, L. (2021). Do goal clarification and process feedback positively affect students' need-based experiences? a quasi-experimental study grounded in self-determination theory. *Physical Education and Sport Pedagogy*, *26*(5), 483–503.

Kruger, A., & DeNisi, A. (1996). The effects of feedback interventions on performance: An historical review, meta-analysis and preliminary feedback theory. *Psychological Bulletin*, *119*, 254–285.

Kumar, J. (2013). Gamification at work: Designing engaging business software. In *International conference of design, user experience, and usability* (pp. 528–537).

Lakens, D. (2022). Sample size justification. *Collabra: psychology*, *8*(1), 33267.

Lam, S.-f., Yim, P.-s., Law, J. S., & Cheung, R. W. (2001). The effects of classroom competition on achievement motivation.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & gaming*, *45*(6), 752–768.

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification science, its history and future: Definitions and a research agenda. *Simulation & Gaming*, *49*(3), 315–337.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A goal setting experiment. *Computers in Human Behavior*, *71*, 508–515.

Landers, R. N., Bauer, K. N., Callan, R. C., & Armstrong, M. B. (2015). Psychological theory and the gamification of learning. *Gamification in education and business*, 165–186.

Landers, R. N., & Landers, A. K. (2014). An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. *Simulation & Gaming*, *45*(6), 769–785.

Largan, C., & Morris, T. (2019). *Qualitative secondary research: A step-by-step guide.* Sage.

Latham, G. P., & Baldes, J. J. (1975). The" practical significance" of locke's theory of goal setting. *Journal of Applied Psychology*, *60*(1), 122.

Lee, T. Y., Dugan, C., Geyer, W., Ratchford, T., Rasmussen, J., Shami, N. S., & Lupushor, S. (2013). Experiments on motivational feedback for crowdsourced workers. In *Proceedings of the international aaai conference on web and social media* (Vol. 7, pp. 341–350).

Leonard, T. C. (2008). *Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness.* Springer.

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and social Psychology*, *28*(1), 129.

Lewin, K. (1948). Resolving social conflicts; selected papers on group dynamics.

Lewin, K. (1951). Field theory in social science: selected theoretical papers (edited by dorwin cartwright.).

Lewis, J. R. (2018). The system usability scale: past, present, and future. *International Journal of Human–Computer Interaction*, *34*(7), 577–590.

Leydesdorff, L., & Milojević, S. (2012). Scientometrics. *arXiv preprint arXiv:1208.4566*.

Li, W., Grossman, T., & Fitzmaurice, G. (2012). Gamicad: a gamified tutorial system for first time autocad users. In *Proceedings of the 25th annual acm symposium on user interface software and technology* (pp. 103–112).

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.

Lipnevich, A. A., & Smith, J. K. (2009). "i really need feedback to learn:" students' perspectives on the effectiveness of the differential feedback messages. *Educational*

Assessment, Evaluation and Accountability, *21*, 347–367.

Liu, D., Li, X., & Santhanam, R. (2013). Digital games and beyond: What happens when players compete? *Mis Quarterly*, 111–124.

Locke, E., & Latham, G. (2006). New directions in goal-setting theory. *Current directions in psychological science*, *15*(5), 265–268.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

Lofland, J., Snow, D., Anderson, L., & Lofland, L. H. (2022). *Analyzing social settings: A guide to qualitative observation and analysis*. Waveland Press.

Louangrath, P. (2017). Minimum sample size method based on survey scales. *Int. J. Res. Methodol. Soc. Sci*, *3*(3), 44–52.

Lunenburg, F. C. (2011). Goal-setting theory of motivation. *International journal of management, business, and administration*, *15*(1), 1–6.

Lymn, J. S., & Mostyn, A. (2010). Audience response technology: Engaging and empowering non-medical prescribing students in pharmacology learning. *BMC medical education*, *10*, 1–10.

Mahendra, I. (2016). Contextual learning approach and performance assessment in mathematics learning. *International Research Journal of Management, IT & Social Sciences*, *3*(3), 7–15.

Mainetti, R., Sedda, A., Ronchetti, M., Bottini, G., & Borghese, N. (2013). Duck-neglect: video-games based neglect rehabilitation. *Technology and Health Care*, *21*(2), 97–111.

Marache-Francisco, C., & Brangier, E. (2014). The gamification experience: Uxd with a gamification background. In *Emerging research and trends in interactivity and the human-computer interface* (pp. 205–223). IGI Global.

Marczewski, A. (2013). *Gamification: a simple introduction*. Andrzej Marczewski.

Marczewski, A. (2017). The ethics of gamification. *XRDS: Crossroads, The ACM Magazine for Students*, *24*(1), 56–59.

Marker, A. M., & Staiano, A. E. (2015). Better together: outcomes of cooperation versus competition in social exergaming. *Games for health journal*, *4*(1), 25–30.

Martinho, D., Carneiro, J., Corchado, J. M., & Marreiros, G. (2020). A systematic review of gamification techniques applied to elderly care. *Artificial Intelligence Review*, *53*(7), 4863–4901.

Massung, E., Coyle, D., Cater, K. F., Jay, M., & Preist, C. (2013). Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 371–380).

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport*, *60*(1), 48–58.

McGonigal, J. (2011). *Reality is broken: Why games make us better and how they can change the world*. Penguin.

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). Do points, levels and leaderboards harm intrinsic motivation? an empirical analysis of common gamification elements. In *Proceedings of the first international conference on gameful design, research, and applications* (pp. 66–73).

Minge, M., Thüring, M., Wagner, I., & Kuhr, C. V. (2017). The mecue questionnaire: a modular tool for measuring user experience. In *Advances in ergonomics modeling, usability & special populations: Proceedings of the ahfe 2016 international conference on ergonomics modeling, usability & special populations, july 27-31, 2016, walt disney world®, florida, usa* (pp. 115–128).

Mitchell, A., Petter, S., & Harris, A. (2017). Learning by doing: Twenty successful active learning exercises for information systems courses. *Journal of Information Technology Education. Innovations in Practice*, *16*, 21.

Miyahara, A. (2004). Toward theorizing japanese interpersonal communication competence from a non-western perspective. *Intercultural communication: A global reader*, 279–292.

Mora, A., Riera, D., Gonzalez, C., & Arnedo-Moreno, J. (2015). A literature review of gamification design frameworks. In *2015 7th international conference on games and virtual worlds for serious applications (vs-games)* (pp. 1–8).

Mora, A., Riera, D., González, C., & Arnedo-Moreno, J. (2017). Gamification: a systematic review of design frameworks. *Journal of Computing in Higher Education*, *29*(3), 516–548.

Mora, A., Zaharias, P., González, C., & Arnedo-Moreno, J. (2015). Fraggle: a framework for agile gamification of learning experiences. In *International conference on games and learning alliance* (pp. 530–539).

Morgan, D. L. (1996). *Focus groups as qualitative research* (Vol. 16). Sage publications.

Morris, M. G., & Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing work force. *Personnel psychology*, *53*(2), 375–403.

Morrison, B. B., & DiSalvo, B. (2014). Khan academy gamifies computer science. In *Proceedings of the 45th acm technical symposium on computer science education* (pp. 39–44).

Morschheuser, B., Hamari, J., Koivisto, J., & Maedche, A. (2017). Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human-Computer Studies*, *106*, 26–43.

Morschheuser, B., Hamari, J., & Maedche, A. (2019). Cooperation or competition–when do people contribute more? a field experiment on gamification of crowdsourcing. *International Journal of Human-Computer Studies*, *127*, 7–24.

Morschheuser, B., Hamari, J., Werder, K., & Abe, J. (2017). How to gamify? a method for designing gamification. In *Proceedings of the 50th hawaii international conference on system sciences 2017.*

Morschheuser, B., Hassan, L., Werder, K., & Hamari, J. (2018). How to design gamification? a method for engineering gamified software. *Information and Software Technology*, *95*, 219–237.

Morschheuser, B., Maedche, A., & Walter, D. (2017). Designing cooperative gamification: Conceptualization and prototypical implementation. In *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing* (pp. 2410–2421).

Morschheuser, B., Riar, M., Hamari, J., & Maedche, A. (2017). How games induce cooperation? a study on the relationship between game features and we-intentions in an augmented reality game. *Computers in human behavior*, *77*, 169–183.

Mouratidis, A., Vansteenkiste, M., Michou, A., & Lens, W. (2013). Perceived structure and achievement goals as predictors of students' self-regulated learning and affect and the mediating role of competence need satisfaction. *Learning and Individual Differences*, *23*, 179–186.

Mousavi, S. A., & Khodadoost, F. (2019). Effects of detergents on natural ecosystems and wastewater treatment processes: a review. *Environmental Science and Pollution Research*, *26*, 26439–26448.

Nacke, L. E., Bateman, C., & Mandryk, R. L. (2014). Brainhex: A neurobiological gamer typology survey. *Entertainment compauting*, *5*(1), 55–62.

Nah, F. F.-H., Telaprolu, V. R., Rallapalli, S., & Venkata, P. R. (2013). Gamification of education using computer games. In *International conference on human interface and the management of information* (pp. 99–107).

Nakajima, T., & Lehdonvirta, V. (2013). Designing motivation using persuasive ambient mirrors. *Personal and ubiquitous computing*, *17*(1), 107–126.

Nicholson, S. (2015). A recipe for meaningful gamification. *Gamification in education and business*, 1–20.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, *31*(2), 199–218.

Nilsen, P. (2020). Making sense of implementation theories, models, and frameworks. In *Implementation science 3.0* (pp. 53–79). Springer.

Nilsen, P., & Bernhardsson, S. (2019). Context matters in implementation science: a scoping review of determinant frameworks that describe contextual determinants for implementation outcomes. *BMC health services research*, *19*(1), 1–21.

O'Donovan, S., Gain, J., & Marais, P. (2013). A case study in the gamification of a university-level games development course. In *Proceedings of the south african institute for computer scientists and information technologists conference* (pp. 242–251).

Oja, H., & Randles, R. H. (2004). Multivariate nonparametric tests.

Oja, M., & Riekki, J. (2012). Ubiquitous framework for creating and evaluating persuasive applications and games. In *International conference on grid and pervasive computing* (pp. 133–140).

O'Keefe, D. J. (2007). Brief report: Post hoc power, observed power, a priori power, retrospective power, prospective power, achieved power: Sorting out appropriate uses of statistical power analyses. *Communication methods and measures*, *1*(4), 291–299.

Oliveira, W., Hamari, J., Shi, L., Toda, A. M., Rodrigues, L., Palomino, P. T., & Isotani, S. (2022). Tailored gamification in education: A literature review and future agenda. *Education and Information Technologies*, 1–34.

Orji, R., Vassileva, J., & Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, *24*(5), 453–498.

Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutors' formative feedback: a qualitative interview study. *Assessment & Evaluation in Higher Education*, *30*(4), 369–386.

Ortega Sánchez, D., & Gómez Trigueros, I. M. (2019). Gamification, social problems, and gender in the teaching of social sciences: Representations and discourse of trainee teachers. *Plos one*, *14*(6), e0218869.

Oskamp, S. (2000). A sustainable future for humanity? how can psychology help? *American Psychologist*, *55*(5), 496.

Ovando, M. N. (1994). Constructive feedback: A key to successful teaching and learning. *International Journal of Educational Management*.

Oyibo, K., Orji, R., & Vassileva, J. (2017a). The influence of culture in the effect of age and gender on social influence in persuasive technology. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 47–52).

Oyibo, K., Orji, R., & Vassileva, J. (2017b). Investigation of the social predictors of competitive behavior and the moderating effect of culture. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 419–424).

Palomino, P., Toda, A., Oliveira, W., Rodrigues, L., Cristea, A., & Isotani, S. (2019). Exploring content game elements to support gamification design in educational systems: narrative and storytelling. In *Brazilian symposium on computers in* *education (simpósio brasileiro de informática na educação-sbie)* (Vol. 30, p. 773).

Palomino, P. T., Toda, A. M., Oliveira, W., Cristea, A. I., & Isotani, S. (2019). Narrative for gamification in education: why should you care? In *2019 ieee 19th international conference on advanced learning technologies (icalt)* (Vol. 2161, pp. 97–99).

Panadero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review*, *35*, 100416.

Pat-El, R., Tillema, H., & Van Koppen, S. W. (2012). Effects of formative feedback on intrinsic motivation: Examining ethnic differences. *Learning and Individual Differences*, *22*(4), 449–454.

Patel, V., Pauli, N., Biggs, E., Barbour, L., & Boruff, B. (2021). Why bees are critical for achieving sustainable development. *Ambio*, *50*, 49–59.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), 45–77.

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2018). Meega+: a method for the evaluation of educational games for computing education. *INCoD–Brazilian Institute for Digital Convergence*, 1–47.

Potts, S. G., Imperatriz-Fonseca, V., Ngo, H. T., Aizen, M. A., Biesmeijer, J. C., Breeze, T. D., … others (2016). Safeguarding pollinators and their values to human well-being. *Nature*, *540*(7632), 220–229.

Pranckutė, R. (2021). Web of science (wos) and scopus: The titans of bibliographic information in today's academic world. *Publications*, *9*(1), 12.

Preist, C., Massung, E., & Coyle, D. (2014). Competing or aiming to be average? normification as a means of engaging digital volunteers. In *Proceedings of the 17th acm conference on computer supported cooperative work & social computing* (pp. 1222–1233).

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, *1*(1), 21–21.

Primack, B. A., Carroll, M. V., McNamara, M., Klem, M. L., King, B., Rich, M., …

Nayak, S. (2012). Role of video games in improving health-related outcomes: a systematic review. *American journal of preventive medicine*, *42*(6), 630–638.

Prochaska, J. O., & DiClemente, C. C. (1983). Stages and processes of self-change of smoking: toward an integrative model of change. *Journal of consulting and clinical psychology*, *51*(3), 390.

Prue, D. M., & Fairbank, J. A. (1981). Performance feedback in organizational behavior management: A review. *Journal of Organizational Behavior Management*, *3*(1), 1–16.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of general psychology*, *14*(2), 154–166.

Putz, L.-M., & Treiblmaier, H. (2015). Creating a theory-based research agenda for gamification.

Quintero, J., Baldiris, S., Cerón, J., Garzón, J., Burgos, D., & Vélez, G. (2022). Gamification as support for educational inclusion: The case of ar-mbot. In *2022 international conference on advanced learning technologies (icalt)* (pp. 269–273).

Ramaprasad, A. (1983). On the definition of feedback. *Behavioral science*, *28*(1), 4–13.

Rao, V. (2013). A framework for evaluating behavior change interventions through gaming. In *International conference on advances in computer entertainment technology* (pp. 368–379).

Rapp, A. (2015). A qualitative investigation of gamification: Motivational factors in online gamified services and applications. In *Gamification: Concepts, methodologies, tools, and applications* (pp. 32–48). IGI Global.

Rauschenberger, M., Willems, A., Ternieden, M., & Thomaschewski, J. (2019). Towards the use of gamification frameworks in learning environments. *Journal of Interactive Learning Research*, *30*(2), 147–165.

Reeves, B., & Read, J. L. (2009). *Total engagement: How games and virtual worlds are changing the way people work and businesses compete.* Harvard Business Press.

Reich, J. (2021). Preregistration and registered reports. *Educational Psychologist*, *56*(2), 101–109.

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of general psychology*, *8*(3), 179–193.

Rhee, E., Uleman, J. S., Lee, H. K., & Roman, R. J. (1995). Spontaneous self-descriptions and ethnic identities in individualistic and collectivistic cultures. *Journal of personality and social psychology*, *69*(1), 142.

Robertson, M. (2010). Can't play, won't play. hide & seek: inventing new kinds of play. *Online verfügbar unter http://hideandseek. net/2010/10/06/cant-play-wont-play/, zuletzt aktualisiert am*, *6*, 2010.

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? understanding the principles of gamification. *Business horizons*, *58*(4), 411–420.

Roccas, S., Sagiv, L., Schwartz, S. H., & Knafo, A. (2002). The big five personality factors and personal values. *Personality and social psychology bulletin*, *28*(6), 789–801.

Rogers, R. (2017). The motivational pull of video game feedback, rules, and social interaction: Another self-determination theory approach. *Computers in Human Behavior*, *73*, 446–450.

Rosenthal, R., Cooper, H., Hedges, L., et al. (1994). Parametric measures of effect size. *The handbook of research synthesis*, *621*(2), 231–244.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53–65.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54–67.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion*, *30*(4), 344–360.

Saggah, A., Atkins, A. S., & Campion, R. J. (2020). A review of gamification design frameworks in education. In *2020 fourth international conference on intelligent computing in data sciences (icds)* (pp. 1–8).

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371–380.

Saleem, A. N., Noori, N. M., & Ozdamli, F. (2022). Gamification applications in e-learning: A literature review. *Technol. Knowl. Learn.*, *27*(1), 139–159. Retrieved from https://doi.org/10.1007/s10758-020-09487-x doi: 10.1007/s10758-020-09487-x

Santos, A. C. G., Oliveira, W., Hamari, J., & Isotani, S. (2021). Do people's user types change over time? an exploratory study. *arXiv preprint arXiv:2106.10148*.

Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-health. *Journal of biomedical informatics*, *71*, 31–48.

Schartel, S. A. (2012). Giving feedback–an integral part of education. *Best practice & research Clinical anaesthesiology*, *26*(1), 77–87.

Schell, J. (2008). *The art of game design: A book of lenses.* CRC press.

Schoech, D., Boyas, J. F., Black, B. M., & Elias-Lambert, N. (2013). Gamification for behavior change: Lessons from developing a social, multiuser, web-tablet based prevention game for youths. *Journal of Technology in Human Services*, *31*(3), 197–217.

Schönen, R. (2014). Gamification in change management processes. *Munich University, Munich. Bachelor Thesis*.

Schrepp, M. (2015). User experience questionnaire handbook. *All you need to know to apply the UEQ successfully in your project*.

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of human-computer studies*, *74*, 14–31.

Seidler, A.-R., Henkel, C., Fiedler, M., Kranz, J., Ixmeier, A., & Strunk, K. S. (2020). Promoting eco-sustainable behavior with gamification: An experimental study on the alignment of competing goals. In *Icis.*

Seif El-Nasr, M., Aghabeigi, B., Milam, D., Erfani, M., Lameman, B., Maygoli, H., & Mah, S. (2010). Understanding and evaluating cooperative games. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 253–262).

Serral, E., De Weerdt, J., Sedrakyan, G., & Snoeck, M. (2016). Automating immediate and personalized feedback taking conceptual modelling education to a next level. In *2016 ieee tenth international conference on research challenges in information* *science (rcis)* (pp. 1–6).

Shin, D. (2017). The role of affordance in the experience of virtual reality learning: Technological and affective affordances in virtual reality. *Telematics Informatics*, *34*(8), 1826–1836. Retrieved from https://doi.org/10.1016/j.tele.2017.05.013 doi: 10.1016/j.tele.2017.05.013

Shindler, J. (2009). *Transformative classroom management: Positive strategies to engage all students and promote a psychology of success.* John Wiley & Sons.

Sillaots, M. (2014). Achieving flow through gamification: A study on re-designing research methods courses. In *European conference on games based learning* (Vol. 2, p. 538).

Smith, R., & Bryant, R. (1975). Metal substitutions in carbonic anhydrase: A halide ion probe study. *Biochemical and Biophysical Research Communications*, *66*(4), 1281-1286. Retrieved from https://www.sciencedirect.com/science/article/pii/0006291X75904982 doi: https://doi.org/10.1016/0006-291X(75)90498-2

Soderberg, C. K., Errington, T. M., Schiavone, S. R., Bottesini, J., Thorn, F. S., Vazire, S., ... Nosek, B. A. (2021). Initial evidence of research quality of registered reports compared with the standard publishing model. *Nature Human Behaviour*, *5*(8), 990–997.

Sordello, R., Ratel, O., Flamerie De Lachapelle, F., Leger, C., Dambry, A., & Vanpeene, S. (2020). Evidence of the impact of noise pollution on biodiversity: A systematic map. *Environmental Evidence*, *9*, 1–27.

Spiller, T. R., & Olff, M. (2018). *Reducing uncertainty in research: Introducing registered reports* (Vol. 9) (No. 1). Taylor & Francis.

Strmecki, D., Bernik, A., & Radosevic, D. (2015). Gamification in e-learning: Introducing gamified design elements into e-learning systems. *J. Comput. Sci.*, *11*(12), 1108–1117.

Ströckl, D. E., Oberrauner, E., Krainer, D., Wohofsky, L., & Oberzaucher, J. (2020). Hedonic and pragmatic qualities of aal pilot region technologies: Evaluation with the attrakdiff survey. In *Ict4awe* (pp. 187–193).

Su, X., Li, X., & Kang, Y. (2019). A bibliometric analysis of research on intangible cultural heritage using citespace. *Sage Open*, *9*(2), 2158244019840119.

Taspinar, B., Schmidt, W., & Schuhbauer, H. (2016). Gamification in education: A board game approach to knowledge acquisition. *Procedia Computer Science*, *99*, 101–116.

Tauer, J. M., & Harackiewicz, J. M. (2004). The effects of cooperation and competition on intrinsic motivation and performance. *Journal of personality and social psychology*, *86*(6), 849.

Teh, N., Schuff, D., Johnson, S., & Geddes, D. (2013). Can work be fun? improving task motivation and help-seeking through game mechanics.

Tekinbas, K. S., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals.* MIT press.

Teng, C.-I., & Chen, W.-W. (2014). Team participation and online gamer loyalty. *Electronic Commerce Research and Applications*, *13*(1), 24–31.

Thiel, S.-K., Reisinger, M., & Röderer, K. (2016). " i'm too old for this!" influence of age on perception of gamified public participation. In *Proceedings of the 15th international conference on mobile and ubiquitous multimedia* (pp. 343–346).

Thorndike, E. L. (1927). The law of effect. *The American journal of psychology*, *39*(1/4), 212–222.

Thousand, J. S., Villa, R. A., & Nevin, A. I. (2002). *Creativity and collaborative learning: The practical guide to empowering students, teachers, and families.* ERIC.

Tisza, G., & Markopoulos, P. (2021a). Funq: Measuring the fun experience of a learning activity with adolescents. *Current Psychology*, 1–21.

Tisza, G., & Markopoulos, P. (2021b). Understanding the role of fun in learning to code. *International Journal of Child-Computer Interaction*, *28*, 100270.

Toda, A., Klock, A., Pereira, F. D., Rodrigues, L. A., Palomino, P. T., Lopes, V., . . . others (2022). Towards the understanding of cultural differences in between gamification preferences: A data-driven comparison between the us and brazil. In *Proceedings of the 15th international conference on educational data mining* (p. 560).

Toda, A. M., Klock, A. C., Oliveira, W., Palomino, P. T., Rodrigues, L., Shi, L., . . .

Cristea, A. I. (2019). Analysing gamification elements in educational environments using an existing gamification taxonomy. *Smart Learning Environments*, *6*(1), 1–14.

Toda, A. M., Oliveira, W., Klock, A. C., Palomino, P. T., Pimenta, M., Gasparini, I., ... Cristea, A. I. (2019). 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, pp. 84–88).

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. an overview of some recommended measures of effect size. *TRENDS in Sport Sciences*.

Tomé Klock, A. C., Cunha, L. F. d., Carvalho, M. F. d., Eduardo Rosa, B., Jaqueline Anton, A., & Gasparini, I. (2015). Gamification in e-learning systems: A conceptual model to engage students and its application in an adaptive e-learning system. In *International conference on learning and collaboration technologies* (pp. 595–607).

Tondello, G. F., Mora, A., & Nacke, L. E. (2017). Elements of gameful design emerging from user preferences. In *Proceedings of the annual symposium on computer-human interaction in play* (pp. 129–142).

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The gamification user types hexad scale. In *Proceedings of the 2016 annual symposium on computer-human interaction in play* (pp. 229–243).

Trinidad, M., Calderón, A., & Ruiz, M. (2021). Gorace: a multi-context and narrative-based gamification suite to overcome gamification technological challenges. *IEEE Access*, *9*, 65882–65905.

Trinidad, M., Ruiz, M., & Calderón, A. (2021). A bibliometric analysis of gamification research. *IEEE Access*, *9*, 46505–46544.

Tuomela, R. (2000). Collective and joint intention. *Mind & Society*, *1*(2), 39–69.

Tuomela, R. (2005). We-intentions revisited. *Philosophical Studies*, *125*(3), 327–369.

Tylor Edward, B. (1871). Primitive culture, vol. 1. *Henry Holt, New York*.

Unep, A., & ASSESSMENT, I. R. R. (2016). The rise of environmental crime. *Nairobi: UNEP*.

Vasconcelos-Raposo, J., Bessa, M., Melo, M., Barbosa, L., Rodrigues, R., Teixeira, C. M., ... Sousa, A. A. (2016). Adaptation and validation of the igroup presence questionnaire (ipq) in a portuguese sample. *Presence*, *25*(3), 191–203.

Vassileva, J. (2012). Motivating participation in social computing applications: a user modeling perspective. *User Modeling and User-Adapted Interaction*, *22*(1), 177–201.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425–478.

Verhoeff, T. (1997). The role of competitions in education. *Future world: Educating for the 21st century*, 1–10.

Vockell, E. (2004). Educational psychology: A practical approach. *Purdue University*.

Walk, W., Görlich, D., & Barrett, M. (2017). Design, dynamics, experience (dde): an advancement of the mda framework for game design. In *Game dynamics* (pp. 27–45). Springer.

Wang, Y.-S., Wu, M.-C., & Wang, H.-Y. (2009). Investigating the determinants and age and gender differences in the acceptance of mobile learning. *British journal of educational technology*, *40*(1), 92–118.

Weisberg, M. (2011). Student attitudes and behaviors towards digital textbooks. *Publishing research quarterly*, *27*(2), 188–196.

Werbach, K., Hunter, D., & Dixon, W. (2012). *For the win: How game thinking can revolutionize your business* (Vol. 1). Wharton digital press Philadelphia.

Whitson, J. R. (2013). Gaming the quantified self. *Surveillance & Society*, *11*(1/2), 163–176.

Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational psychologist*, *52*(1), 17–37.

Wohlin, C., Kalinowski, M., Felizardo, K. R., & Mendes, E. (2022). Successful combination of database search and snowballing for identification of primary studies in systematic literature studies. *Information and Software Technology*, *147*, 106908.

Wood, T. J., Kaplan, I., & Szendrei, Z. (2018). Wild bee pollen diets reveal patterns of seasonal foraging resources for honey bees. *Frontiers in Ecology and Evolution*, *6*, 210.

Woolson, R. F. (2007). Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, 1–3.

Wright, L. T., Nancarrow, C., & Kwok, P. M. (2001). Food taste preferences and cultural influences on consumption. *British Food Journal*.

Xie, P. (2015). Study of international anticancer research trends via co-word and document co-citation visualization analysis. *Scientometrics*, *105*(1), 611–622.

Yee, N. (2006). Motivations for play in online games. *CyberPsychology & behavior*, *9*(6), 772–775.

Yen, P.-Y., Wantland, D., & Bakken, S. (2010). Development of a customizable health it usability evaluation scale. In *Amia annual symposium proceedings* (Vol. 2010, p. 917).

Yigitbas, E., Schmidt, M., Bucchiarone, A., Bassanelli, S., & Engels, G. (2024). *Gamification- and virtual reality-based learning environment for uml class diagram modeling.* (in submission)

Yigitbas, E., Schmidt, M., Bucchiarone, A., Gottschalk, S., & Engels, G. (2024). Gamovr: Gamification-based uml learning environment in virtual reality. *Science of Computer Programming*, *231*, 103029.

Yu, F.-Y., Chang, L.-J., Liu, Y.-H., & Chan, T.-W. (2002). Learning preferences towards computerised competitive modes. *Journal of Computer Assisted Learning*, *18*(3), 341–350.

Yusof, A., Atan, N. A., Harun, J., Salleh, S., & Khamis, N. (2022). Gamified hybrid service-learning for students engagement in computer network course (gamysel). *INNOVATING EDUCATION FOR A BETTER TOMORROW*, 502-506.

Zahedi, L., Batten, J., Ross, M., Potvin, G., Damas, S., Clarke, P., & Davis, D. (2021). Gamification in education: A mixed-methods study of gender on computer science students' academic performance and identity development. *Journal of Computing in Higher Education*, *33*(2), 441–474.

Zhou, W., Chen, Q., & Meng, S. (2019). Knowledge mapping of credit risk research: scientometrics analysis using citespace. *Economic research-Ekonomska istraživanja*, *32*(1), 3451–3478.

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in web and mobile apps.* " O'Reilly Media, Inc.".